A Sample Size Calculation for Training and Certifying Targeting Policies

Duncan Simester*, Artem Timoshenko†, and Spyros I. Zoumpoulis‡

*Sloan School of Management, Massachusetts Institute of Technology

†Kellogg School of Management, Northwestern University

‡Decision Sciences, INSEAD

July 16, 2024

We propose an approach for determining the sample size required when using an experiment to train and certify a targeting policy. Calculating the rate at which the performance of a targeting model improves with additional training data is a complex problem. We address this challenge by assuming that customers are grouped into segments that capture relevant information about their responsiveness to the firm's marketing actions. We consider two problem formulations. The first formulation identifies the sample size required to train a targeting policy and certify that its expected performance exceeds a predefined threshold. The second formulation identifies the sample size required to train a targeting policy and certify that it outperforms a baseline in an out-of-sample statistical test. We establish theoretical properties of these problems, based on which we propose computationally efficient algorithms for optimal sample size calculations. We illustrate our algorithms and analysis using data from a luxury fashion retailer.

1 Introduction

Firms often conduct experiments to measure the effectiveness of their marketing actions. The sample size calculations for these experiments can be guided by established methods, such as power calculations. Firms have recently started using randomized experiments for a different purpose: to train and certify targeting (personalization) policies. For example, retailers mail different promotional offers to customers in different zip codes, online platforms show different digital advertisements to customers with different browsing histories, sales representatives recommend different products to customers with different behavioral profiles. A new question arises: what is the sample size required when using an experiment to train and certify a targeting policy?[1]

Consider, for example, a large department store that sends a catalog to selected customers each Christmas season. Because the catalog is costly to print and distribute, the retailer wanted to optimize which customers to mail the catalog to. The department store first conducted an experiment in which customers were randomly assigned to either receive or not receive the catalog. The experimental data was then used to train and evaluate a targeting policy, which recommends who receives the catalog in future years.

This process exhibits several characteristics typical to many firms in different industries. First, the retailer wants to target its marketing actions, so that different customers receive different marketing actions. This contrasts with a uniform policy, in which all customers receive the same action. Second, the firm uses experimental data to design a targeting policy. The targeting policy allocates marketing actions based on their relative effectiveness for different customers, and the experimental variation ensures that the relationship between the marketing actions and the customer outcomes is causal (and not merely correlational). Finally, the firm collects data in a single batch, with multiple customers engaged in a single experiment in a single time period.[2] This contrasts with an online learning process, which uses a sequence of small batches. One-shot batch experiments are particularly common in non-digital marketing channels (direct mail, telephone, or in-person channels). Before conducting the experiment, firms must decide how many customers to include in the experiment. We propose an approach for answering this question.

We consider two problem formulations that differ in how the firm uses the experimental data. The first formulation focuses on the expected profit improvement of the targeting policy compared to a baseline policy. Targeting policies yield higher expected performance with more training data (*policy training*). We provide algorithms that estimate the minimal sample size required to reach a predefined threshold on the expected performance improvement (*policy certification*).

[1] We introduce the term "policy certification" to describe the process of evaluating whether the performance of a targeting policy satisfies prespecified criteria.

[2] The catalog experiment takes at least four months to design and implement. Catalog mailing decisions have a lead time of six to eight weeks, implementing the mailing decisions requires two weeks, and observing the outcomes typically requires an additional eight weeks. Thus, firms often cannot conduct a series of experiments within the same season. Instead, experimental data are obtained from a single experiment, and the sample size for this experiment is chosen in advance.

The second formulation identifies the sample size required to train a targeting policy and certify that the trained policy can outperform a baseline in an out-of-sample statistical test with predefined confidence and power. Compared to standard power calculations, our framework recognizes that firms can divide the experiment data into two samples and use each sample for a different purpose. One sample is used to update information about the treatment effects prior to the certification, and to improve the targeting policy (*policy training*). The other sample is used to certify the performance of the targeting policy out of sample (*policy certification*). This introduces a tradeoff; more training data can reduce the amount of data required for certification. We develop an approach that recommends both how much experimental data to collect in total, and how to allocate experimental data into training and testing samples. In our empirical application, training and certifying a targeting policy requires a total sample size that is over seven times smaller than the data required to certify an alternative targeting policy defined using only prior beliefs (without additional training).

Both problem formulations require modeling how the expected performance of a targeting policy varies with the amount of training data, which is an inherently challenging problem. In general, the relationship between policy performance and the amount of training data depends on the specifics of the problem (including the joint distribution of the descriptive and outcome variables), as well as the choice of the estimator used to design the targeting policy. Current frameworks cannot characterize this relationship without additional assumptions.

We make the problem tractable by assuming that customers are grouped into predefined segments. The customer segmentation captures all of the relevant information about how customers respond to the firm's marketing actions, including any information potentially contained in covariates.[3] We further assume that the segment treatment effects represent draws from independent segment-specific distributions. Similar segments can have similar prior distributions, but the draws are independent across the segments. As a result, the firm cannot update beliefs about treatment effects in one segment by observing outcomes in a different segment. Together these assumptions represent an approximation and allow us to make progress on an important business problem.

We formalize the two problem formulations as optimization problems. We then establish properties of these problems, based on which we propose computationally efficient algorithms for optimal sample size calculations. The algorithms use prior beliefs about the treatment effects and account for sampling variance to form expectations about the outcomes of the experiment. The priors describe the information that the firm has *a priori* about the size of the treatment effects and the associated uncertainty. Consistent with a Bayesian framework, we consider targeting policies that maximize the firm's expected profit after observing the experimental outcomes, and propose algorithms to identify the minimum sample sizes for policy training and certification.

Our algorithms can either use simulations to obtain the recommended sample size for the experiments, or identify a solution using analytical approximations. Empirically, we verify that the analytical approximation-based solutions yield similar sample size estimates, while achieving a dra-

[3]We assume that the firm does not optimize the definition of the segments during policy training.

matic improvement in runtime (compared to the simulation-based solutions). We conclude that our analytical approximations provide practical solutions that offer computational efficiency with little loss of accuracy. However, if a manager favors accuracy over computational efficiency, we recommend using the simulation-based algorithms.

We illustrate the proposed algorithms and evaluate sensitivity of the algorithms to the input parameters using data from a luxury fashion retailer.

Related literature

We contribute to the literatures on experimental design and targeting. Choosing a sample size is an important experimental design decision, and a standard approach is to conduct a power calculation. Power calculations recommend an experimental sample size so that a statistical test conducted with the sample yields a sufficient probability of correctly rejecting a null hypothesis. For example, Anderson et al. (2018) used power calculations to design an experiment to test a relationship between an entrepreneur's financial performance and business training, and concluded that 750 firms is a sufficient sample for their test. Power calculations can also be used to evaluate whether it is feasible to test a hypothesis using an experiment. For example, Lewis and Rao (2015) demonstrate that it is often infeasible to detect a significant treatment effect for digital advertising with sufficient power, even at a large advertising platform. A related literature proposes sample size calculations for accuracy in parameter estimation (AIPE), see for example Maxwell et al. (2008).

Field experiments can provide data not only for hypothesis testing and parameter estimation, but also for choosing the best treatment. For example, firms can conduct experiments to measure whether price discounts are more effective than free trials, and use the findings to choose which marketing actions to implement. The hypothesis testing and profit optimization objectives are not always aligned. In particular, there is an exploration-exploitation tradeoff: designing an experiment to maximize learning will not maximize immediate profits, and vice versa. Multi-armed bandit methods explicitly model this tradeoff in a sequence of experiments (see for example Hauser et al., 2009; Schwartz et al., 2017). Feit and Berman (2019) provide closed-form expressions for the profit-maximizing sample size for a single experiment. These expressions explicitly consider a tradeoff between the size of the training sample and the remaining population in the implementation phase, and show that the profit-maximizing sample size can be substantially smaller than the sample size typically recommended for hypothesis testing. Blattberg (1979) and Ginter et al. (1981) employed statistical decision theory, as opposed to statistical hypothesis testing, to inform sample size decisions for designing advertising experiments.

A related literature uses statistical decision theory and a minimax regret loss function to inform treatment choice from data generated by a randomized experiment. Manski (2004) derives a closed-form bound on the maximum regret for a class of treatment rules, which in turn yields sufficient sample sizes for productive use of covariate information. Similar to our work, Manski considers the problem of learning optimal *targeting* policies, allowing for population heterogeneity. Tetenov

(2012) extends the work by Manski (2004), by considering decision criteria that asymmetrically treat Type I regret and Type II regret, but without allowing for targeting. Joo and Chiong (2023) develop the "asymptotic minimax-regret" criterion, a general large-sample approximation of the minimax-regret criterion, which may be helpful for sample size decisions, again without allowing for targeting.

Researchers from different academic disciplines have proposed methods for targeting and personalization and have applied targeting methods in different domains (Sutton and Barto, 2018; Wager and Athey, 2018). In marketing, the applications include personalized pricing (Dubé and Misra, 2019), catalog mailing (Hitsch and Misra, 2018), targeted promotions (Zhang and Krishnamurthi, 2004; Simester et al., 2020b; Gabel and Timoshenko, 2022; Daljord et al., 2023), and digital advertising (Rafieian and Yoganarasimhan, 2021). These targeting applications often leverage field experiments to train targeting policies and evaluate their performance.

We contribute to the literatures on experimental design and targeting by proposing an approach for sample size calculation for targeting. In Section 3, we focus on training a targeting policy. Our method estimates how much training data firms need to design a targeting policy that satisfies a predefined performance requirement. This objective is similar in spirit to the profit-maximizing sample size calculation proposed by Feit and Berman (2019). They focus on uniform policies and investigate the sample size that maximizes expected profit over both a learning and an implementation phase, when the total population across the two phases is limited. In contrast, we study targeting policies, in which the recommended actions vary across customer segments, and we investigate the required sample size for training a targeting policy that satisfies an expected performance requirement in the implementation phase.

In Section 4, we focus on certifying a targeting policy through significance testing. We propose a method for calculating a sample size required to certify that a targeting policy can outperform a benchmark policy in a statistical test. Compared to standard power calculations, our framework recognizes that firms can use experimental data to both train a targeting policy and evaluate the policy out-of-sample. Our method answers two questions: how much experimental data to obtain, and how to allocate this dataset between policy training and policy certification. We formulate our approach using a Bayesian framework, which allows us to incorporate the firm's prior information about treatment effects in each segment when calculating sample sizes.

Outline of the paper

The paper continues in Section 2, with the introduction of a formal model. In Section 3, we present our first formulation, which calculates the amount of training data required to satisfy a predefined performance requirement over a benchmark policy. We introduce a second formulation in Section 4, for firms that require out-of-sample certification of the trained policy. An empirical application is presented in Section 5, where we discuss how firms can obtain inputs for our algorithms, and use data from a luxury retailer to illustrate the algorithms. The paper concludes in Section 6 with a review of the findings and discussion of opportunities for future research.

2 Model

We consider a single firm in a market with C exogenously defined customer segments, and denote the proportion of the population in segment c as w_c. The firm has two marketing actions: treatment and control. For customer i in segment c, we define a monetary outcome Y_i as

$$Y_i = \alpha_c + m_c \cdot D_i + \epsilon_i, \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, s_c^2) \tag{1}$$

where m_c is a segment-specific treatment effect for segment c, $D_i = 1$ if customer i receives the treatment, and $D_i = 0$ if customer i receives the control. We assume that the treatment effects m_c incorporate any costs related to implementing the marketing actions, and they are unknown to the firm, independent across segments, and distributed according to a normal prior $m_c \sim \mathcal{N}\left(\mu_c, \sigma_c^2\right)$, $c = 1, .., C$. The parameters μ_c and σ_c summarize the firm's prior beliefs about the size of the treatment effects and the associated uncertainty. The parameters α_c and s_c define segment-specific intercepts and the variance of idiosyncratic noise. We note that the number of parameters in the model grows linearly with the number of segments.

We compare monetary outcomes under the treatment and control conditions, so α_c plays no role in our analysis. In our primary specification, monetary outcomes Y_i can occur over the full support of positive and negative outcomes. We consider an alternative model specification using a log-normal distribution of the monetary outcomes in Appendix C.2.7.

Customer segmentation is an important assumption in our model. We assume that customers are grouped into mutually exclusive customer segments that are exogenously defined, known to the firm, and fully summarize contextual information about the responsiveness to the firm's marketing actions. We emphasize that the assumption that segmentation is known does not require that the segments are "optimal." Instead, we simply require that a segmentation is defined *a priori*. For example, it could capture demographic information, historical purchasing differences between customers, or responsiveness to previous marketing actions. The segment structure could be obtained using segmentation (clustering) algorithms, or a factorial combination of categorical covariates. What is critical is that a segmentation exists, and the firm does not adjust the segments once it receives the training data.

We further assume that the treatment effects distributions are independent across segments. Similar segments can have similar prior distributions, by having similar μ's and/or σ's, but the actual treatment effects represent independent draws from segment-specific prior distributions. The independence assumption implies that firm cannot update beliefs about treatment effects in one segment by observing outcomes in a different segment. The independence assumption enables analytical results for sample size calculation, and it simplifies obtaining priors in practical applications (there is no need to specify priors on the correlations). In Section 3.4, we discuss an extension in which we relax this assumption. In particular, we use a multivariate normal prior to model correlations in treatment effects across the segments.

2.1 A single experiment

The firm plans an experiment to obtain data for targeting applications. After conducting the experiment, the firm will observe the monetary outcome for each customer that participated, and use these outcomes to train and evaluate a targeting policy. Our analysis focuses on calculating the required sample size for this experiment.

The experimental design randomly assigns customers into treatment and control conditions. We assume that the experiment is balanced across the two experimental conditions (treatment and control) in each segment, and the total sample size is allocated proportionally to the segments' relative sizes w_c.[4] We denote the total experimental sample size as N. The firm's goal is to identify the minimum sample size for the experiment, N, such that the trained targeting policy meets a managerial certification requirement. The firm wants to decide on this sample size before conducting the experiment.

The data from the experiment is used differently in our two certification formulations. In Section 3, we estimate the experimental sample size required to train a targeting policy that is expected to outperform a benchmark policy by a threshold margin. In this section, the experimental data is retained in a single sample, and is all used to train the targeting policy. In Section 4, we estimate the experimental sample size required to train a targeting policy and certify that it outperforms a benchmark in an out-of-sample statistical test. In contrast to the formulation in Section 3, the experimental data is *randomly* divided into two subsamples. The training subsample is used to train the targeting policy, and the testing subsample is used to conduct the statistical test.[5] Our analysis determines both the total required experimental sample size, and the size of the training and testing subsamples.[6]

The label "training data" describes both the complete experimental data in Section 3, and the subsample of the experimental data used to train the targeting policy in Section 4. The label "testing data" describes the second subsample of the experimental data used to conduct the statistical test in Section 4. We use N_{tr} to describe the sample size of the training data, and N_{test} to describe the

[4] We focus our analysis on the experiment size and not the experiment design. In Section 3.4 and Appendix C.2.5, we discuss an extension in which we optimally allocate the experimental sample size across the different segments (instead of allocating the sample proportionally according to segment size).

[5] Specifically, we assume that the random sampling is stratified, so that the training data contains $\frac{1}{2}w_cN_{tr}$ customers in the treatment and $\frac{1}{2}w_cN_{tr}$ customers in the control condition in each segment $c \in \{1, .., C\}$. The testing data contains $\frac{1}{2}w_cN_{test}$ customers in the treatment and $\frac{1}{2}w_cN_{test}$ customers in the control condition in each segment $c \in \{1, .., C\}$.

[6] The use of separate datasets for training and evaluating a targeting policy helps to ensure that the evaluation is not susceptible to over-optimism due to training and evaluating the policy with the same idiosyncratic noise. Alternatively, the policy could be trained and evaluated using a k-fold cross-validation approach on the entire experimental sample. The cross-validation approach could reduce the variance of the estimates, and consequently reduce the required sample sizes.

sample size of the testing data. In Section 3, $N = N_{tr}$, while in Section 4, $N = N_{tr} + N_{test}$.

For ease of exposition, throughout the paper we describe how the firm uses the experimental data using the present tense: "The firm uses the experimental data." However, it is helpful to keep in mind that the firm makes the sample size decision before observing the outcomes of the experiment, and when making that decision, it anticipates how it will use the data (once it is available).

2.2 Training a targeting policy

A targeting policy recommends different marketing actions (treatment or control) to different customers. In our setting, we consider targeting policies that operate at the segment level and apply the same action to all customers within a segment. So a targeting policy π is a function $\pi : \{1, \ldots, C\} \longrightarrow \{0, 1\}$, where a mapping to 1 (0) denotes assigning the treatment (control) action to that segment.

The firm uses the training data to design a targeting policy. Recall that the firm randomly selects N_{tr} customers for the training data, including $\frac{1}{2}w_c N_{tr}$ customers in the treatment and $\frac{1}{2}w_c N_{tr}$ customers in the control condition in each segment c, $\forall c \in \{1, .., C\}$. Using the training data, the firm can estimate the observed treatment effect for each segment.

$$\hat{m}_c^{tr} := \bar{Y}_{c|1}^{tr} - \bar{Y}_{c|0}^{tr}, \qquad c = 1, .., C,$$

where, in segment c, $\bar{Y}_{c|1}^{tr}$ and $\bar{Y}_{c|0}^{tr}$ are the average observed outcomes in the treatment and control conditions in the training sample. We note that $\hat{m}_c^{tr}$ is a sufficient statistic for the underlying parameter m_c. By standard Bayesian updating, the firm's posterior for the segment treatment effect m_c given $\hat{m}_c^{tr}$ is:[7]

$$m_c \mid \hat{m}_c^{tr} \sim \mathcal{N}\left(\mu_c^{post}, (\sigma_c^{post})^2\right), \tag{2}$$

where the posterior parameters μ_c^{post} and σ_c^{post} are given by

$$\mu_c^{post} = \left(\frac{1}{\sigma_c^2} + \frac{w_c N_{tr}}{4s_c^2}\right)^{-1} \left(\frac{\mu_c}{\sigma_c^2} + \frac{w_c N_{tr}}{4s_c^2} \cdot \hat{m}_c^{tr}\right) \tag{3}$$

$$(\sigma_c^{post})^2 = \left(\frac{1}{\sigma_c^2} + \frac{w_c N_{tr}}{4s_c^2}\right)^{-1} . \tag{4}$$

Notice that for a given $N_{tr} > 0$, the posterior mean μ_c^{post} follows a normal distribution

$$\mu_c^{post} \sim \mathcal{N}\left(\mu_c, \frac{\sigma_c^2}{k_c^2}\right), \tag{5}$$

[7] We apply Bayesian updating for a normal-normal model where a new data point is the observed average treatment effect in segment c with a likelihood function given by $\hat{m}_c^{tr} \mid m_c \sim \mathcal{N}\left(m_c, \frac{4s_c^2}{w_c N_{tr}}\right)$.

where

$$k_c = \sqrt{1 + \frac{4s_c^2}{w_c N_{tr} \sigma_c^2}},$$

and the posterior standard deviation σ_c^{post} is a constant.

We assume the firm uses the optimal targeting policy within the proposed Bayesian framework given the available information (priors and experimental outcomes). In particular, we study the targeting policy that only treats customers in segments with a non-negative posterior mean:

$$\pi_t(c) := I\left(\mu_c^{post} \geq 0\right) = \begin{cases} 1 & \text{if } \mu_c^{post} \geq 0 \\ 0 & \text{if } \mu_c^{post} < 0, \end{cases} \tag{6}$$

where $\pi_t(c) = 1(0)$ indicates that the policy assigns the treatment (control) to all customers in segment c, and where $I(\cdot)$ denotes the indicator function.[8] The targeting policy π_t fully incorporates the prior beliefs and information from the observed outcomes in the training data, and maximizes the expected monetary outcome for every customer.

2.3 Evaluating a targeting policy

The firm can use the testing data to evaluate the performance of a targeting policy. In particular, for each segment c, the firm can use the testing data to estimate the average per-customer monetary outcome in the treatment and control conditions ($\bar{Y}_{c|1}^{test}$ and $\bar{Y}_{c|0}^{test}$). The average outcomes can then be aggregated across customer segments to produce the average per-customer policy performance $\bar{V}_\pi$ for any policy π:

$$\bar{V}_\pi := \sum_{c=1}^{C} w_c \cdot \bar{Y}_{c|\pi}^{test},$$

where $\bar{Y}_{c|\pi}^{test}$ is the average per-customer monetary outcome in segment c in the condition defined by policy recommendation $\pi(c)$ in the testing sample.

The firm compares the performance of the trained targeting policy π_t to a benchmark policy. We consider the benchmark policy to be a uniform policy of not treating anyone, π_0: $\pi_0(c) = 0$, $\forall c \in \{1, .., C\}$, and we generalize our results to an arbitrary benchmark policy in Appendix C.2.1. The

[8] Alternatively, we could model the firm's decision rule as the more naïve test of whether the observed treatment effect in the segment is non-negative, $\hat{m}_c^{tr} \geq 0$. In the limit of large prior variance σ_c^2 in a segment, the naïve and the Bayesian decision rules coincide.

performance difference between the trained targeting policy and the benchmark control policy is:

$$\begin{aligned}
\bar{V}_{\pi_t} - \bar{V}_{\pi_0} &= \sum_{c=1}^{C} w_c \cdot \left(\bar{Y}^{test}_{c|\pi_t} - \bar{Y}^{test}_{c|\pi_0}\right) \\
&= \sum_{c=1}^{C} w_c \cdot I\big(\mu^{post}_c \geq 0\big) \cdot \left(\bar{Y}^{test}_{c|1} - \bar{Y}^{test}_{c|0}\right) \\
&= \sum_{c=1}^{C} w_c \cdot I\big(\mu^{post}_c \geq 0\big) \cdot \hat{m}^{test}_c, \qquad (7)
\end{aligned}$$

where we define $\hat{m}^{test}_c := \bar{Y}^{test}_{c|1} - \bar{Y}^{test}_{c|0}$ as the average treatment effect in segment c estimated using the testing sample.

Estimating $\bar{V}_{\pi_t} - \bar{V}_{\pi_0}$ requires using a part of the experimental data for policy evaluation. However, firms can also form an expectation about the policy performance without any testing data. After obtaining the training data and updating beliefs about the distribution of the treatment effects, the firm can calculate an expectation of the per-customer policy performance difference:

$$\begin{aligned}
V^{post}_{\pi_t-\pi_0} &:= \mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}^{tr}_1, \ldots, \hat{m}^{tr}_C\right] \qquad (8) \\
&= \sum_{c=1}^{C} w_c \cdot I\big(\mu^{post}_c \geq 0\big)\, \mathbb{E}\left[m_c \mid \hat{m}^{tr}_c\right] \\
&= \sum_{c=1}^{C} w_c \cdot I\big(\mu^{post}_c \geq 0\big) \cdot \mu^{post}_c \qquad (9)
\end{aligned}$$

The expected policy performance difference $V^{post}_{\pi_t-\pi_0}$ is a theoretical measure. It captures the firm's expectations about the per-customer policy performance after observing the training data, as if the policy were implemented on a testing sample. In contrast, $\bar{V}_{\pi_t} - \bar{V}_{\pi_0}$ is an observed policy difference calculated using the actual testing data from the experiment. We consider $V^{post}_{\pi_t-\pi_0}$ in our formulation of policy certification based on the expected performance gains in Section 3, and we incorporate $\bar{V}_{\pi_t} - \bar{V}_{\pi_0}$ into our formulation of policy certification based on a statistical test in Section 4. We provide a complete table of notations in Appendix A.

3 Certification Based on Expected Improvement

In this section, we propose an approach for sample size calculations for training a targeting policy. More training data yield targeting policies with higher expected performance. Our approach estimates how much training data is required to train a targeting policy that is expected to outperform a benchmark policy by a margin δ. We conclude the section by considering several extensions to this formulation.

3.1 Problem formulation

For $\delta > 0$, we say that a trained targeting policy is *δ-improvement certified*, if the expected per-customer out-of-sample performance difference between the trained targeting policy and the benchmark is at least δ: $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] \geq \delta$. The parameter δ is a managerial requirement set by the firm.

Recall that the firm decides on the sample size of the experiment before running the experiment. We formulate the firm's decision problem as the optimization problem[9]

$$\min_{N_{tr}} \quad N_{tr} \tag{10}$$

$$\text{s.t.} \quad \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] \geq \delta \tag{11}$$

$$N_{tr} \geq 0. \tag{12}$$

There are two interpretations of the proposed certification constraint (11). The first interpretation is related to the expected policy performance *before* the experiment. For any sample size N_{tr}, a firm can use prior beliefs about the distribution of the treatment effects to calculate an expectation of the policy performance difference $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$. The firm can then find a minimal sample size that satisfies a predefined performance threshold δ.

An alternative interpretation recognizes that *after* the experiment is conducted, firms can update their prior beliefs about treatment effects, and calculate the expected policy performance difference conditional on the training data. Recall that in Equation (8) we denote the expected policy performance difference *after* the experiment by $V^{post}_{\pi_t-\pi_0}$. Different draws of the training data yield different conditional expectations $V^{post}_{\pi_t-\pi_0}$, so $V^{post}_{\pi_t-\pi_0}$ is a random variable that depends on the draw of the training data. Firms can require that the expectation of $V^{post}_{\pi_t-\pi_0}$ is at least δ:

$$\mathbb{E}\left[V^{post}_{\pi_t-\pi_0}\right] \geq \delta.$$

By the law of iterated expectations, these two interpretations are equivalent:

$$\mathbb{E}\left[V^{post}_{\pi_t-\pi_0}\right] = \mathbb{E}\left[\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}^{tr}_1, \ldots, \hat{m}^{tr}_C\right]\right] = \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$$

The parameter δ is motivated by the cost of adopting a new policy. Consider a risk-neutral firm that is only concerned about expected profit, and only requires that the trained targeting policy has at least as high an expected profit as the benchmark policy. This is equivalent to setting $\delta = 0$. As long as there exists a segment for which the prior beliefs favor deviating from the benchmark policy, the expected profit will strictly improve compared to the benchmark policy. If so, this firm would never collect any training data: $N_{tr} = 0$ would be optimal. Yet this is obviously not what

[9] In Sections 3 and 4 we formulate optimization problems with continuous decision variables for the sample sizes. We acknowledge that the practical problem has integrality requirements, and therefore our proposed algorithms return integer solutions by applying a ceiling function (bracket notation).

we observe in practice, as many firms conduct experiments to train and certify targeting policies.

One explanation is that there is a cost of migrating from the benchmark policy. For example, the process for updating a policy typically starts with the data science team presenting the proposed change and anticipated performance improvement to a business team, who approves the change, and then updates its forecasts. After implementing the change, an additional process is generally required to confirm that the performance of the new policy is meeting expectations. These coordination and monitoring costs provide motivation for the δ-expected improvement formulation, where the (per-customer) cost of adopting the new policy is represented by $\delta > 0$.

We next investigate the properties of the expected performance difference $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$. In particular, we show that (1) the expected profit improvement increases monotonically with the size of the training sample, (2) even without training data, a targeting policy constructed just using priors yields a non-negative expected profit improvements, and (3) there is a finite upper bound on the expected profit improvement. We leverage these properties to propose a computationally efficient algorithm for the firm's optimization problem (10)-(12). In Section 5, we use an empirical application to illustrate these properties (see Figure 1).

3.2 Analysis and algorithm for δ-expected improvement certification

We derive an explicit expression for the left-hand side (LHS) of constraint (11). In Appendix B.1 we demonstrate that the expected performance difference $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$ is

$$\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] = \begin{cases} \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \quad (13a) \\ \sum_{c=1}^{C} w_c \cdot I\left(\mu_c \geq 0\right) \cdot \mu_c & \text{if } N_{tr} = 0, \quad (13b) \end{cases}$$

where

$$k_c := \sqrt{1 + \frac{4 s_c^2}{w_c N_{tr} \sigma_c^2}}.$$

We can characterize the properties of the optimization problem for δ-expected improvement. The first observation is that, at optimality, at least one of the constraints (11) and (12) is binding. This follows by continuity in N_{tr} of the left-hand sides of the constraints. In particular, if neither of the constraints were binding at the optimal solution, then we could slightly decrease the decision variable N_{tr}, while still satisfying all the constraints; this would lead to an improvement in the objective, which is a contradiction.

We also recognize that if δ is small enough, then $N_{tr} = 0$ becomes a feasible and optimal solution. Without any training data, the firm can use the priors to choose a targeted action in each segment. Specifically, the firm can assign action 1 (treatment) in segment c if $\mu_c \geq 0$, and otherwise assign action 0 (control). It is possible that this targeting policy, defined just using the priors, dominates the benchmark policy in expected profits, and thus satisfies (11) when δ is small. Intuitively, the firm may already have enough information about the likely treatment effect in each segment, that it does not need to collect any additional training data to satisfy the δ-expected improvement requirement. In particular, the solution $N_{tr} = 0$ is feasible (and therefore also optimal) if

$$\sum_{c=1}^{C} w_c \cdot I(\mu_c \geq 0) \cdot \mu_c \geq \delta. \tag{14}$$

We next establish the following monotonicity property. We relegate all proofs to the Appendix.

Proposition 1. $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$ *is monotonically increasing in* N_{tr}.

Because $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$ is monotonic in N_{tr}, equation $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] = \delta$ will have up to one solution.

There is a finite upper bound on the expected profit improvement, and therefore the optimization problem may be infeasible. Specifically, there may be no value of N_{tr} for which the constraint (11) is satisfied. This can occur if the no-treatment action dominates, so that the optimal targeting policy is very similar to the uniform control policy. For example, marketing actions that are costly to implement, but yield little benefit, may be dominated by the no-treatment action. In that case, the expected performance difference between the trained targeting policy and the uniform control policy π_0 may never exceed δ, even for very large sample size N_{tr}. Formally, in the regime of large sample size we have

$$\lim_{N_{tr}\to\infty} \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] = \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c}{\sigma_c}\right) + \sigma_c \phi\left(\frac{\mu_c}{\sigma_c}\right)\right], \tag{15}$$

and therefore the optimization problem is feasible if and only if

$$\sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c}{\sigma_c}\right) + \sigma_c \phi\left(\frac{\mu_c}{\sigma_c}\right)\right] > \delta. \tag{16}$$

This analysis has an important practical implication. The firm can calculate the optimal sample size using a simple procedure. Identifying the (up to unique) root of equation $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] = \delta$, is computationally efficient using standard root-finding algorithms. Bisection will yield the solution efficiently. Given the derivative with respect to N_{tr} is available (see proof of Proposition 1), Newton's method can also be employed and can result in even faster convergence.[10] We formalize this procedure as Algorithm 1, where we use N_{tr}^* to refer to an optimal solution.

Proposition 2 establishes several additional monotonicity properties of the δ-expected improvement certification.

[10]A firm could solve the optimization problem even without resorting to sophisticated algorithms. The expression in (13a) is straightforward to calculate; it is easily implemented in Excel. It uses $\{\mu_c, \sigma_c^2, s_c^2, w_c\}_{c=1,\ldots,C}$. With these inputs, a firm can search on N_{tr}, starting at $N_{tr} = 0$ and increasing N_{tr} until the constraint (11) is satisfied.

Algorithm 1 δ-Expected Improvement Sample Size Calculation

Input: Prior parameters for segment treatment effects $\{\mu_c, \sigma_c^2\}_{c=1,\ldots,C}$; customer outcome variances $\{s_c^2\}_{c=1,\ldots,C}$; relative segment sizes $\{w_c\}_{c=1,\ldots,C}$; required expected improvement δ

if $\sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c}{\sigma_c}\right) + \sigma_c \phi\left(\frac{\mu_c}{\sigma_c}\right)\right] \leq \delta$ **then**
  return Infeasible
else
  if $\sum_{c=1}^{C} w_c \cdot I(\mu_c \geq 0) \cdot \mu_c \geq \delta$ **then**
    $N_{tr}^* \leftarrow 0$
  else

$$N_{tr}^* \leftarrow \left[\begin{array}{c} \text{Solve for } N_{tr} : \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] = \delta, \\ \text{where } k_c := \sqrt{1 + \frac{4 s_c^2}{w_c N_{tr} \sigma_c^2}} \end{array}\right]$$

  return N_{tr}^*

Proposition 2. $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$ *is monotonically decreasing in* s_c*; and increasing in* σ_c *and* μ_c*, for all* $c = 1, \ldots, C$.

The LHS of constraint (11) is decreasing in the customer outcome variances s_c^2, $c = 1, \ldots, C$. When there is higher variance in the observed customer outcomes, more training data is required in order to satisfy the constraints. This is what we would expect; larger variance increases the noise in the training data, increasing the risk of error when training the targeting policy. Consequently, more training data is required to ensure that the expected performance of the trained policy exceeds the expected performance of the benchmark.

In contrast, the LHS of constraint (11) is increasing in the prior variances of the segment treatment effects σ_c^2, $c = 1, \ldots, C$. With larger variance in the segment treatment effects, less training data is required in order to satisfy the constraints. Intuitively, targeting policies exploit variation in the treatment effects m_c. Large positive treatment effects m_c increase the performance difference between the targeting policy and a control condition, while the negative treatment effects are truncated at zero.[11]

Finally, larger treatment effect prior means, μ_c, increase the expected profitability of the trained targeting policy compared to the uniform control policy, $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$, making it easier to satisfy the constraint (11).[12]

[11]The variance of the normal random variable μ_c^{post} in (5) is increasing in σ_c, and the expectation of a rectified normal is increasing in the variance of the underlying normal distribution.

[12]This result would be reversed if we compared the targeting policy to the uniform treatment policy (instead of the uniform control). In that comparison, larger treatment effects μ_c would increase the expected performance of the uniform treatment policy, and require a larger sample size to satisfy the δ-expected improvement requirement.

3.3 Alternative benchmark policies

We have framed the firm's problem as training a targeting policy and comparing its expected performance with a uniform control policy. However, the method is easily extended to compare the trained targeting policy with other benchmarks.

A fixed targeting policy as a benchmark

We can compare the trained targeting policy against any pre-existing fixed benchmark policy, including the uniform treatment policy, or policies in which the recommended action varies across customer segments. Consider a fixed and known targeting policy π' that recommends to each segment c action $\pi'(c)$, where $\pi'(c) \in \{0, 1\}$ for every c. The firm wants to calculate a minimal sample size to train a targeting policy π_t, so that the expected performance improvement of π_t over π' is at least δ:

$$\min_{N_{tr}} \quad N_{tr} \tag{17}$$

$$\text{s.t.} \quad \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}] \geq \delta \tag{18}$$

$$N_{tr} \geq 0. \tag{19}$$

In Appendix C.2.1, we derive a closed-form expression for the optimization constraint (18):

$$\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}] = \begin{cases} \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \pi'(c) \cdot \mu_c + \frac{\sigma_c}{k_c} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \right] & \text{if } N_{tr} > 0 \quad \text{(20a)} \\ \sum_{c=1}^{C} w_c \cdot \left[I\left(\mu_c \geq 0\right) - \pi'(c) \right] \cdot \mu_c & \text{if } N_{tr} = 0. \quad \text{(20b)} \end{cases}$$

This expression is simple to calculate, and is again increasing in N_{tr}. If the problem is feasible, then either $N_{tr} = 0$ is the optimal solution; or the optimal required sample size can be identified using standard root-finding algorithms (or a one-dimensional grid search), to solve for the value of N_{tr} at which constraint (18) is binding.

A learnt optimal uniform policy as a benchmark

An alternative benchmark is to use the optimal uniform policy, where the choice of the uniform treatment or uniform control is informed by the training data. In particular, firms can define the optimal uniform policy π_{u^*} after observing the outcomes from the experiment:

$$\pi_{u^*}(c) = I\left(\sum_c w_c \mu_c^{post} \geq 0\right), \forall c.$$

We can reformulate the firm's decision problem as the optimization problem

$$\begin{aligned}
\min_{N_{tr}} \quad & N_{tr} \\
\text{s.t.} \quad & \mathbb{E}[V^{post}_{\pi_t-\pi_{u^*}}] \geq \delta \\
& N_{tr} \geq 0,
\end{aligned}$$

where $V^{post}_{\pi_t-\pi_{u^*}}$ is similar to (9):

$$\begin{aligned}
V^{post}_{\pi_t-\pi_{u^*}} :=& I\left(\sum_c w_c \mu_c^{post} < 0\right) \cdot \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \mu_c^{post} \\
& - I\left(\sum_c w_c \mu_c^{post} \geq 0\right) \cdot \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} < 0\big) \cdot \mu_c^{post}.
\end{aligned}$$

The first term captures the expected performance difference between the targeting policy π_t and the uniform policy π_0, if π_0 outperforms π_1 in the experiment. The second term captures the case when π_1 outperforms π_0. In Appendix C.2.2, we derive a closed-form solution for the expected $V_{\pi_t-\pi_{u^*}}$:

$$\mathbb{E}[V^{post}_{\pi_t-\pi_{u^*}}] = \begin{cases} \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] - \left[\bar{\mu}\Phi\left(\frac{\bar{\mu}}{\bar{\sigma}}\right) + \bar{\sigma}\phi\left(\frac{\bar{\mu}}{\bar{\sigma}}\right)\right] & \text{if } N_{tr} > 0 \\ \sum_{c=1}^{C} w_c \cdot I\left(\mu_c \geq 0\right) \cdot \mu_c - I\left(\sum_c w_c \mu_c \geq 0\right) \cdot \sum_c w_c \mu_c & \text{if } N_{tr} = 0 \end{cases}$$

where

$$\bar{\mu} = \sum_c w_c \cdot \mu_c, \text{ and } \bar{\sigma}^2 = \sum_c \frac{w_c^2 \cdot \sigma_c^2}{k_c^2}.$$

3.4 Additional extensions

We consider four additional extensions to our formulation for policy certification based on δ-expected performance improvement. We summarize the extensions in the main text, and provide additional details and analyses in Appendix C.2.

We previously recognized that the δ threshold in the δ-expected performance improvement formulation could be interpreted as a cost of migrating from the benchmark policy. If the cost of migrating is δ, then a risk-neutral firm would only migrate if the expected performance improvement from adopting the trained targeting policy is at least δ. As an extension of the δ-expected improvement formulation, we allow the firm to specify a desired probability that the expected performance difference between the trained policy and the benchmark policy exceeds δ. We present this extension in Appendix C.2.3.

Our primary formulation allocates the total experimental sample to different segments proportionally to the segments' relative sizes. This can help to simplify implementation of the experimental design. On the other hand, by optimizing the proportion of data allocated to each segment, the firm can potentially reduce expected variance, and improve the expected performance of the trained policy. In Appendix C.2.5 we describe how to extend the analysis to allow for optimizing the sample size in each segment (rather than just optimizing the overall sample size, N_{tr}). In this analysis we continue to assume that the sample size in a given segment will be divided equally between the treatment and control conditions.

We have assumed that treatment effects are independent across segments. This is a strong assumption. If a firm has access to a library of multiple experiments for the same marketing action, it may be possible to test this assumption by investigating the correlation in the treatment effects across segments. We also investigated the possibility of relaxing this assumption. In particular, we considered using a multivariate normal prior to model the full covariance structure of the segments. Although we obtain a closed-form expression for the expected performance improvement, the expression is complex. We present this analysis in Appendix C.2.6.

Finally, in Appendix C.2.7, we consider an alternative model specification with a log-normal distribution of the monetary outcomes. The log-normal distribution only yields non-negative monetary outcomes, which may be appropriate in some settings. We derive a closed-form expression for the expected performance improvement, so that a one-dimensional grid search provides an efficient solution for finding the required sample size. More generally, our problem formulation could incorporate more flexible distributions for the outcome measure and treatment effect priors, including zero-inflated and mixture models. However, in practice, we will generally require a simulation-based solution for these non-linear models. For each sample size and for each draw of the training data, this requires numerically calculating posterior distributions for the treatment effects. These steps are nested and computationally costly, and may be impractical without additional assumptions, such as conjugate priors for efficient Bayesian updating. The normal and log-normal specifications both allow us to derive the posteriors analytically, and to provide tractable algorithms.

3.5 Summary

We have proposed an approach for calculating the sample size required for training a targeting policy so that it satisfies a predefined performance threshold. The procedure incorporates existing knowledge (as it is Bayesian) and is easily implemented. It is also computationally efficient, even with a large number of customer segments.

Identifying the required sample size for training a targeting policy is a potentially complicated problem. As we discuss in Section 1, the performance of the trained targeting policy improves with more training data, and a sample size calculation has to account for the rate of improvement. This rate of improvement depends upon a complex set of interacting factors, including the choice of estimators and the underlying relationships between the covariates, the treatment effects, and the outcomes.

Our ability to offer a simple algorithm for solving this complex problem relies on a key assumption: we restrict attention to fixed, separable segments that summarize contextual information about the customers. This approximation makes the problem analytically tractable. By recognizing how to approximate the problem so that we can model the expected performance of the targeting policy, we are able to provide a computationally efficient algorithm for estimating the amount of experimental data required to certify that the trained policy meets a performance requirement.

The proposed procedure requires several model inputs, including the firm's prior beliefs about the distribution of the treatment effects. We discuss how to obtain these model inputs and illustrate how to implement the procedure in Section 5. Before this empirical application, in the next section we propose a formulation that certifies the trained targeting policy in a statistical test using a holdout (testing) sample.

4 Certification Using a Statistical Test

In this section we study how much experimental data is required for certifying that a trained targeting policy can outperform a uniform benchmark in a statistical test. Our analysis recognizes that firms can use part of the experimental data to update beliefs about the treatment effects and improve the targeting policy, and the rest of the data to compare the targeting policy to the benchmark in a statistical test. We propose an approach that (1) estimates the size of the experiment required to conduct a statistical test with predefined confidence and power, and (2) recommends how to allocate the data to policy training and evaluation.

4.1 Problem formulation

Recall that we consider a firm that plans to conduct an experiment of size N and use the experimental data to demonstrate that a trained targeting policy can outperform a uniform benchmark using a statistical test. We assume that after conducting the experiment, the firm will randomly split the data into training and testing samples, so that the training sample contains N_{tr} observations, and the testing sample contains N_{test} observations, with $N_{tr} + N_{test} = N$. The firm uses the training sample to update its beliefs about the distribution of the treatments effects m_c and train the targeting policy $\pi_t(c) = I(\mu_c^{post} \geq 0)$ for customer segments $c = 1, ..., C$. The firm then uses the testing sample to compare the out-of-sample performance of the trained targeting policy with the uniform benchmark. Our analysis estimates how much data is required to train and certify a targeting policy with high probability.

We assume that the firm makes the comparison between the targeting policy and the uniform benchmark based on a standard one-sided two-sample t-test.[13] After observing the training sample

[13]We use a one-sided test, because the trained targeting policy cannot perform worse (in expectation) than the benchmark policy.

and updating its prior beliefs, the firm uses the testing sample to construct the test statistic T:

$$T = \frac{\bar{V}_{\pi_t} - \bar{V}_{\pi_0}}{\sqrt{\mathrm{Var}\left(\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right)}} \tag{22}$$

where $\bar{V}_{\pi_t}$ and $\bar{V}_{\pi_0}$ are the average per-customer performances for policies π_t and π_0 defined in Section 2.3. In Appendix D.1, we show that we can rewrite the statistic T as

$$T = \frac{\sum_{c=1}^{C} w_c \cdot I(\mu_c^{post} \geq 0) \cdot \hat{m}_c^{test}}{\sqrt{\sum_{c=1}^{C} w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \mathrm{Var}\left(\hat{m}_c^{test} \mid \hat{m}_c^{tr}\right)}} \tag{23}$$

where

$$\mathrm{Var}\left(\hat{m}_c^{test} \mid \hat{m}_c^{tr}\right) = \left(\sigma_c^{post}\right)^2 + \frac{4s_c^2}{w_c N_{test}},$$

and that conditional on training outcomes $\hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}$, the statistic T follows a normal distribution. Notice that a testing sample is required to conduct the statistical test, so we need $N_{test} > 0$. However, the firm can decide to allocate no data to the training sample and just construct the targeting policy using the priors ($N_{tr} \geq 0$).[14]

The variance expression captures two sources of uncertainty. The first term captures the remaining uncertainty about the treatment effects m_c after observing the training data, and the second term captures the idiosyncratic (finite sample) noise in the testing sample. Equation (23) recognizes that the performance difference between the targeting policy and the uniform policy is exactly zero (with zero variance) in segments where two policies recommend the same action (Simester et al., 2020a). This improves the efficiency of the statistical test.

The firm concludes that the targeting policy significantly outperforms the uniform benchmark with confidence $(1 - \alpha) \cdot 100\%$, if the test statistic T exceeds or equals the critical value $z_{1-\alpha}$: $T \geq z_{1-\alpha}$. In such cases, we say that the targeting policy is $(1 - \alpha)$-*certified*. The realization of the test statistic T depends on the draw of the training and testing data. The firm's goal is to calculate the minimum experimental sample size N, such that the targeting policy is $(1 - \alpha)$-certified with probability at least $(1 - \beta)$.

We consider α and β to be managerial requirements set by the firm, which determine the confidence and power, respectively, of the statistical test used to compare the two policies. These parameters allow the firm to separately define the probability that the statistical test either incorrectly favors the trained policy over the benchmark (a false positive outcome), or incorrectly does not (a false negative outcome). In practice, many managers are familiar with these concepts, and some firms have corporate policies specifying the levels of confidence $(1 - \alpha) \cdot 100\%$ and power $(1 - \beta) \cdot 100\%$ required for marketing experiments. Although academic researchers often use 95% confidence and 80% power thresholds, many firms adopt lower requirements, such as 90% confidence

[14]We only define the test statistic T for those training outcomes for which not all segments have $\mu_c^{post} < 0$. When all segments have $\mu_c^{post} < 0$, we say the policy is not certified.

and 70% power.

We formulate the firm's decision problem as the optimization problem

$$\min_{N_{tr}, N_{test}} \quad N_{tr} + N_{test} \tag{24}$$

$$\text{s.t.} \quad \mathbb{P}\left(T \geq z_{1-\alpha}\right) \geq 1 - \beta \tag{25}$$

$$N_{tr} \geq 0, N_{test} > 0. \tag{26}$$

We can compare the proposed problem formulation to the power analysis often used to design experiments for detecting treatment effects. The traditional power calculation estimates the sample size required to yield a predefined power in a statistical significance test. Similarly, our proposed problem formulation defines a constraint (25) on the probability of certifying the performance advantage of the targeting policy compared to the uniform benchmark.[15] However, there is an important difference between the two approaches. Traditionally, the entire experimental sample recommended by the power analysis is used in a statistical test. In contrast, our proposed formulation leverages part of the experimental data for policy training and updating the prior beliefs about the treatment effects in different segments. This introduces a tradeoff: a larger training sample lowers uncertainty about the treatment effects and yields a targeting policy with higher expected performance, which can reduce the data requirements for an out-of-sample statistical test. Our approach calculates *both* how much experimental data to collect, *and* how much of that data to allocate to policy training.

We note that we formulate the sample size calculation problem in the Bayesian framework. In particular, we specify priors and characterize the probability of a certain event using the posterior distributions. The event of interest is focused on policy certification. It might appear inconsistent that the hypothesis testing approach and the statistic T are fundamentally frequentist. This definition represents one possible out-of-sample certification criterion that is aligned with our industry experience. Future research could investigate alternative certification criteria, which would represent different definitions of events in the Bayesian framework.

We can solve the optimization problem in Equations (24)-(26) for (α, β) policy certification in two ways: using simulations and by analytical approximation. For the simulation-based approach, we can use a two-dimensional search to find a minimum experimental sample size $N = N_{tr} + N_{test}$ for which the certification condition (25) is satisfied. For each (N_{tr}, N_{test}) combination, this approach requires generating random samples and checking whether the certification condition holds, which will often be impractical. In Appendix D.2, we provide two efficiency improvements that speed up the computation. In particular, we leverage statistical properties of the test statistic in order to estimate the probability in the certification constraint by only sampling the training data outcomes μ_c^{post}. We also recognize that the probability in the certification constraint is monotonically

[15] The statistic T is consistent with the test statistic used in traditional power analysis. Compared to power calculations, T incorporates variation in the targeted actions across the segments (indicator functions) and uncertainty about the treatment effects (σ_c^{post}). In the case where $\pi_t(c) = 1$ and $\sigma_c = 0$ for all segments c, the statistical test in our analysis is equivalent to the standard t-test.

increasing in N_{test}, which prunes the grid search.

Even with our proposed improvements, the simulation-based solution for the (α, β)-policy certification problem is computationally demanding. In the empirical application in Section 5, we show that the required computation time can be substantial. We next propose an analytical approximation to the optimization problem (24)-(26) that leads to a computationally efficient solution.

4.2 Approximate analytical solution for (α, β)-policy certification

Our analytical approximation for the optimization problem (24)-(26) involves two steps. We first derive a closed-form approximation of the constraint (25). Evaluating the optimization constraint using closed-form expressions is substantially faster than the simulation-based approach. We further improve the optimization by considering an auxiliary optimization problem. The auxiliary problem allows for an explicit solution for an optimal N_{test} as a function of N_{tr}, which helps to prune the grid search in the primary optimization problem.

We approximate the LHS of the certification constraint (25) as

$$\mathbb{P}\left(T \geq z_{1-\alpha}\right) \approx \Phi\left(\kappa \cdot \left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right)\right), \tag{27}$$

where

$$\mu_\Delta := \begin{cases} \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \quad (28a) \\ \sum_{c=1}^{C} w_c \cdot I\left(\mu_c \geq 0\right) \cdot \mu_c & \text{if } N_{tr} = 0, \quad (28b) \end{cases}$$

$$\mu_U := \begin{cases} \sum_{c=1}^{C} w_c^2 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right) \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) & \text{if } N_{tr} > 0 \quad (29a) \\ \sum_{c=1}^{C} w_c^2 \cdot I\left(\mu_c \geq 0\right) \cdot \left(\sigma_c^2 + \frac{4s_c^2}{w_c N_{test}}\right) & \text{if } N_{tr} = 0. \quad (29b) \end{cases}$$

and where $\kappa := \kappa(N_{tr}, N_{test}) \in (0, 1)$ for all $N_{tr} \geq 0$ and $N_{test} > 0$. We provide a closed-form expression for κ in Appendix D.3.1. Term μ_Δ, defined earlier in Section 3, corresponds to the expected performance difference between the trained targeting policy and the uniform benchmark: $\mu_\Delta = \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$. Term μ_U captures the expectation of the conditional variance:

$$\mu_U = \mathbb{E}\left[\operatorname{Var}\left(\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right)\right].$$

We provide the derivation for the proposed approximation in Appendix D.3.2. For $N_{tr} = 0$, Equations (27)-(29) provide an exact expression for $\mathbb{P}(T \geq z_{1-\alpha})$. For $N_{tr} > 0$, our analysis yields an approximation. The derivation employs the central limit theorem and Taylor series expansions to approximate Equation (88).

We next consider the firm's optimization problem with an approximate constraint:

$$\min_{N_{tr},N_{test}} \quad N_{tr} + N_{test} \tag{30}$$

$$\text{s.t.} \quad \Phi\left(\kappa \cdot \left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right)\right) \geq 1 - \beta \tag{31}$$

$$N_{tr} \geq 0, N_{test} > 0. \tag{32}$$

The derived closed-form expressions for μ_Δ, μ_U and κ allow direct evaluation of the optimization constraint (31), leading to substantial gains in computation time compared to the simulation-based solution. However, the optimization still requires a two-dimensional search over (N_{tr}, N_{test}) to identify a minimal sample size that satisfies the constraints.

We can further improve computational efficiency by considering an auxiliary optimization problem:

$$\min_{N_{tr},N_{test}} \quad N_{tr} + N_{test} \tag{33}$$

$$\text{s.t.} \quad \Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right) \geq 1 - \beta \tag{34}$$

$$N_{tr} \geq 0, N_{test} > 0. \tag{35}$$

We solve the auxiliary optimization problem in Appendix D.3.3. In particular, the LHS of the constraint (34) is continuous and monotonically increasing in N_{test}, and this constraint is feasible for a given N_{tr} if and only if the following condition holds:

$$\begin{cases} \mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot (\sigma_c^{post})^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) > 0 & \text{if } N_{tr} > 0 \quad (36a) \\ \mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot I\left(\mu_c \geq 0\right) \cdot \sigma_c^2 > 0 & \text{if } N_{tr} = 0, \quad (36b) \end{cases}$$

where $Z_{\alpha,\beta} := z_{1-\alpha} + z_{1-\beta}$. When the constraint is feasible, we can write the required testing sample size $N_{test}^{\text{aux}}(N_{tr})$ as an explicit function of the training sample size N_{tr}:

$$N_{test}^{\text{aux}}(N_{tr}) = \begin{cases} \dfrac{4Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c \cdot s_c^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)}{\mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot (\sigma_c^{post})^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)} & \text{if } N_{tr} > 0 \quad (37a) \\ \dfrac{4Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c \cdot s_c^2 \cdot I\left(\mu_c \geq 0\right)}{\mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot I\left(\mu_c \geq 0\right) \cdot \sigma_c^2} & \text{if } N_{tr} = 0 \quad (37b) \end{cases}$$

These results have an important implication for the optimization problem (30)-(32). Recall that

$\kappa \in (0,1)$, so the following inequality holds:

$$\Phi\left(\kappa \cdot \left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right)\right) < \Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right).$$

For a given N_{tr}, we have that $\Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right) < 1-\beta$ for any $N_{test} < N_{test}^{\text{aux}}(N_{tr})$. This implies that, for a given N_{tr}, the primary optimization constraint (31) is also not satisfied for any $N_{test} < N_{test}^{\text{aux}}(N_{tr})$.

We leverage this insight in Algorithm 2 to prune the search space for N_{test} when calculating the sample sizes for (α,β)-policy certification. Similar to the simulation-based solution in Algorithm 5, the proposed algorithm conducts a one-dimensional exhaustive search over the integer values of N_{tr}, and requires, as an input, a maximum experiment size for the enumeration. Firms often have a limited customer base, which may constrain the maximum experiment size. However, Algorithm 2 requires less computation time for each value of N_{tr}, as a result of evaluating the approximate certification constraint (31) with explicit functions, instead of simulation-based estimated probabilities. We illustrate this efficiency in our empirical application in Section 5.

4.3 The tradeoff between the training and the testing sample sizes

A distinctive characteristic of the (α,β)-policy certification formulation is that firms not only choose the size of the experiment, but also allocate the experimental data into training and testing samples. Training data is used to update the targeting policy and the prior beliefs about the treatment effects, and testing data is used to conduct an out-of-sample statistical test. This introduces a tradeoff: a larger training sample lowers uncertainty about the treatment effects and yields a targeting policy with higher expected performance, which can reduce the data requirements for the statistical test. The following proposition highlights the tradeoff between N_{tr} and N_{test} around an optimal solution with $N_{tr}^* > 0$.

Proposition 3. *Consider an optimal solution (N_{tr}^*, N_{test}^*) of the optimization problem (24)-(26) with $N_{tr}^* > 0$.*

(i) The derivatives of the implicit functions $N_{test}(N_{tr})$ and $N_{tr}(N_{test})$ satisfy

$$\left.\frac{dN_{test}}{dN_{tr}}\right|_{(N_{tr}^*, N_{test}^*)} = \left.\frac{dN_{tr}}{dN_{test}}\right|_{(N_{tr}^*, N_{test}^*)} = -1.$$

(ii) The second derivatives of the implicit functions $N_{test}(N_{tr})$ and $N_{tr}(N_{test})$ satisfy

$$\left.\frac{d^2N_{test}}{dN_{tr}^2}\right|_{(N_{tr}^*, N_{test}^*)} \geq 0 \quad \text{and} \quad \left.\frac{d^2N_{tr}}{dN_{test}^2}\right|_{(N_{tr}^*, N_{test}^*)} \geq 0$$

Algorithm 2 (α, β) Sample Size Calculation (Analytical Approximation)

Input: Prior parameters for segment treatment effects $\{\mu_c, \sigma_c^2\}_{c=1,\ldots,C}$; customer outcome variances $\{s_c^2\}_{c=1,\ldots,C}$; relative segment sizes $\{w_c\}_{c=1,\ldots,C}$; required confidence $1-\alpha$; required power $1-\beta$; maximum experiment size $N_{\max}$

```
function SEARCH.NTEST(N_tr, N_test^min, N_test^max)
    N*_test ← N_test^max + 1
    for N_test = N_test^min, ..., N_test^max do
        Calculate Prob using (27)-(29) and (90)
        if Prob ≥ 1 − β then
            N*_test ← N_test
            break
    return N*_test

function MAIN
    N_curr = N_max + 1
    for N_tr = 0, 1, ..., N_max do
        if N_tr ≥ N_curr then
            break
        if (36) is satisfied at N_tr then
            Calculate N_test^aux ← ⌈N_test^aux(N_tr)⌉ using (37)
            if N_tr + N_test^aux < N_curr then
                N_test ← SEARCH.NTEST(N_tr, N_test^aux, N_curr − N_tr)
                if N_tr + N_test < N_curr then
                    (N*_tr, N*_test) ← (N_tr, N_test)
                    N_curr ← N_tr + N_test
    if N_curr ≤ N_max then
        return (N*_tr, N*_test)
    else
        return Infeasible
```

If an optimal solution has $N_{tr}^* > 0$, then the marginal increase in the LHS of the constraint (25) when increasing N_{tr} around the optimal solution is exactly offset by the marginal decrease when decreasing N_{test}. Larger training samples (N_{tr}) result in lower uncertainty about the treatment effects m_c and make higher performance of the trained targeting policy more likely, reducing the amount of data required for the statistical test (N_{test}). Alternatively, larger testing samples increase the statistical power of the policy certification, reducing both the performance of the targeting policy and the precision of the treatment effect estimates that are required to satisfy the probability threshold, and therefore lowering training data requirements.

The second part of the proposition focuses on the rate at which the required size of the testing sample decreases, as the size of the training sample increases, around an optimal solution. The decrease in N_{test} becomes smaller as N_{tr} continues to increase. This means that larger and larger increases in N_{tr} are required to achieve the same reduction in N_{test}. As the training sample size increases towards an optimal solution, the size of the testing sample decreases until optimality, at which point an increase in N_{tr} results in an equivalent offsetting decrease in N_{test}. Increases in N_{tr} beyond this point will result in smaller offsetting decreases in N_{test}, which will not be optimal. Similarly, decreases of N_{tr} from optimality will result in larger offsetting increases in N_{test}, which again will not be optimal.

We caution that this is a local property around an optimal solution with $N_{tr}^* > 0$. In general, the (α, β)-policy certification problem can have a corner solution with $N_{tr}^* = 0$, and the relationship between the size of the training data and the optimal testing sample size $N_{test}^*(N_{tr})$ can be non-monotonic. We provide an example in Appendix D.5.

4.4 Summary

We have proposed a formulation for certifying a targeting policy using an out-of-sample statistical test. The analysis anticipates that firms will conduct an experiment, and then use part of the data to train a targeting policy and the remaining data to compare the performance of the trained policy with a benchmark policy out of sample. We propose algorithms for calculating the required size of the experiment. The algorithms also identify the optimal allocation of the total sample size into training and testing samples. The findings highlight an important tradeoff between the size of these two samples. In the next section, we both illustrate this tradeoff, and demonstrate how to implement the algorithms in an empirical application.

5 Empirical Analysis

We use empirical data to demonstrate application of the proposed algorithms for sample size calculation, illustrate the theoretical results and tradeoffs, and investigate sensitivity of the optimal sample sizes to the input parameters. We start with a data overview and a discussion of how to obtain the model inputs.

5.1 Data overview

Our data is provided by a large luxury fashion retailer. The retailer operates stores in many US states and also has a large online business. Customers have unique identifiers, and we observe purchases by individual customers from both the physical stores and from the e-commerce website. The retailer regularly sends direct mail to its customers. These direct mail campaigns include glossy, multi-page catalogs, highlighting luxury products consistent with the retailer's brand image. The retailer conducts experiments to measure the effectiveness (ATEs) of these catalog campaigns. Rather than simply mailing to all eligible customers, they create a control condition by suppressing mailing to a randomly selected sample of customers.

To illustrate the algorithms proposed in our paper, we assume that the retailer is designing an experiment for a new catalog campaign. The experimental data will be used to train and certify a targeting policy, so that in the future, the catalog is mailed to some customers and not mailed to other customers. To provide model inputs for the algorithms, we use 40,000 customers from a past experiment involving a similar catalog. Half of these customers randomly received the catalog, and the other 20,000 customers did not.[16]

Our data identifies which customers received a catalog ($D_i = 1$), and which customers were in the control group ($D_i = 0$). It also contains a transaction history for each customer. Following the retailer's standard practice, we define the individual monetary outcome Y_i as the total gross profit contributed from purchases by customer i in the 60 days after the catalog's in-home date. We adjust the total gross profit by the catalog mailing cost for the customers in the treatment condition, and we scale the total gross profit for confidentiality.

5.2 Defining inputs

In order to implement the algorithms for sample size calculation we require three key model inputs: a mapping of customers into segments (indexed $c = 1, .., C$), the sample variance s_c^2, and prior parameters μ_c and σ_c^2 for the treatment effect in each segment c.[17]

We follow the retailer's standard segmentation practice that segments customers based on their total annual spending. In particular, the customers are split into the following segments based on total spending in the 12 months prior to the in-home date: "Less than or equal to \$0", ($\$0, \$200$], ..., ($\$2,800, \$3,000$], ($\$3,000, \$3,500$], ..., ($\$7,500, \$8,000$], and "Over \$8,000." The segmentation yields $C = 27$ segments with relative segment sizes w_c ranging from 0.7% to 13.8% (mean = 3.7%, median = 2.4%).[18]

[16] A randomization check (not reported) using pre-treatment spending characteristics confirmed the equivalence of the two groups.

[17] All of our algorithms except Algorithm 1 also require a maximum sample size $N_{\max}$ for enumeration. For illustrative purposes, we use $N_{\max} = 100,000$ customers. In practice, firms can use the size of the entire customer base to specify $N_{\max}$.

[18] There is an inherent tradeoff in specifying the number of segments. More segments lead to more granular targeting models, but yield less precise estimates of the segment treatment effects (for a given N_{tr}). Fewer segments lead to

We estimate s_c by segment as the sample standard deviation of the outcomes Y_i for customers in segment c in the control condition. Empirically, the measure of s_c is insensitive to whether we estimate it using the treatment customers, the control customers, or customers from both conditions.

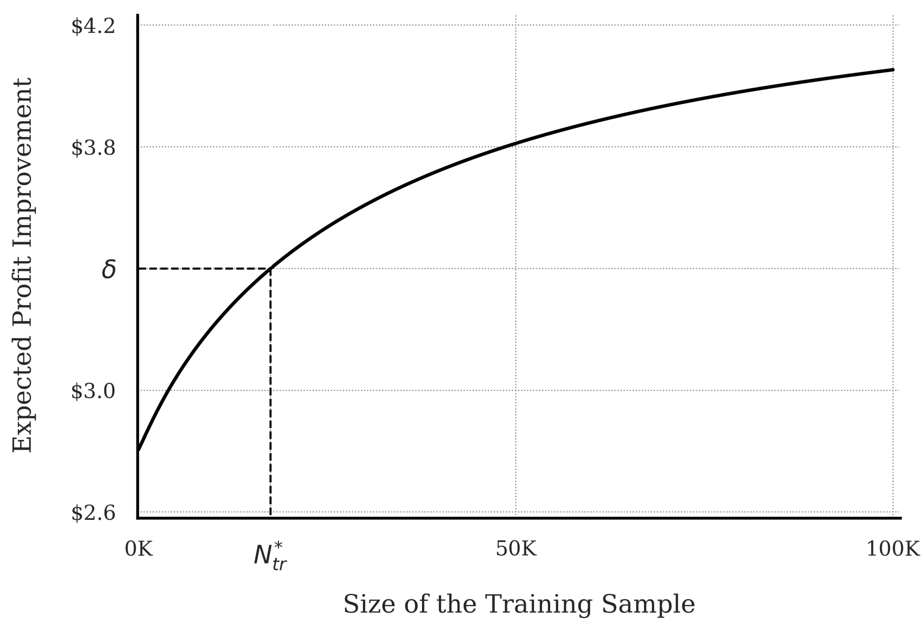
We generate the prior parameters for the expected size of the treatment effect μ_c and the associated uncertainty σ_c^2 for each segment c using a bootstrap approach. In particular, we iteratively draw with replacement 1,000 samples of 40,000 observations (customers) from the field data, and use these samples to calculate a set of 1,000 treatment effect estimates for each segment c. We define μ_c and σ_c^2 as the mean and variance across these estimates (by segment), and provide our empirical model inputs in Appendix E.1.

There are alternative approaches available to obtain these model inputs. For ease of exposition (and to avoid the potential for confusion), we delay discussion of alternative approaches to the end of this empirical section. We next discuss the findings when using these model inputs in the algorithms from Sections 3 and 4.

5.3 δ-expected improvement certification

Figure 1 illustrates the expected profit improvement for different sizes of the training sample, using (13). The x-axis indicates the size of the training data N_{tr} used to train a targeting policy. The y-axis reports the expected profit improvement $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$. The (black) solid line captures the expected profit improvement $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$ as a function of the size of the training data N_{tr} under the model inputs described above.

Figure 1: δ-Expected Improvement Calculation



more precise treatment effect estimates, at the cost of reducing the degrees of freedom for the targeting models.

There are several important observations. First, the solid line does not start at zero on the y-axis. This is consistent with our logic in Section 3.2. Firms can use priors to define a targeting policy, and this policy might already yield a sufficient expected profit improvement over the benchmark policy to satisfy the required performance threshold (δ).

Second, the solid line monotonically increases with the size of the training sample N_{tr}. This is consistent with Proposition 1. Additional training data reduces the variance in the estimates of treatment effects m_c. As a result, the trained targeting policy is more likely to correctly recommend treatment when $m_c \geq 0$, and less likely to recommend treatment when $m_c < 0$. This increases the expected performance of the trained targeting policy compared to the benchmark.

Third, the solid line has a horizontal asymptote, indicating an upper bound on the expected profit improvement. Recall that in Section 3.2 we recognized that the δ-expected improvement optimization problem may not have a feasible solution, and we derived an analytical expression for the least upper bound on the expected performance improvement (15). In our empirical application, there is no finite size of the training sample N_{tr} at which the expected profit improvement reaches \$4.47. As N_{tr} increases, the expected profit improvement approaches this bound asymptotically, but it can never exceed it.

Fourth, Figure 1 illustrates the intuition behind Algorithm 1. The goal is to find the size of the training sample N_{tr} at which the horizontal line $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] = \delta$ intersects with the solid line. Because the solid line is monotonic and continuously differentiable in N_{tr}, we can apply efficient root-finding methods to find the optimal N_{tr}^*.

The inputs to Algorithm 1 include the firm's priors μ_c and σ_c, and the standard deviation of customer outcomes s_c. In Table 1 we investigate how sensitive the sample size calculations are to varying these parameters for $\delta = \$3.40$. In particular, for each parameter, we re-calculate the optimal sample size when increasing or decreasing each parameter separately by 1%, 5% or 10%. We repeat this analysis using two benchmark policies: (a) the uniform control policy, which we have used as the standard benchmark throughout the paper, and (b) the optimal uniform policy learnt using the training data (see Section 3.3).

Our empirical results using the uniform control policy as a benchmark are consistent with Propositions 1 and 2. Recall that for a given sample size N_{tr}, the expected performance improvement over the uniform benchmark decreases with s_c and increases with σ_c (Proposition 2). Because $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$ is monotonically increasing in N_{tr} (Proposition 1), this in turn implies that when s_c decreases or σ_c increases, then smaller sample sizes are required to meet the δ-expected performance improvement threshold.[19] More generally, the required sample size is relatively sensitive to variation in μ_c and σ_c, compared to variation in s_c. Firms that want to ensure they have sufficient sample, may choose to engage in their own sensitivity analysis by varying these empirical inputs.

The comparison to the optimal uniform benchmark highlights that the sample size requirements

[19]Proposition 2 also establishes monotonicity properties with respect to μ_c. However, Table 1 varies μ_c in percentage terms, so μ_c increases or decreases for different segments depending on the sign of μ_c, and Proposition 2 does not directly apply.

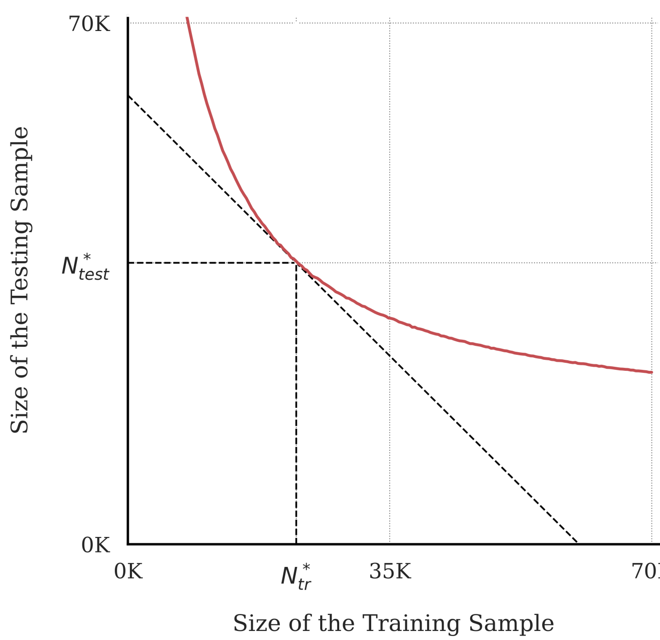
for the optimal uniform policy are larger than for the uniform control. Our sample size calculation when using the optimal uniform benchmark incorporates the probability that the uniform policy improves with training, so larger samples are required to satisfy the performance threshold. In our empirical application, the difference in sample size requirements for optimal uniform and uniform control benchmarks is relatively small. This is because the uniform control policy outperforms the uniform treatment policy on average across the segments.

We next discuss the application of the (α, β) certification algorithms.

5.4 (α, β) certification

Figure 2 illustrates the sample size calculations for (α, β)-certification using $(1-\alpha) = 95\%$ confidence and $(1-\beta) = 70\%$ power. Recall that Algorithms 5 and 2 estimate the optimal size of the experiment N by conducting a one-dimensional search over integers for training data size N_{tr}, estimating the testing data requirements N_{test} for a fixed N_{tr}, and then choosing the optimal N^*_{tr}. The (red) solid line represents the relationship $N^*_{test}(N_{tr})$ estimated using simulations (Algorithm 5). For any size of the training data N_{tr} (x-axis), the figure reports the optimal size of the certification data $N^*_{test}(N_{tr})$ (y-axis).

Figure 2: (α, β)-Certification Calculation



The figure illustrates a tradeoff; around the optimal solution with $N^*_{tr} > 0$, the more training data we obtain, the less testing data we need. Recall from our discussion in Section 4.3, more training data reduces the uncertainty about the treatment effects and makes higher performance of the trained targeting policy more likely, reducing the amount of testing data needed to meet the

Table 1: Sample size requirements for δ-expected improvement and sensitivity to variation in the input parameters, for $\delta = \$3.40$ and two benchmark policies: the uniform control and the optimal uniform.

Input Parameters	Uniform Control		Optimal Uniform	
	N^*_{tr}	% Diff. from Default	N^*_{tr}	% Diff. from Default
Default (Section 5.2)	17,500	—	20,308	—
Sensitivity to μ_c				
−10%	30,463	74.07%	41,380	103.76%
−5%	23,531	34.46%	29,410	44.82%
−1%	18,638	6.50%	21,939	8.03%
+1%	16,394	−6.32%	18,762	−7.61%
+5%	12,284	−29.81%	13,349	−34.27%
+10%	7,819	−55.32%	8,066	−60.28%
Sensitivity to σ_c				
−10%	29,753	70.02%	35,180	73.23%
−5%	22,561	28.92%	26,394	29.97%
−1%	18,382	5.04%	21,363	5.19%
+1%	16,672	−4.73%	19,320	−4.87%
+5%	13,829	−20.98%	15,947	−21.47%
+10%	11,101	−36.57%	12,737	−37.28%
Sensitivity to s_c				
−10%	14,175	−19.00%	16,449	−19.00%
−5%	15,793	−9.75%	18,328	−9.75%
−1%	17,151	−1.99%	19,904	−1.99%
+1%	17,851	2.01%	20,716	2.01%
+5%	19,293	10.25%	22,389	10.25%
+10%	21,174	20.99%	24,572	21.00%

(α, β)-policy certification requirement. Around the optimal solution, a marginal increase in the size of the training sample results in an identical offsetting decrease in the required size of the testing sample. In particular, we see in Figure 2 that at the optimal solution (N^*_{tr}, N^*_{test}), which minimizes the total required sample size $N = N_{tr} + N_{test}$, the (negative) 45°-line is tangential to the solid line.

We also illustrate the tradeoff between the size of the training and testing samples in Figure 3. In this figure we report the power of the statistical test as we vary the allocation of a fixed sample of data ($N = 100,000$) to the training and testing samples. The x-axis is the share of the 100,000 observations allocated to training, with the remaining observations allocated to testing. The y-axis reports the power of the test ($1-\beta$), measuring the probability the test will correctly reject the null hypothesis that there is no difference in the performance of the trained targeting and benchmark policies.

There are several findings of interest. First, at the extreme left of the x-axis, all of the data is allocated to testing. Even at this extreme, where the targeting policy is only designed using the priors, the test has some statistical power. This is because the priors contain information, and so the expected performance of the targeting policy exceeds the expected performance of the benchmark policy. As the allocation of data to training increases, the power of the test initially increases. The availability of data for training improves the expected performance of the targeting model. Although there is less data available for testing, the improvement in performance outweighs this second effect, and so the power of the test increases. However, eventually, the improvement in performance no longer outweighs the reduction in power due to decreasing the size of the testing sample. At this point, the power of the test starts to diminish. Eventually, when all of the data is allocated to training, and there is no data available for testing, the statistical test has no power.

Figure 3: Power of Statistical Test When Varying the Allocation of Data to Training and Testing

In Table 2 we compare the minimum required sample size using the simulation-based and analytical approximation-based solutions. The optimal size of the training sample (N^*_{tr}) under the simulation-based solution is 22,489. The optimal size of the testing sample (N^*_{test}) is 37,797, for a total sample size of 60,286. In comparison, the analytical approximation proposed in Section 4.2 yields similar sample size estimates: 21,934 for training, 38,965 for testing, and 60,899 for total. In Appendix E.2, we replicate Figures 2 and 3 using the simulation and approximation-based approaches. Our approximation closely resembles the simulation-based results, while using the derived closed-form expressions greatly improves computational efficiency: Table 2 reports a dramatic reduction in runtime.[20]

Table 2: Comparison of sample size requirements for (α, β)-certification.

Problem Formulation	N_{tr}^*	N_{test}^*	Total	Runtime
(α, β)-certification, $(1-\alpha) = 95\%$, $(1-\beta) = 70\%$				
Simulation-based solution	22,489	37,797	60,286	1h 59min
Analytical approximation-based solution	21,934	38,965	60,899	13.39s
(α, β)-certification, no training				
Analytical solution (exact)	–	468,048	468,048	0.09s

We can also estimate the sample size requirements if we just defined a targeting policy using the priors, without any training data. When setting $N_{tr} = 0$, (α, β)-certification requires a testing sample with 468,048 customers. Without training data, the expected performance of the targeting policy worsens; the expected profit improvement of the targeting policy with $N_{tr}^* = 22,489$ is \$3.49, while the expected profit improvement without any training data is \$2.81 (see Figure 1). Without training, the required testing data increases, because we need more precise outcome measures from the testing data to establish a significant performance difference over the benchmark policy.

In our empirical setting, the retailer conducted over 180 direct mail experiments. 88% of these campaigns were distributed to over 60,286 customers, but only 4% of the campaigns included over 468,048 customers. Without policy training, (α, β)-certification would require a larger experiment than the retailer typically implements. However, with policy training, the required sample sizes are well within the firm's standard practice. This further emphasizes the importance of the training-testing tradeoff, and highlights the difference between a standard power calculation (which does not include a training component), and the problem that we solve.

In Table 3, we evaluate the sensitivity of the required total sample size for the (α, β)-policy certification. The findings reveal that the sample size for this formulation is much less sensitive to variation in μ_c and σ_c compared to the sample size for the δ-expected improvement certification (Table 1). Firms that use this formulation will generally have less need to engage in their own sensitivity analysis when choosing sample sizes.

In the same table we also report several additional sensitivity results. First, we compare the optimal allocation of the required sample size between training and testing. We see that this

[20] The runtime was measured when implementing each algorithm on a standard 2022 MacBook Air (M2, 8-Core) with 24GB RAM, using Python version 3.9.12.

proportion is relatively stable to varying the input parameters (particularly s_c). Second, we see that the analytical approximation-based solution for the total sample size is within 1.37% of the simulation solution for the different values of the input parameters that we test. We conclude that the accuracy of the analytical approximation is robust to varying these empirical inputs.

The results in Table 3 also allow comparison of the required sample size with and without training data. Specifically, in the last column of the table (N^*_{test}/N^*_{sim}), we report the sample size required without allocating any data to training (N^*_{test}), as a multiple of the optimal simulation-based solution (N^*_{sim}). The finding that training data can sharply reduce the total required sample size is also robust to varying the priors.

Finally, the findings in Table 3 ("—" entries in the "No Training Data" panel) also illustrate the feasibility constraints of the certification problem. Recall that for a *given* value of N_{tr}, there may not be a feasible solution. This is because, while increasing the testing sample N_{test} reduces finite sample noise, the uncertainty about the treatment effects $(\sigma_c^{post})^2$ does not depend on N_{test}. In the "No Training Data" we hold N_{tr} fixed at zero.[21]

5.5 Alternative approaches for obtaining the model inputs

The sample variance s_c^2 is perhaps the easiest of the model inputs to obtain. Our investigations reveal that the variance is very similar when calculated using different approaches, including: the 20,000 treatment customers, the 20,000 control customers, or all 40,000 customers. Further investigation confirms that we can also estimate s_c using historical purchases from other randomly selected customer samples, including customers who did not participate in the prior experiments.

Customer segments can be obtained from multiple potential sources. For example, Simester, Timoshenko, and Zoumpoulis (2020b) compare different approaches to training targeting models for a firm engaged in prospecting for new customers. When prospecting for new customers, firms do not have historical data to segment customers. Instead, it is standard practice to use geographical segmentation. In the study by Simester et al. (2020b), the firm relied on zip codes and postal carrier routes exogenously defined by the US Postal Service. In business-to-business settings, it is common to segment customers using standard industry classification codes, such as NAICS, or segment business customers by their size. Past targeting policies may also suggest customer segments, particularly where the policies are trained using estimators that rely on tree structures. Our approach assumes that customer segmentation summarizes information about responsiveness to marketing actions, so past experiments and targeting policies can also provide valuable information for segmentation. Alternatively, firms may be able to use supervised or unsupervised clustering methods to group customers into segments (Wager and Athey, 2018).

A requirement for priors is a natural characteristic of the problem that we are trying to solve. Firms want to make decisions about sample sizes before running the "main" experiment, and this requires inputs to model the policy learning rate. The challenge of obtaining the segment-specific

[21] If we allow N_{tr} to vary, there is always a feasible solution.

Table 3: Sample size requirements for (α, β)-certification with the simulation-based solution; the analytical approximation; without policy training; and sensitivity to variation in the input parameters, for $1-\alpha = 95\%$, $1-\beta = 70\%$.

Input Parameters	Simulation Solution			Analytical Approximation			No Training Data	
	N^*_{sim}	% Train	%Diff. N^*_{sim} from Default	N^*_{aprx}	% Train	%Diff. N^*_{aprx} from N^*_{sim}	N^*_{test}	N^*_{test}/N^*_{sim}
Default (Section 5.2)	60,286	37.30%	—	60,899	36.02%	1.02%	468,048	7.8×
Sensitivity to μ_c								
−10%	69,672	40.65%	15.57%	70,264	39.32%	0.85%	—	—
−5%	64,913	39.34%	7.68%	65,507	37.77%	0.92%	—	—
−1%	61,183	37.81%	1.49%	61,808	36.41%	1.02%	640,647	10.5×
+1%	59,373	36.73%	−1.51%	59,997	35.66%	1.05%	367,927	6.2×
+5%	55,786	34.16%	−7.46%	56,457	34.06%	1.20%	196,026	3.5×
+10%	51,487	32.13%	−14.60%	51.193	31.79%	1.37%	121,791	2.4×
Sensitivity to σ_c								
−10%	62,488	32.42%	3.65%	63,297	31.23%	1.29%	138,382	2.2×
−5%	61,575	34.49%	2.14%	62,302	33.86%	1.18%	210,596	3.4×
−1%	60,527	35.81%	0.40%	61,205	35.45%	1.12%	374,584	6.2×
+1%	59,982	37.54%	−0.50%	60,581	36.37%	1.00%	625,751	10.4×
+5%	58,661	38.79%	−2.70%	59,215	37.75%	0.94%	—	—
+10%	56,866	40.19%	−5.67%	57,346	38.98%	0.84%	—	—
Sensitivity to s_c								
−10%	48,787	36.02%	−19.07%	49,328	35.98%	1.11%	379,119	7.8×
−5%	54,401	37.00%	−9.76%	54,961	35.95%	1.03%	422,414	7.8×
−1%	59,076	37.35%	−2.01%	59,687	36.00%	1.03%	458,734	7.8×
+1%	61,467	37.34%	1.96%	62,123	36.01%	1.07%	477,456	7.8×
+5%	66,463	37.30%	10.25%	67,141	36.00%	1.02%	516,023	7.8×
+10%	72,937	37.50%	20.98%	73,687	35.88%	1.03%	566,338	7.8×

prior parameters μ_c and σ_c^2 is essentially no different than the challenge of obtaining inputs for a standard power calculation. Conceptually, the prior parameters capture the information about the problem that the firm already has. In practice, this information could come from many different sources. In our empirical application, we demonstrate how to obtain these priors from a past experiment. Recent research suggests that firms may be able to obtain these priors by combining multiple past campaigns on related but different marketing actions (Timoshenko et al., 2020). For example, the retailer that provided the data for this study has conducted several hundred direct mail campaigns in recent years, and in almost all of these campaigns it randomly suppressed mailing to 10% of the customers as a control group. Together, these past campaigns may help inform the distributional assumptions about the treatment effects for each customer segment. The past campaigns can potentially also reveal the correlations between the treatment effects across the segments.[22] However, a limitation is that different marketing campaigns often involve different marketing actions, or use different eligibility criteria to select customers. If so, the experiments may not be measuring the same treatment effects. Other sources for the priors could include managerial intuition and correlations between historical firm actions and the outcome measure. Finally, we point to the research on prior elicitation in Bayesian statistics that proposes a wide range of techniques for formulating existing expert knowledge as a prior probability distribution (O'Hagan et al., 2006).

6 Conclusions

Firms often conduct experiments to train and certify targeting policies. Determining the size of these experiments is a challenging and important problem. The required sample size depends upon how quickly the expected performance of the trained targeting policy improves as the size of the training sample increases. We propose an approach for sample size calculations that relies on simplifying assumptions about customer segmentation. We assume that customers are grouped into segments that fully summarize contextual information about their responsiveness to marketing actions, and that the treatment effects are independent across segments.

We consider two settings. In the first setting, the firm is primarily focused on using the experimental data to train a targeting policy that satisfies a performance threshold. In the second setting, the firm requires certification of the policy in an out-of-sample statistical test. Firms can customize both types of requirements by adjusting the performance threshold and the parameters of the statistical test. Our proposed algorithms are easily applied and efficiently scale to a large number of customer segments. We illustrate the algorithms using data from a field experiment conducted by a retailer.

6.1 Discussion and future research

There are several opportunities to extend our findings in future research.

First, we assume that customer segmentation summarizes all relevant information about the individual treatment effects. Future research could explore how to extend our findings to incorporate contextual information about customers in a more flexible way. For example, we could assume a model for mean monetary outcomes, as a function of customer covariates, treatment, and the unknown segment treatment effect, as is done for example by Alban et al. (2024) for a different context. The firm would still use Bayesian updating to form a posterior on the segment effects based on experimental customer outcomes.

[22]The bootstrap approach presented in Section 5, allows the priors in neighboring segments to be related via similarity in μ_c's and σ_c's. However, it cannot estimate how the treatment effects co-vary across the customer segments. More specifically, our bootstrap approach cannot estimate the off-diagonal elements of the covariance matrix in the extension to correlated treatment effects discussed in Section 3.4. The bootstrap draws do not have a natural ordering, and so the empirical correlation between outcomes in different segments is not meaningful.

Future research could also propose methods for sample size calculations in large action spaces. Our study is motivated by targeting decisions at an offline retailer. The retailer designs marketing campaigns such as catalogs and brochures, and then optimizes which customers should receive the campaigns. In other contexts, firms consider multiple versions of their marketing communications, or optimize continuous variables (such as prices). An extension to large action spaces would be valuable.

Finally, in our setting, the firm conducts a single experiment, and must choose the sample size of the experiment in advance. When the firm requires out-of-sample certification, data from the experiment is randomly assigned to training and testing samples. However, there are settings in which the process is sequential. The firm first conducts an experiment to obtain training data. Once it observes the outcome of this initial experiment, it trains a targeting policy, and then implements a second experiment to further train the policy, or certify the performance of the trained policy. Future research could extend our analysis to multiple sequential experiments.

References

Alban, Andres, Stephen E. Chick, and Spyros I. Zoumpoulis (2024), "Learning personalized treatment strategies with predictive and prognostic covariates in adaptive clinical trials." SSRN, https://ssrn.com/abstract=4160045.

Anderson, Stephen J, Rajesh Chandy, and Bilal Zia (2018), "Pathways to profits: The impact of marketing vs. finance skills on business performance." *Management Science*, 64, 5559–5583.

Beauchamp, Maxime (2018), "On numerical computation for the distribution of the convolution of N independent rectified Gaussian variables." *Journal de la Société Française de Statistique*, 159, 88–111, URL `https://hal-ineris.archives-ouvertes.fr/ineris-03319137`.

Blattberg, Robert C. (1979), "The design of advertising experiments using statistical decision theory." *Journal of Marketing Research*, 16, 191–202, URL `http://www.jstor.org/stable/3150683`.

Daljord, Øystein, Carl F. Mela, Jason M. T. Roos, Jim Sprigg, and Song Yao (2023), "The design and targeting of compliance promotions." *Marketing Science*, 0, null, URL `https://doi.org/10.1287/mksc.2022.1420`.

Dubé, Jean-Pierre and Sanjog Misra (2019), "Personalized pricing and customer welfare." *Available at SSRN 2992257*.

Feit, Elea McDonnell and Ron Berman (2019), "Test & roll: Profit-maximizing A/B tests." *Marketing Science*, 38, 1038–1058.

Gabel, Sebastian and Artem Timoshenko (2022), "Product choice with large assortments: A scalable deep-learning model." *Management Science*, 68, 1808–1827.

Ginter, James L., Martha C. Cooper, Carl Obermiller, and Thomas J. Page (1981), "The design of advertising experiments using statistical decision theory: An extension." *Journal of Marketing Research*, 18, 120–123, URL `http://www.jstor.org/stable/3151323`.

Hauser, John R, Glen L Urban, Guilherme Liberali, and Michael Braun (2009), "Website morphing." *Marketing Science*, 28, 202–223.

Hitsch, Günter J and Sanjog Misra (2018), "Heterogeneous treatment effects and optimal targeting policy evaluation." *Available at SSRN 3111957*.

Joo, Joonhwi and Khai Chiong (2023), "Getting the most out of online a/b tests using the minimax-regret criteria." *B Tests Using the Minimax-Regret Criteria (November 14, 2023)*.

Lewis, Randall A and Justin M Rao (2015), "The unfavorable economics of measuring the returns to advertising." *The Quarterly Journal of Economics*, 130, 1941–1973.

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

Marshall, Albert W and Ingram Olkin (2007), *Life distributions*, volume 13. Springer.

Maxwell, Scott E., Ken Kelley, and Joseph R. Rausch (2008), "Sample size planning for statistical power and accuracy in parameter estimation." *Annual Review of Psychology*, 59, 537—563.

O'Hagan, Anthony, Caitlin E Buck, Alireza Daneshkhah, J Richard Eiser, Paul H Garthwaite, David J Jenkinson, Jeremy E Oakley, and Tim Rakow (2006), *Uncertain Judgements: Eliciting Experts' Probabilities.* John Wiley & Sons.

Owen, D. B. (1980), "A table of normal integrals." *Communications in Statistics - Simulation and Computation*, 9, 389–419, URL `https://doi.org/10.1080/03610918008812164`.

Patriksson, Michael (2008), "A survey on the continuous nonlinear resource allocation problem." *European Journal of Operational Research*, 185, 1–46.

Rafieian, Omid and Hema Yoganarasimhan (2021), "Targeting and privacy in mobile advertising." *Marketing Science*, 40, 193–218.

Schwartz, Eric M, Eric T Bradlow, and Peter S Fader (2017), "Customer acquisition via display advertising using multi-armed bandit experiments." *Marketing Science*, 36, 500–522.

Simester, Duncan, Artem Timoshenko, and Spyros I. Zoumpoulis (2020a), "Efficiently evaluating targeting policies: Improving on champion vs. challenger experiments." *Management Science*, 66, 3412–3424, URL `https://doi.org/10.1287/mnsc.2019.3379`.

Simester, Duncan, Artem Timoshenko, and Spyros I. Zoumpoulis (2020b), "Targeting prospective customers: Robustness of machine-learning methods to typical data challenges." *Management Science*, 66, 2495–2522, URL `https://doi.org/10.1287/mnsc.2019.3308`.

Sutton, Richard S and Andrew G Barto (2018), *Reinforcement learning: An introduction.* MIT press.

Tetenov, Aleksey (2012), "Statistical treatment choice based on asymmetric minimax regret criteria." *Journal of Econometrics*, 166, 157–165.

Timoshenko, Artem, Marat Ibragimov, Duncan Simester, Jonathan Parker, and Antoinette Schoar (2020), "Transferring information between marketing campaigns to improve targeting policies." *Working Paper.*

Wager, Stefan and Susan Athey (2018), "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association*, 113, 1228–1242.

Zhang, Jie and Lakshman Krishnamurthi (2004), "Customizing promotions in online stores." *Marketing science*, 23, 561–578.

Online Appendix to "A Sample Size Calculation for Training and Certifying Targeting Policies"

by Duncan Simester, Artem Timoshenko, and Spyros I. Zoumpoulis

A Table of Notations

Notation	Definition
N_{tr}	Size of training sample.
N_{test}	Size of testing sample.
C	Number of segments.
c	Segment identifier.
i	Customer identifier.
m_c	Treatment effect in segment c.
μ_c	Prior mean of m_c.
σ_c^2	Prior variance of m_c.
Y_i	Outcome for customer i in the experiment.
D_i	Action assignment for customer i in the experiment.
ϵ_i	Individual customer shock in Y_i.
s_c^2	Variance of ϵ_i in segment c.
$\bar{Y}_{c\|1}^{tr}, \bar{Y}_{c\|0}^{tr}$	Average observed outcomes in the treatment and control conditions in the training sample.
$\bar{Y}_{c\|1}^{test}, \bar{Y}_{c\|0}^{test}$	Average observed outcomes in the treatment and control conditions in the testing sample.
$\hat{m}_c^{tr}$	Average treatment effect in segment c estimated using the training sample.
$\hat{m}_c^{test}$	Average treatment effect in segment c estimated using the testing sample.
μ_c^{post}	Posterior mean of m_c after obtaining the training data.
$(\sigma_c^{post})^2$	Posterior variance of m_c after obtaining the training data.
π_t	The trained targeting policy.
π_0	The uniform control policy.
$\bar{V}_\pi$	Average per-customer performance in the testing sample for policy π.
$V_{\pi_t-\pi_0}^{post}$	Expected performance difference between π_t and π_0 conditioned on the training data.
δ	Performance threshold for the δ-expected improvement.
γ	Probability threshold for $(1-\gamma)$-probable δ-expected improvement.
α, β	Thresholds for policy certification with $1-\alpha$ confidence, $1-\beta$ power.
κ	Adjustment term in the analytical approximation for (α, β)-certification; $\kappa := \kappa(N_{tr}, N_{test})$
$\mathcal{N}\left(\mu, \sigma^2\right)$	Normal distribution with mean μ and variance σ^2.
$\phi()$	PDF of the standard normal distribution.
$\Phi()$	CDF of the standard normal distribution.
$z_{1-\alpha}, z_{1-\beta}, z_{1-\gamma}$	z-values such that $\Phi(z_{1-\alpha}) = 1-\alpha$, $\Phi(z_{1-\beta}) = 1-\beta$, $\Phi(z_{1-\gamma}) = 1-\gamma$

To simplify expressions, we also write

$$k_c = \sqrt{1 + \frac{4s_c^2}{w_c N_{tr} \sigma_c^2}}$$

and

$$Z_{\alpha,\beta} = z_{1-\alpha} + z_{1-\beta}.$$

B Details for δ-Expected Improvement Certification

B.1 Derivations for $\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}\right]$

We focus on the conditional expectation of the profit improvement between the trained targeting policy and the benchmark uniform policy, conditioned on the outcomes of the training stage, $V_{\pi_t-\pi_0}^{post} = \mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right]$. Conditioned on random variables $\hat{m}_c^{tr}, c = 1, \ldots, C$, the conditional expectation $V_{\pi_t-\pi_0}^{post}$ is a random variable itself. We next characterize its distribution. Recall Equation (9):

$$V_{\pi_t-\pi_0}^{post} = \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \mu_c^{post}.$$

In the absence of any training data ($N_{tr} = 0$), we have that $\mu_c^{post} = \mu_c$ for every segment c. The conditional expectation then collapses to a constant:

$$\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right] = \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] = \sum_{c=1}^{C} w_c I\left(\mu_c \geq 0\right) \mu_c.$$

We next consider $N_{tr} > 0$. The distribution of $\hat{m}_c^{tr}$ is

$$\hat{m}_c^{tr} \sim \mathcal{N}\left(\mu_c, \sigma_c^2 + \frac{4s_c^2}{w_c N_{tr}}\right). \tag{38}$$

From (3), it follows that the random variable μ_c^{post} is distributed as

$$\mu_c^{post} \sim \mathcal{N}\left(\mu_c, \frac{\sigma_c^4}{\sigma_c^2 + \frac{4s_c^2}{w_c N_{tr}}}\right) = \mathcal{N}\left(\mu_c, \frac{\sigma_c^2}{k_c^2}\right), \tag{39}$$

where

$$k_c := \sqrt{1 + \frac{4s_c^2}{w_c N_{tr} \sigma_c^2}}.$$

We next look into the distribution of random variable $I\left(\mu_c^{post} \geq 0\right) \mu_c^{post} = \max(0, \mu_c^{post})$. For notational convenience, we set $X_c := \mu_c^{post}$, and denote $\mu_c^X := \mathbb{E}[\mu_c^{post}] = \mu_c$, $\sigma_c^X := \sqrt{\mathrm{Var}(\mu_c^{post})} =$

$\frac{\sigma_c}{k_c}$. We also denote $X_c^+ := \max(0, X_c) = I\left(\mu_c^{post} \geq 0\right)\mu_c^{post}$. The random variable X_c^+ has a rectified Gaussian distribution, with mean and variance (Beauchamp, 2018)

$$\mathbb{E}\left[X_c^+\right] = \mu_c^X \Phi\left(\frac{\mu_c^X}{\sigma_c^X}\right) + \sigma_c^X \phi\left(\frac{\mu_c^X}{\sigma_c^X}\right) = \mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \tag{40}$$

$$\begin{aligned}\mathrm{Var}\left(X_c^+\right) &= \left((\mu_c^X)^2 + (\sigma_c^X)^2\right)\Phi\left(\frac{\mu_c^X}{\sigma_c^X}\right) + \mu_c^X\sigma_c^X\phi\left(\frac{\mu_c^X}{\sigma_c^X}\right) - \left(\mathbb{E}\left[X_c^+\right]\right)^2 \\ &= (\mu_c^X)^2\Phi\left(\frac{\mu_c^X}{\sigma_c^X}\right)\left(1 - \Phi\left(\frac{\mu_c^X}{\sigma_c^X}\right)\right) + (\sigma_c^X)^2\left(\Phi\left(\frac{\mu_c^X}{\sigma_c^X}\right) - \phi^2\left(\frac{\mu_c^X}{\sigma_c^X}\right)\right) + \mu_c^X\sigma_c^X\phi\left(\frac{\mu_c^X}{\sigma_c^X}\right)\left(1 - 2\Phi\left(\frac{\mu_c^X}{\sigma_c^X}\right)\right) \\ &= \mu_c^2\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) + \frac{\sigma_c^2}{k_c^2}\left(\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \phi^2\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) + \frac{\mu_c\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - 2\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right).\end{aligned} \tag{41}$$

By the linearity of expectation, it follows that

$$\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] = \mathbb{E}[V_{\pi_t - \pi_0}^{post}] = \begin{cases} \displaystyle\sum_{c=1}^{C} w_c\left[\mu_c\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \\ \displaystyle\sum_{c=1}^{C} w_c I\left(\mu_c \geq 0\right)\mu_c & \text{if } N_{tr} = 0. \end{cases}$$

We note that the expression for $N_{tr} = 0$ coincides with the limit of the expression for $N_{tr} > 0$ as $N_{tr} \to 0$.

We also look into the distribution of random variable $-I\left(\mu_c^{post} < 0\right)\mu_c^{post}$, as it will be useful for the derivations in Appendix C.2.1. We define random variable $\mu_c^{post'} := -\mu_c^{post}$. Note that the random variable $\mu_c^{post'}$ is distributed as

$$\mu_c^{post'} \sim \mathcal{N}\left(-\mu_c, \frac{\sigma_c^4}{\sigma_c^2 + \frac{4s_c^2}{w_c N_{tr}}}\right) = \mathcal{N}\left(-\mu_c, \frac{\sigma_c^2}{k_c^2}\right). \tag{43}$$

We have that $-I\left(\mu_c^{post} < 0\right)\mu_c^{post} = I\left(\mu_c^{post'} > 0\right)\mu_c^{post'}$. We set $X_c' = \mu_c^{post'}$, and denote $\mu_c^{X'} := \mathbb{E}[\mu_c^{post'}] = -\mu_c^X = -\mu_c$. We also denote $X_c^{'+} := \max(0, X_c') = I\left(\mu_c^{post'} > 0\right)\mu_c^{post'} = -I\left(\mu_c^{post} < 0\right)\mu_c^{post}$. The random variable $X_c^{'+}$ has a rectified Gaussian distribution, with mean (Beauchamp, 2018)

$$\begin{aligned}\mathbb{E}\left[X_c^{'+}\right] &= \mu_c^{X'}\Phi\left(\frac{\mu_c^{X'}}{\sigma_c^X}\right) + \sigma_c^X\phi\left(\frac{\mu_c^{X'}}{\sigma_c^X}\right) = -\mu_c^X\left(1 - \Phi\left(\frac{\mu_c^X}{\sigma_c^X}\right)\right) + \sigma_c^X\phi\left(\frac{\mu_c^X}{\sigma_c^X}\right) \\ &= -\mu_c\left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right).\end{aligned} \tag{44}$$

B.2 Proof of Proposition 1

Proof. We focus on segment c. Recall that

$$k_c = \sqrt{1 + \frac{4s_c^2}{w_c N_{tr}\sigma_c^2}}.$$

Note that $k > 1$.

We define

$$k_{c,N_{tr}} := \frac{\partial k_c}{\partial N_{tr}} = -\frac{2s_c^2}{w_c N_{tr}^2 \sigma_c^2 k_c} < 0. \tag{45}$$

We can write the derivative with respect to N_{tr} as

$$\begin{aligned}
\frac{\partial \left[\frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right]}{\partial N_{tr}} &= -\frac{\sigma_c}{k_c^2} k_{c,N_{tr}} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \frac{\sigma_c}{k_c}\frac{\mu_c k_c}{\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{\mu_c}{\sigma_c} k_{c,N_{tr}} + \mu_c \phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{\mu_c}{\sigma_c} k_{c,N_{tr}} \qquad (46)\\
&= -k_{c,N_{tr}} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(\frac{\sigma_c}{k_c^2} + \frac{\mu_c^2}{\sigma_c} - \frac{\mu_c^2}{\sigma_c}\right)\\
&= -k_{c,N_{tr}} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{\sigma_c}{k_c^2}\\
&> 0
\end{aligned}$$

where in the second term of (46) we have used that $\frac{d\phi(x)}{dx} = -x\phi(x)$, and at the last step we have used that $k_{c,N_{tr}} < 0$. It follows that $\frac{\partial \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]}{\partial N_{tr}} > 0$. □

B.3 Proof of Proposition 2

Proof. We focus on segment c and recall $k_c = \sqrt{1 + \frac{4s_c^2}{w_c N_{tr} \sigma_c^2}}$. Letting

$$k_{c,s} := \frac{\partial k_c}{\partial s_c} = \frac{4s_c}{w_c N_{tr} \sigma_c^2 k_c} > 0,$$

we have

$$\begin{aligned}
\frac{\partial \left[\frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right]}{\partial s_c} &= -\frac{\sigma_c}{k_c^2} k_{c,s} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \frac{\sigma_c}{k_c}\frac{\mu_c k_c}{\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{\mu_c}{\sigma_c} k_{c,s} + \mu_c \phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{\mu_c}{\sigma_c} k_{c,s} \qquad (47)\\
&= -k_{c,s} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(\frac{\sigma_c}{k_c^2} + \frac{\mu_c^2}{\sigma_c} - \frac{\mu_c^2}{\sigma_c}\right)\\
&= -k_{c,s} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{\sigma_c}{k_c^2}\\
&< 0.
\end{aligned}$$

We define

$$k_{c,\sigma_c} := \frac{\partial k_c}{\partial \sigma_c} = -\frac{4s_c^2}{w_c N_{tr} \sigma_c^3 k_c} < 0.$$

We have that

$$
\begin{aligned}
\frac{\partial\left[\frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)+\mu_c\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right]}{\partial\sigma_c} &= \frac{1}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)-\frac{\sigma_c}{k_c^2}k_{c,\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)-\frac{\sigma_c}{k_c}\frac{\mu_c k_c}{\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\mu_c\frac{k_{c,\sigma_c}\sigma_c-k_c}{\sigma_c^2} \\
&\quad+\mu_c\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\mu_c\frac{k_{c,\sigma_c}\sigma_c-k_c}{\sigma_c^2} \\
&= \frac{1}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)-\frac{\sigma_c}{k_c^2}k_{c,\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)-\frac{\mu_c^2 k_{c,\sigma_c}}{\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)+\frac{\mu_c^2 k_c}{\sigma_c^2}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \\
&\quad+\frac{\mu_c^2 k_{c,\sigma_c}}{\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)-\frac{\mu_c^2 k_c}{\sigma_c^2}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \\
&= \frac{1}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)-\frac{\sigma_c}{k_c^2}k_{c,\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \\
&> 0.
\end{aligned}
$$

We have that

$$
\begin{aligned}
\frac{\partial\left[\frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)+\mu_c\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right]}{\partial\mu_c} &= -\frac{\sigma_c}{k_c}\frac{\mu_c k_c}{\sigma_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{k_c}{\sigma_c}+\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)+\mu_c\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\frac{k_c}{\sigma_c} \\
&= \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \\
&> 0.
\end{aligned}
$$

$\square$

C Extensions

C.1 Condition-specific customer outcome variance

In our empirical application, the measure of s_c and the optimal sample size are insensitive to whether we estimate s_c using the treatment customers, the control customers, or customers from both conditions. However, we recognize that this will not always be the case. Our model, analysis, and algorithms can be easily extended to accommodate condition-specific customer outcome variances. In particular, the noise in Equation (1) would be distributed according to $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, s_{c|1}^2)$ for treatment ($D_i = 1$) and $\mathcal{N}(0, s_{c|0}^2)$ for control ($D_i = 0$). The likelihood function for Bayesian updating would then be given by

$$
\hat{m}_c^{tr} \mid m_c \sim \mathcal{N}\left(m_c, \frac{2\left(s_{c|1}^2+s_{c|0}^2\right)}{w_c N_{tr}}\right),
$$

and all our derivations would follow through by replacing the $4s_c^2$ term by $2\left(s_{c|1}^2+s_{c|0}^2\right)$ throughout. This applies to the analysis and algorithms of all of Sections 2, 3, 4, 5, and respective appendices.

C.2 Extensions to δ-expected improvement certification

C.2.1 Using a fixed targeting policy as a benchmark

Consider a fixed and known targeting policy π' that recommends to each segment c action $\pi'(c)$, where $\pi'(c) \in \{0, 1\}$ for every c. The firm wants to calculate a minimal sample size to train a targeting policy π_t, so that the expected performance improvement of π_t over π' is at least δ:

$$
\begin{aligned}
\min_{N_{tr}} \quad & N_{tr} && (48)\\
\text{s.t.} \quad & \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}] \geq \delta && (49)\\
& N_{tr} \geq 0. && (50)
\end{aligned}
$$

We can write the conditional expected performance improvement as:

$$
\begin{aligned}
\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi'} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right] &= \sum_{c:\pi'(c)=0} w_c \cdot I\big(\mu_c^{post} \geq 0\big)\, \mathbb{E}\left[m_c \mid \hat{m}_c^{tr}\right] \\
&\quad - \sum_{c:\pi'(c)=1} w_c \cdot I\left(\mu_c^{post} < 0\right) \mathbb{E}\left[m_c \mid \hat{m}_c^{tr}\right] \\
&= \sum_{c:\pi'(c)=0} w_c I\left(\mu_c^{post} \geq 0\right) \mu_c^{post} - \sum_{c:\pi'(c)=1} w_c I\left(\mu_c^{post} < 0\right) \mu_c^{post}.
\end{aligned}
$$

In the absence of any training data ($N_{tr} = 0$), $\mu_c^{post} = \mu_c$ for every segment c. The conditional expectation then collapses to a constant:

$$
\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi'} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right] = \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}] = \sum_{c=1}^{C} w_c \left[I\left(\mu_c \geq 0\right) - \pi'(c)\right] \mu_c.
$$

For $N_{tr} > 0$, the expectation $\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi'}\right]$ can be written as

$$
\begin{aligned}
\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi'}\right] &= \mathbb{E}\left[\mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi'} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right]\right] \\
&= \sum_{c:\pi'(c)=0} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] + \sum_{c:\pi'(c)=1} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \mu_c + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] \\
&= \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \pi'(c)\mu_c + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right],
\end{aligned}
$$

where in the second step we make use of (40) and 44. We note that the expression for $N_{tr} = 0$ coincides with the limit of the expression for $N_{tr} > 0$ as $N_{tr} \to 0$.

We conclude that

$$
\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}] = \begin{cases} \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \pi'(c) \cdot \mu_c + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \quad (51a) \\ \sum_{c=1}^{C} w_c \cdot \left[I\left(\mu_c \geq 0\right) - \pi'(c)\right] \cdot \mu_c & \text{if } N_{tr} = 0, \quad (51b) \end{cases}
$$

where $k_c = \sqrt{1 + \frac{4s_c^2}{w_c N_{tr} \sigma_c^2}}$.

This expression is simple to calculate, and is again increasing in N_{tr}. If the problem is feasible, then either $N_{tr} = 0$ is the optimal solution; or the optimal required sample size can be identified using standard root-finding algorithms (or a one-dimensional grid search), to solve for the value of N_{tr} at which constraint (49) is binding.

Similar to Proposition 2, $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}]$ is decreasing in s_c and increasing in σ_c for all $c = 1, ..., C$. However, the monotonicity properties $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}]$ with respect to μ_c depend on $\pi'(c)$. The expected performance difference $\mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi'}]$ is increasing in μ_c in the segments where $\pi'(c) = 0$, and is decreasing in μ_c otherwise.

C.2.2 Using a learnt optimal uniform policy as the benchmark

We define the best uniform policy π_{u^*} after observing the outcomes from the experiment:

$$\pi_{u^*}(c) = I\left(\sum_c w_c \mu_c^{post} \geq 0\right), \forall c.$$

We can formulate the firm's decision problem as the optimization problem

$$\min_{N_{tr}} \quad N_{tr} \tag{52}$$

$$\text{s.t.} \quad \mathbb{E}[V_{\pi_t - \pi_{u^*}}^{post}] \geq \delta \tag{53}$$

$$N_{tr} \geq 0, \tag{54}$$

where $V_{\pi_t - \pi_{u^*}}^{post}$ is similar to (9) in the main text:

$$\begin{aligned} V_{\pi_t - \pi_{u^*}}^{post} :=& I\left(\sum_c w_c \mu_c^{post} < 0\right) \cdot \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \mu_c^{post} \\ &- I\left(\sum_c w_c \mu_c^{post} \geq 0\right) \cdot \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} < 0\big) \cdot \mu_c^{post}. \end{aligned}$$

The first term captures the expected performance difference between the targeting policy π_t and the uniform policy π_0, if π_0 outperforms π_1 in the experiment. The second term captures the case when π_1 outperforms π_0. We can rewrite the expression for $V_{\pi_t - \pi_u^*}^{post}$ as follows:

$$V_{\pi_t - \pi_{u^*}}^{post} = \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \mu_c^{post} - I\left(\sum_c w_c \mu_c^{post} \geq 0\right) \cdot \sum_c w_c \mu_c^{post}. \tag{55}$$

In the main text, we show that μ_c^{post} follows a normal distribution, so the random variable $V_{\pi_t - \pi_{u^*}}^{post}$ is a sum of rectified normal distributions, with a closed-form solution for its mean:

$$\mathbb{E}[V_{\pi_t - \pi_{u^*}}^{post}] = \begin{cases} \sum_{c=1}^{C} w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c} \phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] - \left[\bar{\mu} \Phi\left(\frac{\bar{\mu}}{\bar{\sigma}}\right) + \bar{\sigma} \phi\left(\frac{\bar{\mu}}{\bar{\sigma}}\right)\right] & \text{if } N_{tr} > 0 \quad \text{(56a)} \\ \sum_{c=1}^{C} w_c \cdot I\left(\mu_c \geq 0\right) \cdot \mu_c - I\left(\sum_c w_c \mu_c \geq 0\right) \cdot \sum_c w_c \mu_c & \text{if } N_{tr} = 0 \quad \text{(56b)} \end{cases}$$

where

$$\bar{\mu} = \sum_c w_c \cdot \mu_c,\ \bar{\sigma}^2 = \sum_c \frac{w_c^2 \cdot \sigma_c^2}{k_c^2},\text{ and } k_c = \sqrt{1 + \frac{4s_c^2}{w_c N_{tr} \sigma_c^2}}.$$

C.2.3 Certification based on probable δ-expected improvement

The δ-expected improvement formulation focuses on the expectation of the performance improvement of the trained targeting policy over the benchmark. Recall that, for each possible draw of the training data, the firm uses the training data to update its posterior beliefs about the treatment effects m_c's and, as a consequence, the actions recommended by the targeting policy. Together, the updated beliefs and policy yield an expected performance difference

$$V^{post}_{\pi_t-\pi_0} = \mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right].$$

The expected performance difference $V^{post}_{\pi_t-\pi_0}$ is a random variable that depends on the draw of the training data. In the δ-expected improvement formulation (Section 3.1), we required that on average, the expected performance difference be at least δ: $\mathbb{E}[V^{post}_{\pi_t-\pi_0}] = \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}] \geq \delta$.

As an extension of the δ-expected improvement formulation, we allow the firm to specify a desired probability that the expected performance difference exceeds δ. In particular, for each draw of the training data $\hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}$, the conditional expected performance difference $V^{post}_{\pi_t-\pi_0}$ either exceeds the δ threshold or does not. We can characterize the probability that it does. We say that a trained targeting policy satisfies the $(1-\gamma)$-*probable δ-expected improvement* requirement if, with probability at least $1-\gamma$, the expected per-customer performance improvement of the trained targeting policy over the benchmark is at least δ.

Formally, the firm's optimization problem for the $(1-\gamma)$-probable δ-expected improvement requirement is

$$\min_{N_{tr}} \quad N_{tr} \tag{57}$$

$$\text{s.t.} \quad \mathbb{P}\left(V^{post}_{\pi_t-\pi_0} \geq \delta\right) \geq 1-\gamma \tag{58}$$

$$N_{tr} \geq 0. \tag{59}$$

We can solve the optimization problem in two ways: using simulations and by analytical approximation. We start with the simulation-based solution.

Simulation-based solution. Recall from Section 2 that the expected policy difference $V^{post}_{\pi_t-\pi_0}$ can be represented as a weighted sum of independent random variables:

$$V^{post}_{\pi_t-\pi_0} = \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \mu_c^{post},$$

where μ_c^{post} is the expected treatment effect in segment c after observing the training data $\hat{m}_c^{tr}$. For a given N_{tr}, we can generate B samples $\mu^{post}_{c,b}, b = 1, \ldots, B$ from the normal distribution (5), and calculate $V^{post}_{\pi_t-\pi_0}(b)$ for each $b = 1, .., B$. Condition (58) is satisfied for N_{tr}, if $V^{post}_{\pi_t-\pi_0}(b)$ is greater or equal than δ for at least $(1-\gamma)B$ samples. We can use enumeration to find a minimum N_{tr} for which the condition is satisfied. Note that similar to Section 3.2, the optimization problem might be infeasible for any N_{tr}. Our approach requires a predefined maximum sample size for the enumeration.[23] We summarize the proposed simulation-based

[23] We can improve the computational efficiency of the simulation solution by evaluating the optimization con-

solution in Algorithm 3.

Algorithm 3 PROBABLE δ-EXPECTED IMPROVEMENT SAMPLE SIZE CALCULATION (SIMULATION)

Input: Prior parameters for segment treatment effects $\{\mu_c, \sigma_c^2\}_{c=1,...,C}$; customer outcome variances $\{s_c^2\}_{c=1,...,C}$; relative segment sizes $\{w_c\}_{c=1,...,C}$; required expected improvement δ and probability requirement $1-\gamma$; simulation iterations B; maximum experiment size $N_{\max}$

```
function ESTIMATE.PROB(N_tr)
    for b = 1,...,B do
        for c = 1,...,C do
            Sample μ^post_{c,b} from distribution (5)
        Calculate V_b using (9)
    Prob ← Σ_b I(V_b ≥ δ) / B
    return Prob

function MAIN
    N*_tr = N_max + 1
    for N_tr = 0,1,...,N_max do
        Prob ← ESTIMATE.PROB(N_tr)
        if Prob ≥ 1 − γ then
            N*_tr ← N_tr
            Break for-loop
    if N*_tr ≤ N_max then
        return N*_tr
    else
        return Infeasible
```

The simulation solution requires randomly sampling posterior parameters $\mu_{c,b}^{post}$ for each iteration $b = 1,..,B$ and segment $c = 1,...,C$. Exhaustive search over the integers (up to a maximum) for N_{tr} yields the exact solution, but it can be computationally demanding. As an alternative, we propose an analytical approximation for the LHS of (58) that yields a faster solution for the $(1-\gamma)$-probable δ-expected improvement formulation.

Analytical approximation. We approximate the distribution of $V_{\pi_t-\pi_0}^{post}$ by a normal distribution:

$$V_{\pi_t-\pi_0}^{post} \overset{approx}{\sim} \mathcal{N}\left(\mu_\Delta, \sigma_\Delta^2\right), \tag{60}$$

straint (58) for multiple values of N_{tr} in parallel. However, this solution has to account for computer memory constraints. For example, generating $B = 100,000$ samples of $\mu_{c,b}^{post}$ for $C = 27$ segments and $1,000$ values of N_{tr} requires over 20GB of memory.

where $\mu_\Delta := \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$ is provided in Equation (13) and

$$\sigma^2_\Delta := \begin{cases} \sum_{c=1}^{C} w_c^2 \left[\mu_c^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1-\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) + \frac{\sigma_c^2}{k_c^2}\cdot\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right. \\ \qquad \left. + \frac{\mu_c \sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1-2\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) - \frac{\sigma_c^2}{k_c^2}\cdot\phi^2\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \quad (61a) \\ 0 & \text{if } N_{tr} = 0. \quad (61b) \end{cases}$$

Our approximation in (60) relies on the central limit theorem (CLT). It requires that there are sufficient segments C, and there are no segments for which w_c and/or μ_c^{post} are extreme outliers.[24] As (61) suggests, the derivation of the variance of the expected performance improvement $V^{post}_{\pi_t-\pi_0}$ is complicated. This is because the variance calculation accounts for two sources of variance. There is variance in the training sample coming from customer outcomes, and represented by s_c^2; in addition, there is uncertainty about the segment treatment effects m_c, represented by σ_c^2. Consequently, the variance of the expected performance improvement is a non-trivial function of μ_c, σ_c^2, s_c^2, and N_{tr}.[25]

We investigated the quality of the analytical approximation for the empirical application reported in Section 5. In Figure 4, we compare how accurately the analytical approximation estimates the probability that the expected performance difference $V^{post}_{\pi_t-\pi_0}$ exceeds $\delta = \$3.40$, where we calculate the probability using simulations. On the x-axis is the size of the training data N_{tr} used to train the targeting policy. On the y-axis is the probability that the expected performance difference $V^{post}_{\pi_t-\pi_0}$ exceeds δ. The (red) solid line captures the probability $\mathbb{P}(V^{post}_{\pi_t-\pi_0} \geq \delta)$ estimated using the simulation-based approach. The (blue) dashed line captures the probability estimated using the analytical approximation. The figure indicates that the quality of the approximation is high with as few as $C = 27$ customer segments.

We approximate the LHS of the optimization constraint (58) by replacing the expected performance improvement $V^{post}_{\pi_t-\pi_0}$ by a random variable distributed as $\mathcal{N}(\mu_\Delta, \sigma^2_\Delta)$:

$$\mathbb{P}\left(V^{post}_{\pi_t-\pi_0} \geq \delta\right) \approx \Phi\left(\frac{\mu_\Delta - \delta}{\sigma_\Delta}\right) \quad \text{for } N_{tr} > 0,$$

and reformulate the optimization problem for the $(1-\gamma)$-probable δ-expected improvement as follows:

$$\min_{N_{tr}} \quad N_{tr} \tag{62}$$

$$\text{s.t.} \quad \mu_\Delta - \sigma_\Delta \cdot z_{1-\gamma} \geq \delta \tag{63}$$

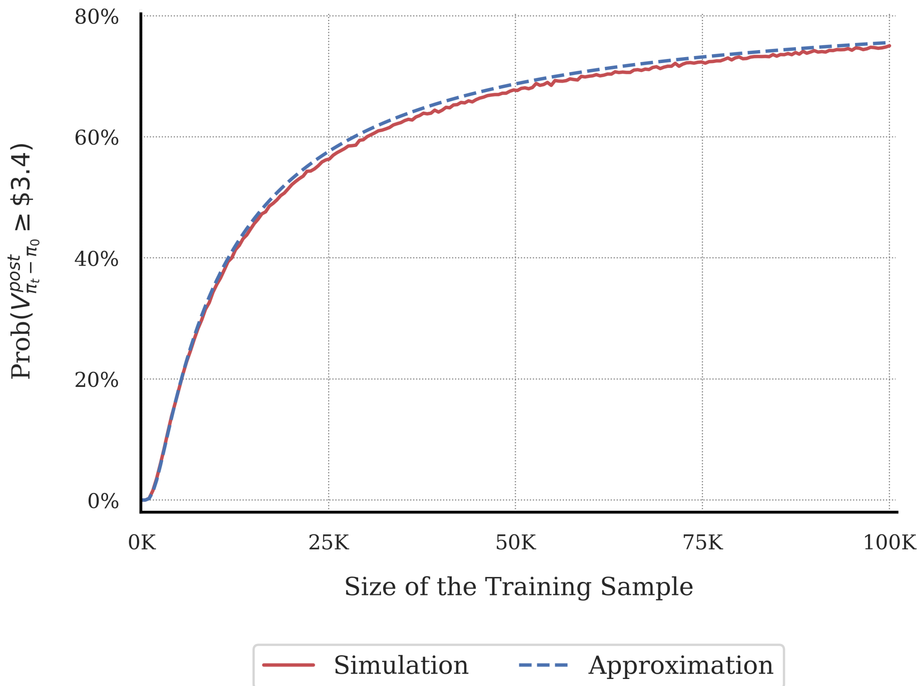
$$N_{tr} \geq 0 \tag{64}$$

This approximation highlights that we can interpret the $(1-\gamma)$-probable δ-expected improvement formulation as a generalization of the δ-expected improvement formulation. Indeed, if we consider $\gamma = 0.5$, then the optimization constraint in (63) is equivalent to $\mu_\Delta \geq \delta$. Recall that by definition, $\mu_\Delta = \mathbb{E}[\bar{V}_{\pi_t} - \bar{V}_{\pi_0}]$, so the constraint in (63) with $\gamma = 0.5$ is equivalent to the optimization constraint in (11) in the δ-expected improvement formulation. Intuitively, the $(1-\gamma)$-probable δ-expected improvement formulation identifies the minimal sample size so that the $\gamma \times 100$-th percentile of the distribution of $V^{post}_{\pi_t-\pi_0}$ is at least δ. For a

[24] Beauchamp (2018) checks that, for the sum of independent, but not necessarily identically distributed, rectified normals, the conditions of the Lyapunov version of the central limit theorem are satisfied under mild assumptions.

[25] The sample size N_{tr} is included in the term k_c.

Figure 4: Approximation for $(1-\gamma)$-Probable δ-Expected Improvement



normal distribution, the 50% percentile (median) is equal to the mean, so the (analytically approximated) 50%-probable δ-expected improvement formulation matches the δ-expected improvement formulation.

We can find the solution to the approximate $(1-\gamma)$-probable δ-expected improvement formulation by enumeration. For each N_{tr}, we can calculate μ_Δ and σ_Δ using the closed-form expressions (13) and (61), and check the optimization constraint (63). By iterating over integer values for N_{tr}, we can identify the minimum sample size for which the optimization constraint is satisfied. We note that, compared to our results in Section 3.2, we cannot prove monotonicity of the optimization constraint in (63), because both μ_Δ and σ_Δ are increasing in N_{tr} (see proofs in Appendix B.2 and C.2.4). We note that the optimization constraint may be infeasible, so our algorithm requires a maximum sample size (as an input) for the enumeration to stop.

The approximation-based solution is substantially faster than the simulation solution, as it relies on the closed-form expressions for evaluating the optimization constraint, without random sampling. Using these expressions, our implementation evaluates the optimization constraint in (63) for multiple values of N_{tr} in a single pass using matrix operations, and avoids the need for a for-loop over integer values of N_{tr}. We formalize our analytical approximation-based solution for the $(1-\gamma)$-probable δ-expected improvement formulation as Algorithm 4.

C.2.4 Monotonicity of σ_Δ in N_{tr}

Proposition 4. *σ_Δ^2 is monotonically increasing in N_{tr}.*

Algorithm 4 Probable δ-Expected Improvement Sample Size Calculation (Analytical Approximation)

Input: Prior parameters for segment treatment effects $\{\mu_c, \sigma_c^2\}_{c=1,\ldots,C}$; customer outcome variances $\{s_c^2\}_{c=1,\ldots,C}$; relative segment sizes $\{w_c\}_{c=1,\ldots,C}$; required expected improvement δ and probability requirement $1-\gamma$; maximum experiment size $N_{\max}$

if $\sum_c w_c \cdot I(\mu_c \geq 0) \cdot \mu_c \geq \delta$ **then**
  $N_{tr}^* \leftarrow 0$
else
  $N_{tr}^* \leftarrow \infty$
  for $N_{tr} = 0, 1, \ldots, N_{\max}$ **do**
    Compute μ_Δ and σ_Δ^2 using (13),(61)
    if $\mu_\Delta - \sigma_\Delta \cdot z_{1-\gamma} \geq \delta$ **then**
      $N_{tr}^* \leftarrow N_{tr}$
      Break for-loop
if $N_{tr}^* \leq N_{\max}$ **then**
  return N_{tr}^*
else
  return Infeasible

Proof. Recall that for $N_{tr} > 0$ we have

$$\sigma_\Delta^2 = \sum_{c=1}^{C} w_c^2 \left[\mu_c^2 \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) - \frac{\sigma_c^2}{k_c^2}\phi^2\left(\frac{\mu_c k_c}{\sigma_c}\right) \right.$$
$$\left. + \frac{\sigma_c^2}{k_c^2}\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\mu_c \sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - 2\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right)\right].$$

We differentiate σ_Δ^2 with respect to N_{tr} to obtain

$$\frac{\partial \sigma_\Delta^2}{\partial N_{tr}} = -\sum_{c=1}^{C} \frac{2 \cdot w_c^2 \cdot \sigma_c^2}{k_c^3} \cdot k_{c,N_{tr}} \left[\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) - \phi^2\left(\frac{\mu_c k_c}{\sigma_c}\right) - \frac{\mu_c k_c}{\sigma_c} \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right],$$

where we recall from (45) that $k_{c,N_{tr}} := \frac{\partial k_c}{\partial N_{tr}} = -\frac{2 s_c^2}{w_c N_{tr}^2 \sigma_c^2 k_c} < 0$. It suffices to show that the expression in the square brackets is positive. To show this, we next show that

$$\Phi(x) - \phi^2(x) - x\Phi(x)\phi(x) > 0 \quad \text{for any} \quad x \in \mathbb{R}. \tag{65}$$

Marshall and Olkin (2007) demonstrate that the hazard rate of the standard normal distribution is log-concave. The definition of log-concavity for a function f is that $f(\theta x + (1-\theta)y) \geq f(x)^\theta f(y)^{1-\theta}$ for all x, y in the domain of f and $0 < \theta < 1$. We apply this definition on the hazard rate of the standard normal, $\phi(x)/(1-\Phi(x))$, for $y = -x$ and $\theta = 1/2$, and we get

$$\frac{\phi(x)}{1-\Phi(x)} \cdot \frac{\phi(-x)}{1-\Phi(-x)} \leq \left(\frac{\phi(0)}{1-\Phi(0)}\right)^2 = \frac{2}{\pi} < 1,$$

which implies $\phi^2(x) < \Phi(x)\,(1-\Phi(x))$. We substitute this inequality for $\phi^2(x)$ into the LHS of (65) to obtain

$$\begin{aligned}\Phi(x)-\phi^2(x)-x\Phi(x)\phi(x) &> \Phi(x)-\Phi(x)\,(1-\Phi(x))-x\Phi(x)\phi(x)\\ &= \Phi^2(x)-x\Phi(x)\phi(x)\\ &= \Phi(x)\,(\Phi(x)-x\phi(x))\,.\end{aligned}$$

The expression $\Phi(x)-x\phi(x)$ is positive for all $x\in\mathbb{R}$, because (i) its derivative satisfies

$$\frac{d\,(\Phi(x)-x\phi(x))}{dx} = x^2\phi(x) \geq 0,$$

with strict inequality everywhere other than at $x=0$; while (ii) $\lim_{x\to-\infty}\,(\Phi(x)-x\phi(x)) = 0$. The claim follows. □

C.2.5 Optimizing sample sizes for each segment

We use notation $N_{tr,c}$ to denote the firm's decision for the sample size corresponding to segment c for policy training. We assume that, within each segment, the sample size is equally divided between the treatment and control experimental conditions. The firm's decision problem for certification based on δ-expected improvement is then formulated as

$$\min_{N_{tr,1},\ldots,N_{tr,C}} \quad \sum_{c=1}^{C} N_{tr,c} \tag{66}$$

$$\text{s.t.} \quad \mathbb{E}[\bar{V}_{\pi_t}-\bar{V}_{\pi_0}] \geq \delta \tag{67}$$

$$N_{tr,c} \geq 0 \quad \forall c, \tag{68}$$

where the expected performance difference can be expressed as

$$\mathbb{E}[\bar{V}_{\pi_t}-\bar{V}_{\pi_0}] = \mathbb{E}\left[V^{post}_{\pi_t-\pi_0}\right] = \sum_{c=1}^{C} w_c\cdot\mathbb{E}\left[I\big(\mu^{post}_c \geq 0\big)\cdot\mu^{post}_c\right],$$

by (9), with

$$\mathbb{E}\left[I\big(\mu^{post}_c \geq 0\big)\cdot\mu^{post}_c\right] = \begin{cases} \mu_c\Phi\left(\dfrac{\mu_c k_c}{\sigma_c}\right)+\dfrac{\sigma_c}{k_c}\phi\left(\dfrac{\mu_c k_c}{\sigma_c}\right) & \text{if } N_{tr,c}>0 \quad (69a)\\ I\left(\mu_c\geq 0\right)\cdot\mu_c & \text{if } N_{tr,c}=0, \quad (69b)\end{cases}$$

where

$$k_c := \sqrt{1+\frac{4s_c^2}{N_{tr,c}\sigma_c^2}}.$$

Similarly to the case with proportional allocation of the sample into segments, the expression in (69a) is separable and monotonically increasing in $N_{tr,c}$ (see proof in Appendix B.2). The optimization problem (66)-(68) is thus generally tractable. In particular, our optimization problem can be reformulated as a multivariate maximization problem with a separable concave objective. There are established algorithms for solving this class of problems (Patriksson, 2008).

C.2.6 Correlations in treatment effects across segments

We assume that the segment treatment effects are correlated and follow a multivariate normal prior distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In this specification, $\boldsymbol{\mu}$ is a C-dimensional column vector of prior means. $\boldsymbol{\Sigma}$ is a $C \times C$ symmetric, positive semi-definite matrix, where the c-th diagonal entry is σ_c^2, and the off-diagonal elements are given by $\Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, for $i \neq j$. When $N_{tr} > 0$ and conditional on the by-segment treatment effects m_c, the likelihood function for the observed average treatment effect $\hat{m}_c^{tr}$ in segment c is given by

$$\hat{m}_c^{tr} \mid m_c \sim \mathcal{N}\left(m_c, \frac{4s_c^2}{w_c N_{tr}}\right).$$

Standard Bayesian updating yields that the firm's posterior for the segment treatment effects $m_1, \dots, m_c$ given $\hat{m}_1^{tr}, \dots, \hat{m}_C^{tr}$ is

$$m_1, \dots, m_C \mid \hat{m}_1^{tr}, \dots, \hat{m}_C^{tr} \sim \mathcal{N}\left(\boldsymbol{\mu}^{post}, \boldsymbol{\Sigma}^{post}\right), \tag{70}$$

where the posterior parameters are given by

$$\boldsymbol{\mu}_c^{post} = \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{S}^{-1}\left(\hat{m}_1^{tr}, \dots, \hat{m}_C^{tr}\right)^{\top}\right) \tag{71}$$

$$\boldsymbol{\Sigma}^{post} = \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)^{-1}, \tag{72}$$

and where we denote by $\boldsymbol{S}$ the $C \times C$ diagonal matrix where the c-th diagonal entry is $\frac{4s_c^2}{w_c N_{tr}}$ (and all the non-diagonal entries are zero).

When $N_{tr} > 0$, the distribution of vector $\hat{m}_1^{tr}, \dots, \hat{m}_C^{tr}$ is

$$\hat{m}_1^{tr}, \dots, \hat{m}_C^{tr} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{S}\right). \tag{73}$$

From (71), it follows that the distribution of random variable $\boldsymbol{\mu}^{post}$ is multivariate normal, with

$$\begin{aligned}
\boldsymbol{\mu}^{post} &\sim \mathcal{N}\left(\boldsymbol{\mu}, \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)^{-1}\boldsymbol{S}^{-1}\left(\boldsymbol{\Sigma} + \boldsymbol{S}\right)\left(\boldsymbol{S}^{-1}\right)^{\top}\left(\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)^{-1}\right)^{\top}\right) \\
&= \mathcal{N}\left(\boldsymbol{\mu}, \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)^{-1}\boldsymbol{S}^{-1}\left(\boldsymbol{\Sigma} + \boldsymbol{S}\right)\boldsymbol{S}^{-1}\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)^{-1}\right) \\
&= \mathcal{N}\left(\boldsymbol{\mu}, \left(\boldsymbol{I} + \boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{I} + \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\left(\boldsymbol{I} + \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)^{-1}\right) \\
&= \mathcal{N}\left(\boldsymbol{\mu}, \left(\boldsymbol{I} + \boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}\right),
\end{aligned} \tag{74}$$

where we use $\boldsymbol{I}$ to denote the $C \times C$ identity matrix. We note that the expression in (74) recovers (5) in the paper in the univariate case.

As in our main analysis, we study the targeting policy that recommends treating customers if and only if the posterior mean in the segment is non-negative:

$$\pi_t(c) := I\left(\mu_c^{post} \geq 0\right) = \begin{cases} 1 & \text{if } \mu_c^{post} \geq 0 \\ 0 & \text{if } \mu_c^{post} < 0, \end{cases} \tag{75}$$

where $\pi_t(c) = 1(0)$ indicates that the policy assigns the treatment (control) to all customers in segment c, and where $I(\cdot)$ denotes the indicator function.

Central to our analysis is the expectation of

$$V_{\pi_t-\pi_0}^{post} = \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \mu_c^{post},$$

where each segment is accounted for separately. By the multivariate normality of $\boldsymbol{\mu}^{post}$ and (74), the distribution of μ_c^{post} is normal, with mean μ_c, and variance equal to the c-th diagonal element of $\left(\boldsymbol{I} + \boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}$.

Although the expression for $\mathbb{E}\left[V_{\pi_t-\pi_0}^{post}\right]$ admits a closed form, the expression is much more complex than when the segment treatment effects are independent. To illustrate, we provide the expression for the variance of random variable μ_1^{post} for the case of $C = 2$ segments. This is the first diagonal element of the covariance matrix of $\boldsymbol{\mu}^{post}$, i.e., the first diagonal element of $\left(\boldsymbol{I} + \boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}$. The variance is

$$\sigma_1^2 \frac{w_1 N_{tr} w_2 N_{tr} \sigma_1^2 \sigma_2^2 (1-\rho_{12}^2) + w_1 N_{tr} \sigma_1^2 4 s_2^2 + w_2 N_{tr} \sigma_2^2 \rho_{12}^2 4 s_1^2}{w_1 N_{tr} w_2 N_{tr} \sigma_1^2 \sigma_2^2 (1-\rho_{12}^2) + w_1 N_{tr} \sigma_1^2 4 s_2^2 + w_2 N_{tr} \sigma_2^2 4 s_1^2 + 16 s_1^2 s_2^2},$$

for $\rho_{12} \neq 1$. Notice that, for $\rho_{12} = 0$, this expression collapses to σ_1^2/k_1^2, matching the variance in (5).

This analysis also has a practical limitation. Defining a multivariate normal prior for the segment effects requires specifying the correlations across segments. These may be hard to obtain in practice.

C.2.7 Log-normal distribution for monetary outcomes

We consider an alternative specification for the monetary outcomes Y_i by assuming a log-normal distribution:

$$Y_i = \exp\left(\alpha_c + m_c \cdot D_i + \epsilon_i\right), \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, s_c^2) \tag{76}$$

where m_c is a segment-specific treatment effect for segment c distributed according to a normal prior $m_c \sim \mathcal{N}\left(\mu_c, \sigma_c^2\right)$, $c = 1, .., C$, $D_i = 1$ if customer i receives the treatment, and $D_i = 0$ if customer i receives the control. Compared to the main text, this specification assumes positive values for Y_i in both treatment and control conditions, and a multiplicative treatment effect m_c.

Using the training data, the firm can estimate the treatment effect coefficient for each segment c:

$$\tilde{m}_c^{tr} := \overline{\ln Y}_{c|1}^{tr} - \overline{\ln Y}_{c|0}^{tr}, \qquad c = 1, .., C,$$

where $\overline{\ln Y}_{c|1}^{tr}$ and $\overline{\ln Y}_{c|0}^{tr}$ are the average log-transformed outcomes in the treatment and control conditions in the training sample. We recognize that the logarithm of the log-normally distributed random variable Y_i is normally distributed, so by standard Bayesian updating, we can obtain the posterior distribution for the segment treatment effect m_c conditional on $\tilde{m}_c^{tr}$:

$$m_c \mid \tilde{m}_c^{tr} \sim \mathcal{N}\left(\mu_c^{post}, (\sigma_c^{post})^2\right), \tag{77}$$

where the posterior parameters μ_c^{post} and σ_c^{post} are given by

$$\mu_c^{post} = \left(\frac{1}{\sigma_c^2} + \frac{w_c N_{tr}}{4 s_c^2}\right)^{-1} \left(\frac{\mu_c}{\sigma_c^2} + \frac{w_c N_{tr}}{4 s_c^2} \cdot \tilde{m}_c^{tr}\right) \tag{78}$$

$$(\sigma_c^{post})^2 = \left(\frac{1}{\sigma_c^2} + \frac{w_c N_{tr}}{4 s_c^2}\right)^{-1}. \tag{79}$$

Similar to the main text, we study the targeting policy that only treats customers in segments with a non-negative posterior mean:

$$\pi_t(c) := I\left(\mu_c^{post} \geq 0\right) = \begin{cases} 1 & \text{if } \mu_c^{post} \geq 0 \\ 0 & \text{if } \mu_c^{post} < 0, \end{cases} \tag{80}$$

where $\pi_t(c) = 1(0)$ indicates that the policy assigns the treatment (control) to all customers in segment c, and where $I(\cdot)$ denotes the indicator function.

We can formulate the firm's decision problem as the optimization problem

$$\min_{N_{tr}} \quad N_{tr} \tag{81}$$

$$\text{s.t.} \quad \mathbb{E}[V_{\pi_t-\pi_0}^{post}] \geq \delta \tag{82}$$

$$N_{tr} \geq 0, \tag{83}$$

where $V_{\pi_t-\pi_0}^{post}$ is similar to (9):

$$\begin{aligned} V_{\pi_t-\pi_0}^{post} &:= \mathbb{E}\left[\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \tilde{m}_1^{tr}, \ldots, \tilde{m}_C^{tr}\right] \\ &= \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \mathbb{E}\left[\bar{Y}_{c|1}^{test} - \bar{Y}_{c|0}^{test} \mid \tilde{m}_c^{tr}\right] \\ &= \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \exp\left(\alpha_c + \frac{s_c^2}{2}\right) \cdot \left(\exp\left(\mu_c^{post} + \frac{(\sigma_c^{post})^2}{2}\right) - 1\right), \end{aligned} \tag{84}$$

where the last step uses the moment generating function of the normal distribution: $\mathbb{E}[\exp(\epsilon_i)] = \exp\left(\frac{s_c^2}{2}\right)$, and $\mathbb{E}[\exp(m_c) \mid \tilde{m}_c^{tr}] = \exp\left(\mu_c^{post} + \frac{(\sigma_c^{post})^2}{2}\right)$. Notice that for a given $N_{tr} > 0$, the posterior mean μ_c^{post} follows a normal distribution

$$\mu_c^{post} \sim \mathcal{N}\left(\mu_c, \frac{\sigma_c^2}{k_c^2}\right) \text{ with } k_c = \sqrt{1 + \frac{4s_c^2}{w_c N_{tr} \sigma_c^2}}$$

and the posterior standard deviation σ_c^{post} is a constant. We can thus derive a closed-form expression for the expectation $\mathbb{E}[V_{\pi_t-\pi_0}^{post}]$ for $N_{tr} > 0$:

$$\begin{aligned} \mathbb{E}[V_{\pi_t-\pi_0}^{post}] &= \sum_{c=1}^{C} w_c \cdot \exp\left(\alpha_c + \frac{s_c^2}{2}\right) \cdot \mathbb{E}\left[I\big(\mu_c^{post} \geq 0\big) \cdot \left(\exp\left(\mu_c^{post} + \frac{(\sigma_c^{post})^2}{2}\right) - 1\right)\right] \\ &= \sum_{c=1}^{C} w_c \cdot \exp\left(\alpha_c + \frac{s_c^2}{2}\right) \cdot \left(\exp\left(\frac{(\sigma_c^{post})^2}{2}\right) \cdot \mathbb{E}\left[I\big(\mu_c^{post} \geq 0\big) \cdot \exp\left(\mu_c^{post}\right)\right] - \mathbb{P}\left(\mu_c^{post} \geq 0\right)\right) \\ &= \sum_{c=1}^{C} w_c \cdot \exp\left(\alpha_c + \frac{s_c^2}{2}\right) \cdot \left(\exp\left(\mu_c + \frac{\sigma_c^2}{2}\right) \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c} + \frac{\sigma_c}{k_c}\right) - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right). \end{aligned} \tag{85}$$

We explain the last step. First, we have $\mathbb{P}\left(\mu_c^{post} \geq 0\right) = \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)$. For the expectation, we use the law of total expectation to write

$$\mathbb{E}\left[I\big(\mu_c^{post} \geq 0\big) \cdot \exp\left(\mu_c^{post}\right)\right] = \mathbb{E}\left[\exp\left(\mu_c^{post}\right) \mid \mu_c^{post} \geq 0\right] \cdot \mathbb{P}\left(\mu_c^{post} \geq 0\right) + 0 \cdot \mathbb{P}\left(\mu_c^{post} < 0\right),$$

and we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\mu_c^{post}\right) \mid \mu_c^{post} \geq 0\right] &= \mathbb{E}\left[\exp\left(\mu_c^{post}\right) \mid \exp\left(\mu_c^{post}\right) \geq 1\right] \\ &= \exp\left(\mu_c + \frac{\sigma_c^2}{2k_c^2}\right) \cdot \frac{\Phi\left(\frac{\mu_c + \sigma_c^2/k_c^2}{\sigma_c/k_c}\right)}{\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)},\end{aligned}$$

by the formula for the conditional expectation of the log-normal. With a bit of algebra one can see that

$$\frac{\left(\sigma_c^{post}\right)^2}{2} + \frac{\sigma_c^2}{2k_c^2} = \frac{\sigma_c^2}{2}.$$

Then the expression in (85) follows.

The closed-form expression for the constraint (82) allows a one-dimensional grid search to solve the sample size optimization problem relatively quickly. Notice also that the log-normal specification requires firms to specify the parameter α_c, in addition to parameters s_c, μ_c, and σ_c. In practice, for each segment c, α_c can be obtained from an average log-transformed outcome in the control condition, as the expectation satisfies that $\mathbb{E}[\ln Y_i \mid D_i = 0] = \alpha_c$.

D Details for (α, β) Policy Certification

D.1 Rewriting the test statistic

In this subsection we consider the test statistic T defined in Equation (22):

$$T = \frac{\bar{V}_{\pi_t} - \bar{V}_{\pi_0}}{\sqrt{\operatorname{Var}\left(\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right)}}$$

For the numerator of the test statistic, we recall Equation (7) from Section 2:

$$\bar{V}_{\pi_t} - \bar{V}_{\pi_0} = \sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \hat{m}_c^{test}.$$

For the denominator of the test statistic, we note that conditioned on random variables $\hat{m}_c^{tr}, c = 1, \ldots, C$, the variance $\operatorname{Var}\left(\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right)$ is a random variable itself. We calculate this conditional variance to be

$$\begin{aligned}\operatorname{Var}\left(\bar{V}_{\pi_t} - \bar{V}_{\pi_0} \mid \hat{m}_1^{tr}, \dots, \hat{m}_C^{tr}\right) &= \operatorname{Var}\left(\sum_{c=1}^{C} w_c \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \hat{m}_c^{test} \mid \hat{m}_1^{tr}, \dots, \hat{m}_C^{tr}\right) \\ &= \sum_{c=1}^{C} w_c^2 \cdot I\big(\mu_c^{post} \geq 0\big) \cdot \operatorname{Var}\left(\hat{m}_c^{test} \mid \hat{m}_c^{tr}\right).\end{aligned}$$

We can thus rewrite the test statistic T as

$$T = \frac{\sum_{c=1}^{C} w_c \cdot I(\mu_c^{post} \geq 0) \cdot \hat{m}_c^{test}}{\sqrt{\sum_{c=1}^{C} w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \operatorname{Var}\left(\hat{m}_c^{test} \mid \hat{m}_c^{tr}\right)}}$$

where

$$\begin{aligned}\operatorname{Var}\left(\hat{m}_c^{test} \mid \hat{m}_c^{tr}\right) &= \operatorname{Var}\left(m_c + \bar{\epsilon}_{c|1}^{test} - \bar{\epsilon}_{c|0}^{test} \,\middle|\, \hat{m}_c^{tr}\right) \\ &= \operatorname{Var}(m_c \mid \hat{m}_c^{tr}) + \frac{4s_c^2}{w_c N_{test}} \\ &= \left(\sigma_c^{post}\right)^2 + \frac{4s_c^2}{w_c N_{test}},\end{aligned}$$

and where for each segment c, $\bar{\epsilon}_{c|1}^{test}$ and $\bar{\epsilon}_{c|0}^{test}$ capture the average idiosyncratic noise in the certification data in the treatment and control conditions.

We can characterize the distribution of the test statistic, conditioned on training outcomes $\hat{m}_1^{tr}, \dots, \hat{m}_C^{tr}$:

$$\begin{aligned}T &= \frac{\sum_{c=1}^{C} w_c \cdot I(\mu_c^{post} \geq 0) \cdot \hat{m}_c^{test}}{\sqrt{\sum_{c=1}^{C} w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \operatorname{Var}\left(\hat{m}_c^{test} \mid \hat{m}_c^{tr}\right)}} \\ &= \sum_{c=1}^{C} \left[\frac{w_c \cdot I(\mu_c^{post} \geq 0)}{\sqrt{\sum_{c=1}^{C} w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right)}} \cdot \hat{m}_c^{test} \right],\end{aligned}$$

so that

$$T \mid \hat{m}_1^{tr}, \dots, \hat{m}_C^{tr} \sim \mathcal{N}\left(\frac{\sum_c w_c \cdot I(\mu_c^{post} \geq 0) \cdot \mu_c^{post}}{\sqrt{\sum_c w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right)}}, 1 \right). \tag{86}$$

The last step recognizes that conditional on the training data, $\hat{m}_c^{test}$ are normally distributed,

$$\hat{m}_c^{test} \mid \hat{m}_c^{tr} \sim \mathcal{N}\left(\mu_c^{post}, (\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right),$$

and that the sum of independent normally distributed random variables follows a normal distribution.

D.2 Simulation-based solution for (α, β)-policy certification

For a given N_{tr} and N_{test}, we can generate B samples $\mu_{c,b}^{post}, b = 1, \dots, B$ from the normal distribution in (5), and conditional on these samples, generate the outcomes of the testing data $\hat{m}_{c,b}^{test}$ from the following distribution:

$$\hat{m}_{c,b}^{test} \mid \mu_{c,b}^{post} \sim \mathcal{N}\left(\mu_{c,b}^{post}, (\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right).$$

We can then calculate $T(b)$ for each $b = 1, .., B$ and check if the certification condition (25) holds. For a given N_{tr} and N_{test}, the certification condition is satisfied if $T(b)$ is greater or equal than $z_{1-\alpha}$ for at least $(1-\beta)B$ samples. We can use a two-dimensional grid search to find a minimum experimental sample size $N = N_{tr} + N_{test}$ for which the condition is satisfied. For each (N_{tr}, N_{test}) combination, this approach requires generating B random samples and checking whether the certification condition holds for each sample, which will often be impractical.

We propose two improvements to this simulation-based solution. First, in Appendix D.1, we demonstrate that conditional on the training data, the statistic T follows a normal distribution $\mathcal{N}(\mathcal{M}, 1)$, where $\mathcal{M}$ is itself a random variable that depends on the outcomes in the training data:

$$\mathcal{M} := \frac{\sum_c w_c \cdot I(\mu_c^{post} \geq 0) \cdot \mu_c^{post}}{\sqrt{\sum_c w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right)}}. \tag{87}$$

By applying the total probability theorem, we rewrite

$$\mathbb{P}\left(T \geq z_{1-\alpha}\right) = \mathbb{E}\left[\mathbb{P}\left(T \geq z_{1-\alpha} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right)\right] = \mathbb{E}\left[\Phi\left(\mathcal{M} - z_{1-\alpha}\right)\right]. \tag{88}$$

The proposed efficiency improvement is to use Equation (88) to estimate the LHS of the certification constraint (25) by only sampling the training data outcomes $\mu_{c,b}^{post}$. This eliminates the step of additionally generating the testing outcomes $\hat{m}_{c,b}^{test}$.

Notice that the random variable $\mathcal{M}$ is always non-negative. If $\alpha \geq 1-\beta$, the optimization constraint (25) is satisfied for any N_{tr} and N_{test}, and the (α, β)-policy certification requires no experimentation:

$$\mathbb{E}\left[\Phi\left(\mathcal{M} - z_{1-\alpha}\right)\right] \geq \mathbb{E}\left[\Phi\left(-z_{1-\alpha}\right)\right] = 1 - \Phi(z_{1-\alpha}) = \alpha \geq 1 - \beta.$$

We thus consider the practically interesting case $\alpha + \beta < 1$ hereafter.

The second efficiency improvement follows from recognizing that the random variable $\mathcal{M}$, and therefore also the expectation in (88), are monotonically increasing in N_{test}. For a given N_{tr}, we can employ root-finding algorithms such as bisection to identify a minimum value of $N_{test}^*(N_{tr})$ for which the certification constraint (25) is satisfied.[26] Root-finding methods for monotonic functions substantially improve efficiency compared to enumeration. This observation reduces the need for exhaustive search from two dimensions to only one.

We summarize the proposed simulation-based solution in Algorithm 5. The algorithm relies on a one-dimensional exhaustive search over integer values for N_{tr} (up to a maximum), and requires, as an input, a maximum experiment size for the enumeration. Firms often have a limited customer base, which may constrain the maximum experiment size. The certification problem might be practically infeasible with the available sample. However, theoretically, the certification problem is always feasible if N_{tr} and N_{test} are both

[26] Our Algorithm 5 estimates the probability in (88) by sampling. In our implementation, we use a large number of simulation iterations ($B = 100,000$), which ensures that the outcome of our root-finding procedure is reliable despite simulation noise.

allowed to be large enough. Intuitively, the uncertainty about the treatment effects σ_c^{post} resolves with large N_{tr}, and the variance of the average idiosyncratic noise in the testing sample vanishes with large N_{test}:

$$\lim_{N_{test}\to\infty}\lim_{N_{tr}\to\infty}\mathbb{P}\left(T \geq z_{1-\alpha}\right) = \lim_{N_{test}\to\infty}\lim_{N_{tr}\to\infty}\mathbb{E}\left[\Phi\left(\mathcal{M} - z_{1-\alpha}\right)\right] = 1 > 1-\beta \tag{89}$$

where we have used that $\lim_{N_{tr}\to\infty}(\sigma_c^{post})^2 = 0$ and $\sum_c w_c \cdot I(\mu_c^{post} \geq 0) \cdot \mu_c^{post} \geq 0$.

Algorithm 5 (α, β) SAMPLE SIZE CALCULATION (SIMULATION)

Input: Prior parameters for segment treatment effects $\{\mu_c, \sigma_c^2\}_{c=1,\ldots,C}$; customer outcome variances $\{s_c^2\}_{c=1,\ldots,C}$; relative segment sizes $\{w_c\}_{c=1,\ldots,C}$; simulation iterations B; required confidence $1-\alpha$; required power $1-\beta$; maximum experiment size $N_{\max}$

```
function ESTIMATE.PROB(N_tr, N_test)
    for b = 1, ..., B do
        for c = 1, ..., C do
            Sample μ^post_{c,b} from distribution (5)
        Calculate M_b using (87)
    Prob ← (Σ_b Φ(M_b − z_{1−α})) / B
    return Prob

function SOLVE.NTEST(N_tr, N^max_test)
    if ESTIMATE.PROB(N_tr, N^max_test) < 1 − β then
        (infeasible) N*_test = N^max_test + 1
    else
        N*_test ← ⌈Solve for N_test : ESTIMATE.PROB(N_tr, N_test) = 1 − β⌉
    return N*_test

function MAIN
    N_curr = N_max + 1
    for N_tr = 0, 1, ..., N_max do
        if N_tr ≥ N_curr then
            break
        N_test ← SOLVE.NTEST(N_tr, N_curr − N_tr)
        if N_tr + N_test < N_curr then
            (N*_tr, N*_test) ← (N_tr, N_test)
            N_curr ← N_tr + N_test
    if N_curr ≤ N_max then
        return (N*_tr, N*_test)
    else
        return Infeasible
```

The simulation-based solution for the (α, β)-policy certification problem is computationally demanding even with our proposed improvements. The solution involves randomly sampling μ_c^{post} to estimate the probability in the optimization constraint (25) for given N_{tr} and N_{test}. For a given N_{tr}, we estimate the probability for multiple values of N_{test} in order to identify $N_{test}^*(N_{tr})$ using a root-finding algorithm. We have to consider all integer values (up to a maximum) for N_{tr}, because the implicit function $N^*_{test}(N_{tr})$ can be non-monotonic.[27] In our empirical application in Section 5 we show that the required computation time can be substantial. In Section 4.2 we propose an analytical approximation to the optimization problem (24)-(26) that leads to a computationally efficient solution.

D.3 Details for approximate analytical solution for (α, β)-policy certification

D.3.1 Complete specification of the certification constraint approximation

In this subsection we provide a complete specification for the proposed approximation for the optimization constraint in the (α, β) certification formulation. We have

$$\mathbb{P}(T \geq z_{1-\alpha}) \approx \Phi\left(\kappa \cdot \left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right)\right),$$

where

$$\begin{aligned}
\mu_\Delta &= \begin{cases} \sum_{c=1}^C w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \\ \sum_{c=1}^C w_c \cdot I(\mu_c \geq 0) \cdot \mu_c & \text{if } N_{tr} = 0, \end{cases} \\
\mu_U &= \begin{cases} \sum_{c=1}^C w_c^2 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right) \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) & \text{if } N_{tr} > 0 \\ \sum_{c=1}^C w_c^2 \cdot I(\mu_c \geq 0) \cdot \left(\sigma_c^2 + \frac{4s_c^2}{w_c N_{test}}\right) & \text{if } N_{tr} = 0, \end{cases} \\
\kappa &:= \kappa(N_{tr}, N_{test}) = \frac{1}{\sqrt{1 + \frac{\mu_\Delta^2}{\mu_U} \cdot \sigma_{\Delta U}^2}},
\end{aligned} \tag{90}$$

and

$$\begin{aligned}
\sigma_{\Delta U}^2 &= \begin{cases} \frac{\sigma_\Delta^2}{\mu_\Delta^2} + \frac{\sigma_U^2}{4\mu_U^2} \\ \quad - \sum_c \frac{w_c^3}{\mu_\Delta \mu_U} \left[\left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right)\left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right)\right. \\ \qquad \left. \cdot \left(\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right)\right] & \text{if } N_{tr} > 0 \\ 0 & \text{if } N_{tr} = 0, \end{cases} \\
\sigma_\Delta^2 &= \begin{cases} \sum_{c=1}^C w_c^2 \left[\mu_c^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) + \frac{\sigma_c^2}{k_c^2} \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right. \\ \qquad \left. + \frac{\mu_c \sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - 2\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) - \frac{\sigma_c^2}{k_c^2} \cdot \phi^2\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \\ 0 & \text{if } N_{tr} = 0, \end{cases} \\
\sigma_U^2 &= \begin{cases} \sum_{c=1}^C w_c^4 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right)^2 \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) & \text{if } N_{tr} > 0 \\ 0 & \text{if } N_{tr} = 0. \end{cases}
\end{aligned}$$

From (90), it is clear that $\kappa(N_{tr}, N_{test}) \in (0, 1)$ for all $N_{tr} \geq 0$ and $N_{test} > 0$.

[27] We provide an example of a non-monotone relationship $N^*_{test}(N_{tr})$ in Appendix D.5.

D.3.2 Derivation of the approximation

In this subsection we derive an approximation for the optimization constraint in the (α, β) certification formulation. Recall from (86) that conditional on the training data, T follows a normal distribution. By the total probability theorem, we write

$$\begin{aligned}\mathbb{P}(T \geq z_{1-\alpha}) &= \mathbb{E}\left[\mathbb{P}\left(T \geq z_{1-\alpha} \mid \hat{m}_1^{tr}, \ldots, \hat{m}_C^{tr}\right)\right] \\ &= \mathbb{E}\left[\Phi\left(\frac{\sum_c w_c \cdot I(\mu_c^{post} \geq 0) \cdot \mu_c^{post}}{\sqrt{\sum_c w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right)}} - z_{1-\alpha}\right)\right], \end{aligned} \tag{91}$$

where the expectations are with respect to random variable μ_c^{post}, which for $N_{tr} > 0$ is normally distributed,

$$\mu_c^{post} \sim \mathcal{N}\left(\mu_c, \frac{\sigma_c^4}{\sigma_c^2 + \frac{4s_c^2}{w_c N_{tr}}}\right) = \mathcal{N}\left(\mu_c, \frac{\sigma_c^2}{k_c^2}\right),$$

and for $N_{tr} = 0$ has all its mass at $\mu_c^{post} = \mu_c$.

We now consider the random variable inside the $\Phi(\cdot)$ term:

$$\mathcal{M} := \frac{\sum_c w_c \cdot I(\mu_c^{post} \geq 0) \cdot \mu_c^{post}}{\sqrt{\sum_c w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right)}} = \frac{V_{\pi_t - \pi_0}^{post}}{\sqrt{U}}$$

where $V_{\pi_t-\pi_0}^{post} = \sum_c w_c \cdot I(\mu_c^{post} \geq 0) \cdot \mu_c^{post}$ by (9), and we denote

$$U := \sum_c w_c^2 \cdot I(\mu_c^{post} \geq 0) \cdot \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right).$$

We can characterize the mean and variance of random variables $V_{\pi_t-\pi_0}^{post}$ and U:

$$\mu_\Delta := \mathbb{E}\left[V_{\pi_t-\pi_0}^{post}\right] = \begin{cases} \sum_{c=1}^C w_c \left[\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \\ \sum_{c=1}^C w_c \cdot I\left(\mu_c \geq 0\right) \cdot \mu_c & \text{if } N_{tr} = 0, \end{cases}$$

$$\sigma_\Delta^2 := \mathrm{Var}\left(V_{\pi_t-\pi_0}^{post}\right) = \begin{cases} \sum_{c=1}^C w_c^2 \left[\mu_c^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) + \frac{\sigma_c^2}{k_c^2} \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \right. \\ \qquad \left. + \frac{\mu_c \sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\left(1 - 2\Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) - \frac{\sigma_c^2}{k_c^2} \cdot \phi^2\left(\frac{\mu_c k_c}{\sigma_c}\right)\right] & \text{if } N_{tr} > 0 \\ 0 & \text{if } N_{tr} = 0, \end{cases}$$

and

$$\mu_U := \mathbb{E}\left[U\right] = \begin{cases} \sum_{c=1}^{C} w_c^2 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}} \right) \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) & \text{if } N_{tr} > 0 \\ \sum_{c=1}^{C} w_c^2 \cdot I\left(\mu_c \geq 0\right) \cdot \left(\sigma_c^2 + \frac{4s_c^2}{w_c N_{test}}\right) & \text{if } N_{tr} = 0, \end{cases}$$

$$\sigma_U^2 := \mathrm{Var}\left(U\right) = \begin{cases} \sum_{c=1}^{C} w_c^4 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}} \right)^2 \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) \left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) & \text{if } N_{tr} > 0 \\ 0 & \text{if } N_{tr} = 0, \end{cases}$$

where we have used that $\mathbb{P}\left(\mu_c^{post} \geq 0\right) = \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)$.

To approximate $\mathcal{M}$, we apply the first-order Taylor series approximations $\ln x \approx x - 1$ around $x = 1$ and $\exp(x) \approx 1 + x$ around $x = 0$:

$$\begin{aligned} \mathcal{M} &= \exp\left(\ln \frac{V_{\pi_t - \pi_0}^{post}}{\sqrt{U}}\right) \\ &= \exp\left(\ln V_{\pi_t - \pi_0}^{post} - \frac{1}{2}\ln U\right) \\ &= \exp\left(\ln \frac{\mu_\Delta}{\sqrt{\mu_U}} + \ln\left(1 + \frac{V_{\pi_t - \pi_0}^{post} - \mu_\Delta}{\mu_\Delta}\right) - \frac{1}{2}\ln\left(1 + \frac{U - \mu_U}{\mu_U}\right)\right) \\ &\approx \frac{\mu_\Delta}{\sqrt{\mu_U}} \exp\left(\frac{V_{\pi_t - \pi_0}^{post} - \mu_\Delta}{\mu_\Delta} - \frac{U - \mu_U}{2\mu_U}\right) \\ &\approx \frac{\mu_\Delta}{\sqrt{\mu_U}} \left(1 + \frac{V_{\pi_t - \pi_0}^{post} - \mu_\Delta}{\mu_\Delta} - \frac{U - \mu_U}{2\mu_U}\right) \end{aligned}$$

We employ the Lyapunov CLT to approximate the distribution of $\frac{V_{\pi_t - \pi_0}^{post} - \mu_\Delta}{\mu_\Delta} - \frac{U - \mu_U}{2\mu_U}$ by a normal distribution:

$$\frac{V_{\pi_t - \pi_0}^{post} - \mu_\Delta}{\mu_\Delta} - \frac{U - \mu_U}{2\mu_U} \overset{approx}{\sim} \mathcal{N}\left(0, \sigma_{\Delta U}^2\right), \tag{92}$$

where

$$\begin{aligned} \sigma_{\Delta U}^2 &= \mathrm{Var}\left(\frac{V_{\pi_t - \pi_0}^{post} - \mu_\Delta}{\mu_\Delta} - \frac{U - \mu_U}{2\mu_U}\right) \\ &= \frac{\sigma_\Delta^2}{\mu_\Delta^2} + \frac{\sigma_U^2}{4\mu_U^2} - \frac{1}{\mu_\Delta \mu_U} \mathrm{Cov}(V_{\pi_t - \pi_0}^{post}, U) \\ &= \frac{\sigma_\Delta^2}{\mu_\Delta^2} + \frac{\sigma_U^2}{4\mu_U^2} - \frac{1}{\mu_\Delta \mu_U} \sum_c w_c^3 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right) \mathrm{Cov}\left(I\left(\mu_c^{post} \geq 0\right) \cdot \mu_c^{post}, I\left(\mu_c^{post} \geq 0\right)\right) \\ &= \frac{\sigma_\Delta^2}{\mu_\Delta^2} + \frac{\sigma_U^2}{4\mu_U^2} - \frac{1}{\mu_\Delta \mu_U} \sum_c w_c^3 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right) \mathbb{E}\left[I\left(\mu_c^{post} \geq 0\right) \cdot \mu_c^{post}\right] \left(1 - \mathbb{P}\left(\mu_c^{post} \geq 0\right)\right) \\ &= \begin{cases} \frac{\sigma_\Delta^2}{\mu_\Delta^2} + \frac{\sigma_U^2}{4\mu_U^2} - \frac{1}{\mu_\Delta \mu_U} \sum_c w_c^3 \left((\sigma_c^{post})^2 + \frac{4s_c^2}{w_c N_{test}}\right) \left(\mu_c \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) + \frac{\sigma_c}{k_c}\phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) \left(1 - \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)\right) & \text{if } N_{tr} > 0 \\ 0 & \text{if } N_{tr} = 0, \end{cases} \end{aligned}$$

where the fourth step uses that, for random variables X_1, X_2, we have $\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2]$.

Putting this together, we obtain an approximation for the probability $\mathbb{P}(T \geq z_{1-\alpha})$ in (91):

$$\begin{aligned}
\mathbb{P}(T \geq z_{1-\alpha}) &= \mathbb{E}\left[\Phi\left(\mathcal{M} - z_{1-\alpha}\right)\right] \\
&\approx \mathbb{E}\left[\Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}}\left(1 + \sigma_{\Delta U} Z\right) - z_{1-\alpha}\right)\right] \\
&= \mathbb{E}\left[\Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha} + \frac{\mu_\Delta \sigma_{\Delta U}}{\sqrt{\mu_U}} Z\right)\right] \\
&= \int_{-\infty}^{\infty} \Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha} + \frac{\mu_\Delta \sigma_{\Delta U}}{\sqrt{\mu_U}} x\right)\phi(x)dx \\
&= \Phi\left(\frac{\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}}{\sqrt{1 + \left(\frac{\mu_\Delta}{\sqrt{\mu_U}}\sigma_{\Delta U}\right)^2}}\right)
\end{aligned}$$

where random variable $Z \sim \mathcal{N}(0,1)$ is a standard normal, and where we use the integral

$$\int_{-\infty}^{\infty} \phi(x)\Phi(a + bx)dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

from (Owen, 1980).

We can now approximate the certification constraint (25) of the (α, β) certification problem as

$$\Phi\left(\frac{\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}}{\sqrt{1 + \left(\frac{\mu_\Delta}{\sqrt{\mu_U}}\sigma_{\Delta U}\right)^2}}\right) \geq 1 - \beta. \tag{93}$$

In our empirical application, applying this approximation yields almost exact solutions in the optimization problem, where exact solutions are recovered via simulation. This approximation is also computationally faster than running a simulation to generate many draws of T in order to compute $\mathbb{P}(T \geq z_{1-\alpha})$ numerically.

If we apply a lower-order Taylor series approximation for the $\exp(\cdot)$ function in the derivation above, we can further approximate

$$\mathcal{M} \approx \frac{\mu_\Delta}{\sqrt{\mu_U}} \exp\left(\frac{V^{post}_{\pi_t - \pi_0} - \mu_\Delta}{\mu_\Delta} - \frac{U - \mu_U}{2\mu_U}\right) \approx \frac{\mu_\Delta}{\sqrt{\mu_U}}.$$

Then $\mathbb{P}(T \geq z_{1-\alpha}) = \mathbb{E}\left[\Phi\left(\mathcal{M} - z_{1-\alpha}\right)\right] \approx \Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right)$, and the certification constraint becomes

$$\Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right) \geq 1 - \beta.$$

This leads to an auxiliary optimization problem that admits an explicit closed-form solution $N^{\text{aux}}_{test}(N_{tr})$ for the required testing sample size as a function of the training sample size. We derive this solution in Appendix D.3.3.

D.3.3 Solving the auxiliary optimization problem

We characterize the properties of the auxiliary optimization problem with an approximate constraint:

$$\min_{N_{tr},N_{test}} \quad N_{tr} + N_{test} \tag{94}$$

$$\text{s.t.} \quad \Phi\left(\frac{\mu_\Delta}{\sqrt{\mu_U}} - z_{1-\alpha}\right) \geq 1-\beta \tag{95}$$

$$N_{tr} \geq 0, N_{test} > 0. \tag{96}$$

where μ_Δ and μ_U are provided in Appendix D.3.2.

The LHS of (95) is continuous and monotonically increasing in N_{test}, because μ_Δ does not depend on N_{test}, and μ_U is decreasing in N_{test}. Continuity implies that, at any optimal solution, constraint (95) is binding. Otherwise, by decreasing N_{test}, we could obtain a smaller sample size $N_{tr} + N_{test}$, while still satisfying the constraints of the optimization problem.

Recall that we consider the practically interesting case $\alpha + \beta < 1$, which is equivalent to $Z_{\alpha,\beta} := z_{1-\alpha} + z_{1-\beta} > 0$. Under this condition, we can restate the inequality constraint (95) as an equality constraint:

$$\mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \mu_U = 0. \tag{97}$$

For a given N_{tr}, the optimization constraint (95) may be infeasible even for large values of N_{test}. By increasing the testing sample N_{test}, we can reduce the idiosyncratic noise, but the uncertainty about the treatment effects $(\sigma_c^{post})^2$ does not depend on N_{test}. Formally, in the limit of large N_{test} we have

$$\lim_{N_{test}\to\infty} \left(\mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \mu_U\right) = \begin{cases} \mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot (\sigma_c^{post})^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) & \text{if } N_{tr} > 0 \quad (98a) \\ \mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot I\left(\mu_c \geq 0\right) \cdot \sigma_c^2 & \text{if } N_{tr} = 0. \quad (98b) \end{cases}$$

The constraint (95) is feasible for a given N_{tr} if and only if the following condition holds:

$$\begin{cases} \mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot (\sigma_c^{post})^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right) > 0 & \text{if } N_{tr} > 0 \quad (99a) \\ \mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot I\left(\mu_c \geq 0\right) \cdot \sigma_c^2 > 0 & \text{if } N_{tr} = 0. \quad (99b) \end{cases}$$

When the constraint is feasible, we can write the required testing sample size $N_{test}^{\text{aux}}(N_{tr})$ as an explicit function of the training sample size N_{tr}:

$$N_{test}^{\text{aux}}(N_{tr}) = \begin{cases} \dfrac{4Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c \cdot s_c^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)}{\mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot (\sigma_c^{post})^2 \cdot \Phi\left(\frac{\mu_c k_c}{\sigma_c}\right)} & \text{if } N_{tr} > 0 \quad (100a) \\ \dfrac{4Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c \cdot s_c^2 \cdot I\left(\mu_c \geq 0\right)}{\mu_\Delta^2 - Z_{\alpha,\beta}^2 \cdot \sum_{c=1}^{C} w_c^2 \cdot I\left(\mu_c \geq 0\right) \cdot \sigma_c^2} & \text{if } N_{tr} = 0 \quad (100b) \end{cases}$$

Similar to Appendix D.2, even though the approximate optimization problem may not be feasible for a *given* N_{tr}, there always exists a feasible solution if N_{tr} and N_{test} are both allowed to be large enough.

D.4 Proof of Proposition 3

Proof. Claim (i): We set $f = \mathbb{P}(T \geq z_{1-\alpha}) - (1-\beta)$ and note that f is increasing in N_{test}. We first show by contradiction that

$$\left.\frac{\partial f}{\partial N_{tr}}\right|_{(N_{tr}^*, N_{test}^*)} = \left.\frac{\partial f}{\partial N_{test}}\right|_{(N_{tr}^*, N_{test}^*)}. \tag{101}$$

Assume that $\left.\frac{\partial f}{\partial N_{tr}}\right|_{(N_{tr}^*, N_{test}^*)} > \left.\frac{\partial f}{\partial N_{test}}\right|_{(N_{tr}^*, N_{test}^*)}$. Then a marginal decrease in N_{test} can be offset by a smaller marginal increase in N_{tr}. This would lead to an improvement in the objective, while still satisfying all the constraints, which is a contradiction. Similarly, assume that $\left.\frac{\partial f}{\partial N_{tr}}\right|_{(N_{tr}^*, N_{test}^*)} < \left.\frac{\partial f}{\partial N_{test}}\right|_{(N_{tr}^*, N_{test}^*)}$. Then a marginal decrease in N_{tr} can be offset by a smaller marginal increase in N_{test}. This would lead to an improvement in the objective, while still satisfying all the constraints, which is a contradiction. Equation (101) follows.

Because f is continuous in N_{test}, it follows that at any optimal solution the constraint (25) is binding. Otherwise, by decreasing N_{test}, we could obtain a smaller sample size $N_{tr} + N_{test}$, while still satisfying the constraints of the optimization problem. Therefore at optimality, $f = 0$. We can then calculate the derivative of the implicit functions as follows:

$$\frac{dN_{test}}{dN_{tr}} = -\frac{\frac{\partial f}{\partial N_{tr}}}{\frac{\partial f}{\partial N_{test}}}$$

and

$$\frac{dN_{tr}}{dN_{test}} = -\frac{\frac{\partial f}{\partial N_{test}}}{\frac{\partial f}{\partial N_{tr}}}.$$

Substituting using (101) we obtain $\left.\frac{dN_{test}}{dN_{tr}}\right|_{(N_{tr}^*, N_{test}^*)} = -1$ and $\left.\frac{dN_{tr}}{dN_{test}}\right|_{(N_{tr}^*, N_{test}^*)} = -1$.

Claim (ii): We again establish a contradiction. By claim (i), at the optimal solution we have $\left.\frac{dN_{test}}{dN_{tr}}\right|_{(N_{tr}^*, N_{test}^*)} = -1$, and so a marginal increase in N_{tr} yields an equivalent marginal decrease in N_{test}, in order to ensure that $f = 0$. However, if $\left.\frac{d^2 N_{test}}{dN_{tr}^2}\right|_{(N_{tr}^*, N_{test}^*)} < 0$, the next marginal increase in N_{tr} would require a larger marginal decrease in N_{test} to ensure that the binding constraint remains binding. This would mean that there is an alternative solution that both satisfies the constraints and for which the objective $N_{tr} + N_{test}$ is smaller, which is a contradiction. Therefore we have that $\left.\frac{d^2 N_{test}}{dN_{tr}^2}\right|_{(N_{tr}^*, N_{test}^*)} \geq 0$. The reasoning for showing that $\left.\frac{d^2 N_{tr}}{dN_{test}^2}\right|_{(N_{tr}^*, N_{test}^*)} \geq 0$ is analogous. □

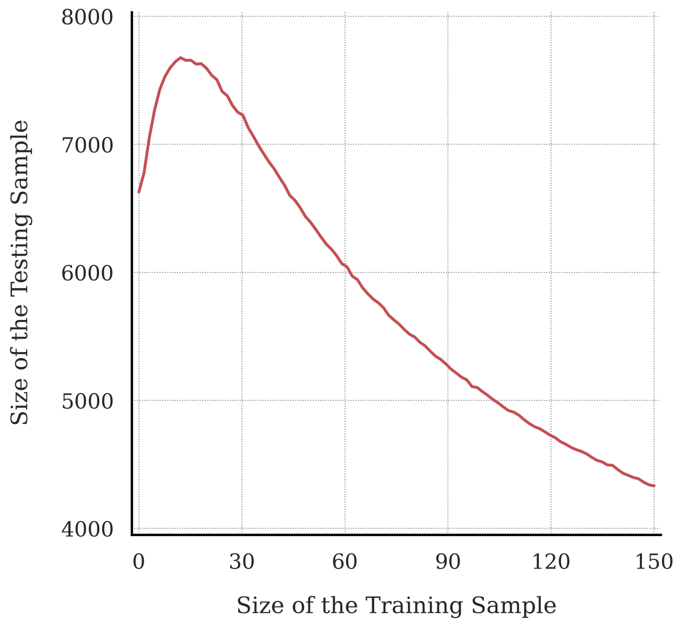
D.5 Non-monotonic $N_{test}^*(N_{tr})$

The relationship between the optimal size of the testing sample and the size of the training data $N_{test}^*(N_{tr})$ derived in Section 4 can be non-monotonic. We provide an empirical example in Figure 5. Figure 5 is similar to Figure 2 in the main text, but it is estimated using different input parameters μ_c, σ_c^2, s_c^2, and w_c. For reproducibility, we generate the input parameters using the following code:

```python
import numpy as np
np.random.seed(2)
C = 30
mu_c = np.random.normal(0,1,C)
sigma_c = np.abs(np.random.normal(0,1,C))
s_c = np.ones(C)*5
w_c = np.ones(C)/C
```

We caution that the figure zooms in on a parameter region with small N_{tr} to illustrate the non-monotonicity of $N^*_{test}(N_{tr})$. The optimal solution is $(N^*_{tr}, N^*_{test}) = (931, 2440)$, which is well beyond the region shown in the figure. In our simulated and empirical analyses, we have only observed non-monotonicity for very small values of N_{tr}.

Figure 5: Non-monotonic relationship $N^*_{test}(N_{tr})$



E Empirical Analysis

E.1 Input parameters in the empirical application

Table 4 provides all input parameters used in our empirical analysis in Section 5.

E.2 Approximation for (α, β)-policy certification

Figure 6 illustrates the quality of the approximation in (27)-(29). The left panel replicates Figure 2 from the main text. The (red) solid line captures the optimal size of the testing data N_{test} as a function of the size of the training data N_{tr} estimated using the simulation-based approach. The (blue) dashed line captures the relationship $N^*_{test}(N_{tr})$ from the analytical approximation-based solution as implemented by Algorithm 2. Similarly, the right panel replicates Figure 3 from the main text. The (red) solid line captures the power of the statistical test as a function of the share of the training data, as estimated using the simulation-based approach, and the (blue) dashed line shows the analytical approximation-based estimates. In both cases, our proposed approximation closely matches the simulation-based estimates.

Figure 6: Approximation for (α, β) Policy Certification

(a) $N^*_{test}(N_{tr})$ (b) Statistical Power

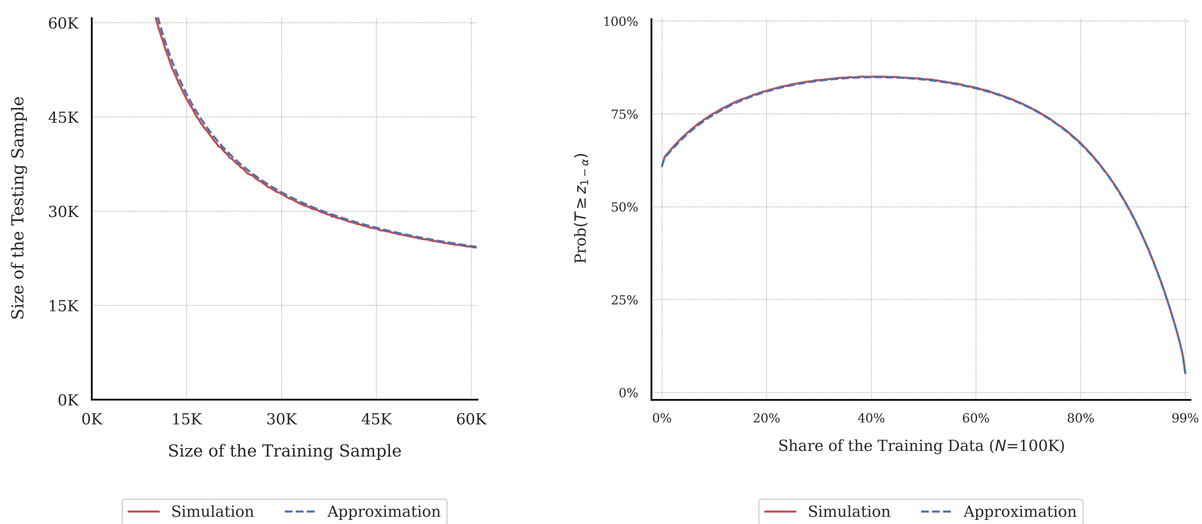


Table 4: Input Parameters in Section 5

Segment ID	Past Spending Bracket	w_c	μ_c	σ_c	s_c
0	Less than or equal to \$0	3.54%	−2.38	6.20	114.34
1	$(\$0, \$200]$	1.62%	16.23	10.69	118.07
2	$(\$200, \$400]$	1.91%	−9.20	10.15	150.51
3	$(\$400, \$600]$	9.46%	−0.68	3.09	91.86
4	$(\$600, \$800]$	13.82%	−4.93	2.80	110.74
5	$(\$800, \$1,000]$	9.89%	−5.20	3.94	133.29
6	$(\$1,000, \$1,200]$	8.06%	−4.53	4.90	141.20
7	$(\$1,200, \$1,400]$	6.66%	−4.32	5.96	154.49
8	$(\$1,400, \$1,600]$	5.57%	−6.66	6.62	166.61
9	$(\$1,600, \$1,800]$	4.27%	7.96	7.64	157.29
10	$(\$1,800, \$2,000]$	3.89%	−4.79	9.34	195.75
11	$(\$2,000, \$2,200]$	3.12%	13.71	12.83	202.41
12	$(\$2,200, \$2,400]$	2.71%	19.51	12.80	195.29
13	$(\$2,400, \$2,600]$	2.36%	11.90	13.43	182.20
14	$(\$2,600, \$2,800]$	2.18%	−1.09	16.60	248.29
15	$(\$2,800, \$3,000]$	1.88%	−25.92	16.60	240.58
16	$(\$3,000, \$3,500]$	3.84%	0.87	12.87	241.13
17	$(\$3,500, \$4,000]$	2.86%	−3.02	15.06	254.05
18	$(\$4,000, \$4,500]$	2.25%	−23.17	18.13	286.34
19	$(\$4,500, \$5,000]$	1.75%	−8.50	21.16	282.36
20	$(\$5,000, \$5,500]$	1.65%	28.68	24.96	316.10
21	$(\$5,500, \$6,000]$	1.30%	9.71	29.00	324.56
22	$(\$6,000, \$6,500]$	1.04%	−10.22	33.62	350.19
23	$(\$6,500, \$7,000]$	0.92%	16.60	31.39	311.15
24	$(\$7,000, \$7,500]$	0.78%	1.63	40.05	366.66
25	$(\$7,500, \$8,000]$	0.72%	−8.80	42.67	358.83
26	Over \$8,000	1.95%	8.44	27.77	384.38