INSEAD
The Business School for the World®

Working Paper Series

2017/63/DSC

Efficiently Evaluating Targeting Policies Using Field Experiments

Duncan Simester
MIT, simester@mit.edu

Artem Timoshenko
MIT, atimoshe@mit.edu

Spyros Zoumpoulis
INSEAD, Spyros.zoumpoulis@insead.edu

August 16, 2017

Firms often want to target different customers with different actions, and the marketing literature contains many models designed to optimize targeting policies. While these models are sometimes validated using simulations or historical data, the gold-standard approach to comparing alternative targeting policies is to implement the alternative policies on randomly selected groups of customers and then compare the aggregate outcomes. We show that we can improve the efficiency of these comparisons by using the same experiments but changing the analysis. Instead of comparing the aggregate outcomes, it is more efficient to compare the outcomes within each customer segment, where the segments are defined using the candidate policies. Differencing within segment reduces variation introduced by between-segment differences, and we can then aggregate the within-segment differences across segments. Our key contribution is to show that the choice of the segments is important. By segmenting using the policies' recommended actions, we know that for some segments the difference in outcomes is zero by construction. This can greatly improve efficiency by eliminating random variation in the comparison of these segments. We illustrate these benefits theoretically and empirically, using data from a recent field experiment. We also show how this segmentation approach extends to an alternative experimental design and identify limitations that may hinder implementation of the proposed approach.

Keywords: targeting, field experiments, segmentation, causal inference, potential outcomes, standard errors
JEL Classification: C12, C93

1 Introduction

Targeting policies match different actions to different populations. For example, farmers may want to give different fields different fertilizers, doctors give different patients different treatments, NGOs offer different interventions to different students, or retailers target different customers with different promotions. The gold-standard approach to comparing alternative targeting policies is to implement the alternative policies on randomly selected groups of participants and then compare the resulting performance. For example, Skiera and Nabout (2013) propose a model that targets different search engine keywords with different bids by an advertiser. They test their model using a field experiment in which bids for twenty keywords were submitted using either the current policy or the proposed model. Similarly, Mantrala et al. (2006) proposed a model for setting different prices for different automobile parts. They validated their model using a field experiment in which 200 stores were randomly assigned to the proposed policy, and 300 stores were randomly assigned to the current policy. Simester et al. (2006) test a policy for targeting customers with a sequence of catalogs. A total of 60,000 customers were randomly assigned to receive catalogs using either the proposed model or the current policy.[1]

It is common to analyze these experiments by calculating the aggregate outcome for each policy and then comparing these aggregate outcomes across policies. In this paper, we show that the efficiency of these comparisons can be improved by changing the analysis, without changing the experiments. Rather than comparing the aggregate outcome from each targeting policy, we propose to first segment the customers and then compare the outcome within each segment. This reduces variation introduced by between-segment differences. We show that the segments can be designed so that for some segments the difference in outcomes between policies is zero by construction. This eliminates variation introduced by random noise within these segments.

These changes provide a more efficient comparison than simply comparing the aggregate outcomes. More specifically, when holding the total number of participants fixed, one can obtain a more accurate estimate of the difference in the performance of any set of policies compared to comparing aggregate outcomes. Alternatively, one can obtain the same level of precision using fewer participants. The efficiency improvement can be large, particularly when comparing targeting policies that are more similar. We show theoretically that our proposed method strictly improves the standard error of the estimator of the performance of a single policy, as well as the standard error of the estimator of the difference between policies. We also show empirically that our proposed method reduces the standard error by more than 50% when evaluating policies and their differences, using data from an actual large-scale field experiment (4.1 million households) run to investigate how a retailer should target prospective customers with promotions (Simester et al., 2017). This reduction has substantive importance in our example: the difference between policies becomes statistically significant, whereas the difference was not significant when using the standard approach.

[1] Similar recent examples include: Belloni et al. (2012); Lu et al. (2016); Neslin et al. (2009); Simester et al. (2017); Urban et al. (2014). We use data from the Simester et al. (2017) study in this paper.

The design of the segments is important. We propose segmenting customers using the actions recommended by the candidate targeting policies. Targeting policies represent a mapping from customer covariates to recommended actions and can be thought of as candidate methods for summarizing the information provided by the covariates. The mapping to actions provides a segmentation of customers. With two or more targeting policies, the intersection of their recommended actions provides an even more fine-grained segmentation of customers.

These segments not only provide a natural basis for decomposing the comparison of the methods, but also offer two important efficiency advantages that lie at the core of our paper. Consider the segments in which the candidate policies recommend the same action. First, when evaluating a single policy, we should use all of the customers in these segments, even if they were randomly assigned to a different policy. Pooling across customers assigned to different policies will increase the sample size used to measure the outcome and improve efficiency. Second, when comparing policies within this segment, we know the *true* difference in the performance of the policies is precisely zero, because the policies are identical for this segment of customers. Because of pooling customers across the policies in these segments, the *observed* difference between the policies is also zero. In contrast, under the standard approach, random error would have been introduced, and the observed outcomes in the two conditions would have been different, although we know the true outcomes should be identical in segments where the policies recommend the same action.

Our results also suggest a further implication. An experiment provides no information about the difference in the performance of two policies in segments in which the two policies recommend the same action. Therefore, if we are only interested in comparing the relative performance of candidate targeting policies (rather than evaluating the absolute performance of any single policy), then we can omit these segments of customers entirely from the experimental design. We can thus save implementation costs with no loss of information about the relative performance of the policies.

Recall that the traditional approach to comparing targeting policies is to use one experimental condition for each policy. We also consider an alternative experimental design that uses one experimental condition for each possible *action*. This design allows evaluation of any targeting policy, and so any new policy can be evaluated using the existing data, without being implemented. Our proposed segmentation approach can also yield its benefits under this alternative experimental design. We finally identify limitations that may hinder implementation of the proposed approach.

Related Literature

The efficiency advantage of aggregating within-segment differences was recently also proposed by Imbens and Rubin (2015) and Athey and Imbens (2017). They recognize that the randomization justifies comparison within segments. However, they contemplate segmenting participants by covariates, rather than segmenting according to the actions recommended by alternative targeting policies, as we propose. This distinction is important; the efficiency improvements that we describe depend critically upon how customers are segmented. By segmenting using the recommended actions from the candidate policies, we identify segments of customers for whom the policies recommend the same actions.

More generally, the paper contributes to a recent stream of research investigating how firms can use field experiments to help optimize their marketing decisions. For example, Barajas et al. (2016) investigate how to use field experiments to attribute the response to online display advertisements. They propose two new experimental designs that allow firms to estimate the overall campaign attribution without using placebo advertisements. These designs also allow the firm to disentangle the effect of the campaign from potential confounds. Li et al. (2015) study the feasibility of using field experiments to make category pricing decisions when items could be complements or substitutes. They ask how many experiments are required as the number of products in the category grows. They show that firms may be able to obtain meaningful estimates using a practically feasible number of experiments, even in categories with a large number of products. Other studies have investigated the required size of field experiments. For example, Lewis and Rao (2015) demonstrate that in digital advertising settings where the effect sizes are small and the response measures are highly stochastic, very large field experiments may be required to generate information.

The recent development economics literature also includes proposals to improve experimental design. Congdon et al. (2017) propose conducting "mechanism experiments" designed to help understand the underlying mechanism rather than the outcome of a specific policy. They argue that these mechanism experiments may be useful both to screen potential policies, and to suggest policy improvements.

Our findings are also related to work on optimal experimental design. This includes an extensive literature studying how to design conjoint experiments (see Louviere et al., 2004 for a review). Experimental design has also been interpreted as a multi-armed bandit problem (see for example Mersereau et al., 2009); pulling arms of a multi-armed bandit samples the outcomes from different marketing actions. The optimal experimental design introduces a trade-off between exploration and exploitation.

2 An Illustrative Example

Before presenting a formal analysis of the efficiency advantages, we first illustrate these advantages using an example. Consider a retailer who wants to target customers with promotional offers. Consider two targeting policies that both recommend sending promotion a to male customers. The policies differ in their recommendations for female customers. Both policies recommending sending either promotion b or promotion c to female customers, but which female customers receive each promotion varies across the policies. Randomly assigning half of the customers to each policy will mean that half of the males will be assigned to each policy, and all of the males will receive promotion a.

The analysis approach we propose segments the sample population according to the actions recommended by the two policies. Male customers are all in one segment because both policies recommended promotion a for male customers. Female customers are divided into different segments according to the promotions recommended by each policy. This example allows us to illustrate each of our arguments.

First, comparing the difference in outcomes between policies within a segment allows us to eliminate between-segment differences. For example, if males respond on average x% of the time but females respond y% of the time, then calculating an aggregate outcome for each policy will include variation introduced by gender. As a result, when comparing the average outcome between policies, the standard error of the difference will also include variance introduced by gender differences. We eliminate this variation if instead we compare the difference in outcomes within a segment, and then aggregate these within-segment differences across segments. The differencing within segment ensures we only compare males with males, and females with females. It is this insight that led Imbens and Rubin (2015) and Athey and Imbens (2017) to also recommend analyzing randomized experiments by stratum, i.e., by segment.

Second, because both policies recommend the same action for males, when measuring the absolute outcome of either policy for males, we should pool across all of the males, irrespective of which of the two policies they were assigned to. In our illustrative example, this will double the sample size used to measure the outcome for male customers.

Third, because we segmented by the recommended actions and both policies recommend the same actions for male customers, the true difference in the performance of the policies for males is precisely zero, by construction. This provides an opportunity to improve efficiency by eliminating the random errors introduced when calculating the aggregate outcome and comparing these aggregate outcomes. This benefit will be particularly important when comparing two policies that are more similar. The more similar the policies, the larger the segments of customers for which the two policies recommend the same action, and for which the true difference in the performance of the two policies is precisely zero.

This example also illustrates the opportunity to further improve efficiency through design of the experiment. An experiment provides no information about the difference in the performance of the two policies for males. If we are only interested in the relative performance of the two policies, there is no need to implement either policy on the male customers; we know that the policies are not different for male customers. We can omit male customers with no loss of information about relative performance.

Notice that if we omit male customers, we can no longer calculate the absolute outcome across all customers for any single policy. It will often be desirable not just to measure the relative performance of two policies, but also to measure each policy's absolute performance. For this reason, instead of omitting male customers, it may be preferable to simply under-sample male customers.

The Importance of Segmenting by Recommended Actions

The simplicity of our male/female example may give the misleading impression that we recommend segmenting customers based upon gender (which is a covariate). Instead, we propose segmenting based upon the actions recommended by the policies, which in this example coincides with gender for the males (although not for the females). This distinction is important as many of the efficiency gains are only achieved because we segmented customers using the actions recommended by the candidate policies.

We can demonstrate this point by modifying our illustrative example. Imagine that we segmented customers by age instead of gender, and that each age segment includes a combination of both genders. Now the segments no longer contain a single recommended action for each policy.

Analyzing by segment still offers the first advantage of removing between-segment differences. If there are differences in the probability of response by young customers and old customers, calculating the difference in outcomes within each age segment and then aggregating the differences across age segments will help to remove this source of variation. This will be more efficient than simply calculating the aggregate outcomes and comparing these aggregate outcomes.

However, the other efficiency gains are no longer available. In particular, because each age segment contains both genders, and the policies recommend different actions for females, there will be no age segment for which the two policies recommend the same action. Therefore, pooling across conditions to increase the sample size is not possible, and relatedly there are no segments in which we avoid the introduction of random error when calculating the difference between the two policies. Segmenting by the recommended actions solves this issue. Because each policy recommends an action for each customer, it is straightforward to segment customers according to these recommended actions. Once the recommended actions are known, the segmentation can be accomplished without reference to the covariates. Moreover, this segmentation immediately identifies any customers for which the policies recommend the same action.

We formalize these arguments next by demonstrating how to estimate the outcomes and their variance using both the standard approach and the proposed approach.

3 A Formal Analysis of Efficiency

We contemplate an experimental design in which each customer i is assigned to one of T policies, or not assigned to any of the policies, by randomly selecting customers from a super-population of size N. Letting W_i be a random variable indicating what policy customer i is assigned to, we have $W_i \in \{P_1, P_2, \ldots, P_T, \emptyset\}$, where $P_1, \ldots, P_T$ denote the policies and $W_i = \emptyset$ denotes the outcome when customer i is not sampled to any of the policies (and therefore does not participate in the experiment). This random selection and assignment to the policies is the source of variation in the system. We use $N_{P_1}, \ldots, N_{P_T}, N_\emptyset$ to denote the number of customers assigned to each of $\{P_1, P_2, \ldots, P_T, \emptyset\}$, so that $N_{P_1} + \ldots + N_{P_T} + N_\emptyset = N$.

We assume a completely randomized experiment:

Assumption 1 (Random sampling and assignment without replacement)**.** Conditional on $N_{P_1}, \ldots, N_{P_T}, N_\emptyset$, the vector $\mathbf{W}$ has multinomial distribution, with

$$\Pr(\mathbf{W} = \mathbf{w} \mid N_{P_1}, \ldots, N_{P_T}, N_\emptyset) = \begin{cases} \binom{N}{N_{P_1}, \ldots, N_{P_T}, N_\emptyset}^{-1} & \text{for all } \mathbf{w} \text{ with} \\ \quad \sum_{i=1}^N \mathbb{1}_{w_i = P_1} = N_{P_1}, \ldots, \sum_{i=1}^N \mathbb{1}_{w_i = P_T} = N_{P_T}, \sum_{i=1}^N \mathbb{1}_{w_i = \emptyset} = N_\emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Customers are described by a set of covariates. We think of the policies as mappings from the space of covariates to a selection of an action. For example, in the context of prospecting new customers for a retailer, the actions are the types of promotional offers the retailer considers mailing to different households according to their covariates, including the control treatment of no mail. Suppressing the covariates from our notation, we denote by $P_\ell(i)$ the action that policy P_ℓ assigns to user i.

We consider a finite set of possible actions. We follow the potential outcomes framework: for each customer i and action s, we define the outcome $Y_i(s)$ that would have occurred, had customer i been treated with action s. We define $Y_i(s)$ regardless of whether customer i is actually treated with action s or not. Only one of the potential outcomes is realized and observed for customer i, which we denote by Y_i^{obs}. We do not observe what would have happened had customer i been exposed to other actions. The uncertainty therefore does not only come from random sampling from the super-population, but also from the unobserved potential outcomes, in the spirit of Abadie et al. (2014).

We first describe how to evaluate a single policy under both the standard and proposed approaches. We then describe how to compare two policies. We summarize the resulting mean and variance calculations under each approach in Table 1.

Evaluating a Single Policy Using the Standard Approach

We label the policy of interest as P_1. Our target (estimand) when evaluating policy P_1 is the population-level measure

$$y_{P_1} = \frac{1}{N} \sum_{i=1}^N Y_i \left(P_1(i)\right).$$

We do not observe the outcomes for the entire population under each policy. Therefore, the standard approach is to estimate y_{P_1} by

$$\hat{y}_{P_1} = \frac{1}{N_{P_1}} \sum_{i=1}^N \mathbb{1}_{W_i = P_1} Y_i^{obs}.$$

It is straightforward to establish that $\hat{y}_{P_1}$ is an unbiased estimator of y_{P_1}. As we show in the Appendix, the variance of this estimator is given by

$$\mathrm{Var}_W(\hat{y}_{P_1}) = \frac{S^2_{P_1,N}}{N_{P_1}}\left(1 - \frac{N_{P_1}}{N}\right),$$

where

$$S^2_{P_1,N} = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i\left(P_1(i)\right) - \frac{1}{N}\sum_{j=1}^{N}Y_j\left(P_1(j)\right)\right)^2. \tag{1}$$

We estimate $\mathrm{Var}(\hat{y}_{P_1})$ by

$$\widehat{\mathrm{Var}}_W(\hat{y}_{P_1}) = \frac{s^2_{P_1,N}}{N_{P_1}}\left(1 - \frac{N_{P_1}}{N}\right),$$

where

$$s^2_{P_1,N} = \frac{1}{N_{P_1}-1}\sum_{i:W_i=P_1}\left(Y_i^{obs} - \hat{y}_{P_1}\right)^2 \tag{2}$$

is an unbiased estimator for $S^2_{P_1,N}$ (Imbens and Rubin, 2015, Chapter 6.5, Appendix A).

We next discuss how to estimate y_{P_1} under the proposed approach.

Evaluating a Single Policy Using the Proposed Approach

Even though we are evaluating a single policy, the proposed approach recommends first segmenting customers using the recommended actions from at least two policies. For this illustration we will continue to evaluate policy P_1, but will use the recommended actions from both policies P_1 and P_2. We must first introduce some additional notation to identify segments of customers, where the segments are constructed using the recommended actions from each policy.

Define $g_{P_1:s,P_2:t}$ as the segment of customers in the super-population to whom policy P_1 would assign action s and policy P_2 would assign action t. Moreover, $N_{P_1:s,P_2:t}$ is the number of customers in segment $g_{P_1:s,P_2:t}$, and $N^{P_1}_{P_1:s,P_2:t}$ is the number of customers in segment $g_{P_1:s,P_2:t}$ that are randomly assigned to receive policy P_1.

Our goal is unchanged. We are after the population-level measure

$$y_{P_1} = \frac{1}{N}\sum_{i=1}^{N}Y_i\left(P_1(i)\right),$$

which can also be written as

$$y_{P_1} = \frac{\sum_{s,t}\sum_{i\in g_{P_1:s,P_2:t}}Y_i(s)}{\sum_{s,t}N_{P_1:s,P_2:t}} = \frac{\sum_{s,t}N_{P_1:s,P_2:t}\cdot y_{\underline{P_1}:s,P_2:t}}{N},$$

where $y_{\underline{P_1}:s,P_2:t} = \frac{\sum_{i\in g_{P_1:s,P_2:t}}Y_i(s)}{N_{P_1:s,P_2:t}}$.

Under the proposed approach we estimate y_{P_1} by calculating the outcome in each segment and then aggregating across segments:

$$\hat{y}_{P_1} = \frac{\sum_{s,t} N_{P_1:s,P_2:t} \cdot \hat{y}_{\underline{P_1}:s,P_2:t}}{N},$$

with

$$\hat{y}_{\underline{P_1}:s,P_2:t} = \begin{cases} \frac{\sum_{i \in g_{P_1:s,P_2:t}} \mathbb{1}_{W_i=P_1} Y_i^{obs}}{N^{P_1}_{P_1:s,P_2:t}} & \text{if } s \neq t, \\ \frac{\sum_{i \in g_{P_1:s,P_2:s}} (\mathbb{1}_{W_i=P_1} + \mathbb{1}_{W_i=P_2}) Y_i^{obs}}{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s}} & \text{if } s = t, \end{cases}$$

where the first case is for segments for which the two policies recommend different actions, while the second case is for segments for which the two policies recommend the same action.

Pooling customers across the two conditions in segments for which the two policies recommend the same action, irrespective of which of the two policies they were assigned to, is one of the core ideas of our paper. The efficiency gain of pooling is twofold. First, the sample size used to measure the outcome for customers for whom the two policies would recommend the same action increases. Second, as we illustrate later on in this section, the difference in the performance of the two policies in the segments for which the two policies recommend the same action is estimated to be zero, as it truly is. This eliminates the random error in these segments when comparing the performance of the two policies.

It is straightforward to establish that $\hat{y}_{P_1}$ is an unbiased estimator of y_{P_1}. (We relegate the proof to the Appendix.) We can also write expressions for the variance of the estimator $\hat{y}_{P_1}$ within each segment. To do so we again distinguish between segments in which the two policies recommend the same or different actions.

For segments for which the two policies recommend different actions (i.e., $s \neq t$), we explain in the Appendix that the variance is given by

$$\text{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right) = \frac{S^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} \left(1 - \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}}\right),$$

where

$$S^2_{P_1,g_{P_1:s,P_2:t}} = \frac{1}{N_{P_1:s,P_2:t} - 1} \sum_{i \in g_{P_1:s,P_2:t}} \left(Y_i(s) - \frac{1}{N_{P_1:s,P_2:t}} \sum_{j \in g_{P_1:s,P_2:t}} Y_j(s)\right)^2. \quad (3)$$

We estimate $\text{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right)$ by

$$\widehat{\text{Var}}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right) = \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} \left(1 - \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}}\right),$$

where

$$s^2_{P_1, g_{P_1:s,P_2:t}} = \frac{1}{N^{P_1}_{P_1:s,P_2:t} - 1} \sum_{i \in g_{P_1:s,P_2:t}, W_i = P_1} \left(Y_i^{obs} - \hat{y}_{\underline{P_1}:s,P_2:t} \right)^2 \tag{4}$$

is an unbiased estimator for $S^2_{g_{P_1:s,P_2:t}}$ (Imbens and Rubin, 2015, Chapter 9.5).

We next consider segments in which the two policies recommend the same action (i.e., $s = t$). As we show in the Appendix, the variance in these segments is given by

$$\mathrm{Var}_W \left(\hat{y}_{P_1:s,P_2:s} \right) = \frac{S^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s}} \left(1 - \frac{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}} \right),$$

where

$$S^2_{g_{P_1:s,P_2:s}} = \frac{1}{N_{P_1:s,P_2:s} - 1} \sum_{i \in g_{P_1:s,P_2:s}} \left(Y_i(s) - \frac{1}{N_{P_1:s,P_2:s}} \sum_{j \in g_{P_1:s,P_2:s}} Y_j(s) \right)^2.$$

We estimate $\mathrm{Var}_W \left(\hat{y}_{P_1:s,P_2:s} \right)$ by

$$\widehat{\mathrm{Var}}_W \left(\hat{y}_{P_1:s,P_2:s} \right) = \frac{s^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s}} \left(1 - \frac{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}} \right),$$

where

$$s^2_{g_{P_1:s,P_2:s}} = \frac{1}{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s} - 1} \sum_{i \in g_{P_1:s,P_2:s}, W_i \in \{P_1, P_2\}} \left(Y_i^{obs} - \hat{y}_{P_1:s,P_2:s} \right)^2$$

is an unbiased estimator for $S^2_{g_{P_1:s,P_2:s}}$.

Notice the benefit from pooling across both experimental conditions. In particular, the variance is reduced because the sample includes data from the experimental condition associated with both policy P_1 and policy P_2. This is an efficiency advantage over the standard approach, which does not use any data from the experimental condition associated with policy P_2 when evaluating policy P_1.

Overall, the variance of estimator $\hat{y}_{P_1}$ across all segments is

$$\mathrm{Var}_W(\hat{y}_{P_1}) = \sum_{s,t} \left(\frac{N_{P_1:s,P_2:t}}{N} \right)^2 \mathrm{Var}_W \left(\hat{y}_{\underline{P_1}:s,P_2:t} \right),$$

which we estimate by

$$\widehat{\mathrm{Var}}_W(\hat{y}_{P_1}) = \sum_{s,t} \left(\frac{N_{P_1:s,P_2:t}}{N} \right)^2 \widehat{\mathrm{Var}}_W \left(\hat{y}_{\underline{P_1}:s,P_2:t} \right).$$

Next we describe estimators for estimating the difference in two policies.

Comparing Two Policies Using the Standard Approach

When comparing two policies, the standard approach is to first calculate the mean outcome for each policy and then calculate the difference in these means. For this illustration we will calculate this difference as the outcome for policy P_1 minus the outcome for policy P_2 .

The target estimand is the population-level measure

$$y_{P_1-P_2} = y_{P_1} - y_{P_2} = \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i\left(P_1(i)\right) - Y_i\left(P_2(i)\right)\Big).$$

Because we do not observe the outcomes for the entire population under each policy, we estimate $y_{P_1-P_2}$ using the unbiased estimator

$$\hat{y}_{P_1-P_2} = \hat{y}_{P_1} - \hat{y}_{P_2} = \frac{1}{N_{P_1}}\sum_{i=1}^{N}\mathbb{1}_{W_i=P_1}Y_i^{obs} - \frac{1}{N_{P_2}}\sum_{i=1}^{N}\mathbb{1}_{W_i=P_2}Y_i^{obs}.$$

As shown in the Appendix, the variance of this estimator is given by

$$\mathrm{Var}_W\left(\hat{y}_{P_1-P_2}\right) = \frac{S^2_{P_1,N}}{N_{P_1}} + \frac{S^2_{P_2,N}}{N_{P_2}} - \frac{S^2_{P_1,P_2,N}}{N}, \tag{5}$$

where variances $S^2_{P_1,N}, S^2_{P_2,N}$ are given by Equation (1), and the variance of the customer-level differences between policies is

$$S^2_{P_1,P_2,N} = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(P_1(i)) - Y_i(P_2(i)) - \frac{1}{N}\sum_{j=1}^{N}\Big(Y_j(P_1(j)) - Y_j(P_2(j))\Big)\right)^2.$$

We estimate $S^2_{P_1,N}, S^2_{P_2,N}$ with $s^2_{P_1,N}, s^2_{P_2,N}$ respectively, as defined in Equation (2). The term $S^2_{P_1,P_2,N}$ is in general impossible to estimate empirically because we never observe the outcome of both policies P_1 and P_2 for the same customer. It is common practice to thus approximate the variance $\mathrm{Var}_W\left(\hat{y}_{P_1-P_2}\right)$ using the first two terms and ignoring the third term of Equation (5), with the Neyman estimator (Neyman, 1934)

$$\mathrm{Var}^{\text{Neyman}}_{P_1-P_2} = \frac{s^2_{P_1,N}}{N_{P_1}} + \frac{s^2_{P_2,N}}{N_{P_2}},$$

which is generally upwardly biased.

Comparing Two Policies Using the Proposed Approach

We complete this formal description by showing how to compare two policies under the proposed approach. Our target estimand is unchanged:

$$y_{P_1-P_2} = y_{P_1} - y_{P_2} = \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i\left(P_1(i)\right) - Y_i\left(P_2(i)\right)\Big),$$

which, accounting for the segmentation of the customers according to the recommended actions, can also be written as

$$y_{P_1-P_2} = \frac{\sum_{s,t}\sum_{i\in g_{P_1:s,P_2:t}}\Big(Y_i(s) - Y_i(t)\Big)}{\sum_{s,t} N_{P_1:s,P_2:t}}.$$

One of the main contributions of the paper is to recognize that in segments in which the two policies recommend the same action, the difference in the performance of the two policies is equal to zero. We can thus write

$$y_{P_1-P_2} = \frac{\sum_{\substack{s,t\\ s\neq t}}\sum_{i\in g_{P_1:s,P_2:t}}\Big(Y_i(s) - Y_i(t)\Big)}{\sum_{s,t} N_{P_1:s,P_2:t}} = \frac{\sum_{\substack{s,t\\ s\neq t}} N_{P_1:s,P_2:t}\Big(y_{\underline{P_1}:s,P_2:t} - y_{P_1:s,\underline{P_2}:t}\Big)}{N},$$

where, for $s \neq t$,

$$y_{\underline{P_1}:s,P_2:t} = \frac{\sum_{i\in g_{P_1:s,P_2:t}} Y_i(s)}{N_{P_1:s,P_2:t}}, \qquad y_{P_1:s,\underline{P_2}:t} = \frac{\sum_{i\in g_{P_1:s,P_2:t}} Y_i(t)}{N_{P_1:s,P_2:t}}.$$

We estimate $y_{P_1-P_2}$ with the unbiased estimator

$$\hat{y}_{P_1-P_2} = \frac{\sum_{s\neq t} N_{P_1:s,P_2:t}\Big(\hat{y}_{\underline{P_1}:s,P_2:t} - \hat{y}_{P_1:s,\underline{P_2}:t}\Big)}{N},$$

where, for $s \neq t$,

$$\hat{y}_{\underline{P_1}:s,P_2:t} = \frac{\sum_{i\in g_{P_1:s,P_2:t}} \mathbb{1}_{W_i=P_1} Y_i^{obs}}{N^{P_1}_{P_1:s,P_2:t}}, \qquad \hat{y}_{P_1:s,\underline{P_2}:t} = \frac{\sum_{i\in g_{P_1:s,P_2:t}} \mathbb{1}_{W_i=P_2} Y_i^{obs}}{N^{P_2}_{P_1:s,P_2:t}}.$$

In particular, the *true* difference in the outcome between the two policies is zero in segments for which the two policies recommend the same action. Remember that because of pooling customers across the two conditions in these segments, the *observed* difference in the outcomes between the two policies in these segments is also zero. In contrast, under the standard approach, random error would have been introduced, and the observed outcomes in the two conditions would have been different, although we know the true outcomes should be identical in segments where the policies recommend the same action.

In segments for which the two policies recommend different actions ($s \neq t$), the variance of the difference between the two policies is given by[2]

$$\mathrm{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t} - \hat{y}_{P_1:s,\underline{P_2}:t}\right) = \frac{S^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} + \frac{S^2_{P_2,g_{P_1:s,P_2:t}}}{N^{P_2}_{P_1:s,P_2:t}} - \frac{S^2_{P_1,P_2,g_{P_1:s,P_2:t}}}{N_{P_1:s,P_2:t}},$$

where variances $S^2_{P_1,g_{P_1:s,P_2:t}}, S^2_{P_2,g_{P_1:s,P_2:t}}$ are given by Equation (3), and the variance of the customer-level differences between policies is

$$S^2_{P_1,P_2,g_{P_1:s,P_2:t}} = \frac{1}{N_{P_1:s,P_2:t} - 1} \sum_{i \in g_{P_1:s,P_2:t}} \left(Y_i(s) - Y_i(t) - \frac{1}{N_{P_1:s,P_2:t}} \sum_{j \in g_{P_1:s,P_2:t}} \Big(Y_j(s) - Y_j(t)\Big)\right)^2. \quad (6)$$

We estimate $S^2_{P_1,g_{P_1:s,P_2:t}}, S^2_{P_2,g_{P_1:s,P_2:t}}$ with $s^2_{P_1,g_{P_1:s,P_2:t}}, s^2_{P_2,g_{P_1:s,P_2:t}}$ respectively, as defined in Equation (4). The term $S^2_{P_1,P_2,g_{P_1:s,P_2:t}}$ is in general impossible to estimate empirically because we never observe the outcome of both actions s, t for the same customer. This is the same issue that arises with the standard approach. We again use the Neyman variance estimator

$$\mathrm{Var}^{\mathrm{Neyman}}_{g_{P_1:s,P_2:t}} = \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} + \frac{s^2_{P_2,g_{P_1:s,P_2:t}}}{N^{P_2}_{P_1:s,P_2:t}},$$

which is upwardly biased.

Overall, the variance of estimator $\hat{y}_{P_1-P_2}$ across all segments is

$$\mathrm{Var}_W\left(\hat{y}_{P_1-P_2}\right) = \sum_{\substack{s,t \\ s\neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \mathrm{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t} - \hat{y}_{P_1:s,\underline{P_2}:t}\right),$$

which we estimate with

$$\mathrm{Var}^{\mathrm{Neyman}}_{P_1-P_2} = \sum_{\substack{s,t \\ s\neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \mathrm{Var}^{\mathrm{Neyman}}_{g_{P_1:s,P_2:t}}.$$

This is upwardly biased and, as a result, confidence intervals will be conservative.

Summary

As we mentioned at the start of this section, in Table 1 we summarize the estimators and variances for the performance of a single policy and comparison of two policies under the standard and proposed approaches. We note that under both the standard approach and the proposed approach, the calculations for the variances include an adjustment for the size of the super-population. We also include this adjustment in our calculations in the example in the next section. However, the adjustment asymptotes to zero as the super-population grows large and so for large enough populations it is often ignored.

[2]The derivation for the variance of the difference in each segment is similar to the derivation of Equation (5) for the overall variance under the standard approach.

Table 1: Evaluating and Comparing Policies: Estimators and Variances

	Standard Approach	**Proposed Approach**
Single Policy Evaluation		
Estimator $\hat{y}_{P_1}$	$\frac{1}{N_{P_1}} \sum_{i=1}^{N} \mathbb{1}_{W_i=P_1} Y_i^{obs}$	$\frac{\sum_{s,t} N_{P_1:s,P_2:t} \cdot \hat{y}_{\underline{P_1}:s,P_2:t}}{N}$
Variance $\widehat{\mathrm{Var}}_W\left(\hat{y}_{P_1}\right)$	$\frac{s^2_{P_1,N}}{N_{P_1}}\left(1-\frac{N_{P_1}}{N}\right)$	$\sum_{\substack{s,t \\ s\neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} \left(1-\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}}\right) + \sum_s \left(\frac{N_{P_1:s,P_2:s}}{N}\right)^2 \frac{s^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}} \cdot \left(1-\frac{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}}\right)$
Comparison of Two Policies		
Estimator $\hat{y}_{P_1-P_2}$	$\frac{1}{N_{P_1}} \sum_{i=1}^{N} \mathbb{1}_{W_i=P_1} Y_i^{obs} - \frac{1}{N_{P_2}} \sum_{i=1}^{N} \mathbb{1}_{W_i=P_2} Y_i^{obs}$	$\frac{\sum_{\substack{s,t \\ s\neq t}} N_{P_1:s,P_2:t}\left(\hat{y}_{\underline{P_1}:s,P_2:t}-\hat{y}_{P_1:s,\underline{P_2}:t}\right)}{N}$
Variance $\mathrm{Var}^{\mathrm{Neyman}}_{P_1-P_2}$	$\frac{s^2_{P_1,N}}{N_{P_1}}+\frac{s^2_{P_2,N}}{N_{P_2}}$	$\sum_{\substack{s,t \\ s\neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \left(\frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}}+\frac{s^2_{P_2,g_{P_1:s,P_2:t}}}{N^{P_2}_{P_1:s,P_2:t}}\right)$

The table summarizes the estimators and variances for the performance of a single policy and comparison of two policies under the standard and proposed approaches.

We conclude this section with a formal comparison of the efficiency of the standard and the proposed approaches and several comments, including a brief discussion of limitations.

Comparison of the Efficiency of the Proposed vs. the Standard Approach

The expressions in Table 1 allow us to analytically compare the efficiency of the proposed and standard approaches. We have argued that our proposed method has two efficiency advantages: it reduces variance introduced by between-segment differences; and it ensures the true and observed difference between two policies in segments in which they recommend the same action is zero, instead of introducing random noise. We now show analytically that the standard errors of the estimators strictly improve under the proposed method.

Our argument requires two assumptions. The first assumption is that the experimental design is balanced, i.e., for a policy we are evaluating, the proportion of customers assigned to that policy within a segment is the same as the proportion of customers assigned to that policy in any other segment:

$$\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}} = \frac{N_{P_1}}{N}, \text{ for all } s,t,$$

and the same is true for all policies we are evaluating. For example, if we are also evaluating policy P_2, then

$$\frac{N^{P_2}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}} = \frac{N_{P_2}}{N}, \text{ for all } s,t.$$

The second assumption is that for all policies we are evaluating, the observed variance within any segment that received the policy is not larger than the observed variance across all observations in that condition:

$$s^2_{P_1,g_{P_1:s,P_2:t}} \leq s^2_{P_1,N}, \quad s^2_{g_{P_1:s,P_2:s}} \leq s^2_{P_1,N}, \qquad \text{for all } s,t.$$

This assumption is aligned with the well-known intended benefit of stratification (e.g., Imbens and Rubin, 2015): we segment in order to achieve balance in the covariates. This means that the units within each block would be similar with respect to the covariates or some functions of the covariates. It therefore makes sense to expect that within any segment, the observed variance is not larger than in the aggregate.

The following theorem formalizes our result that the proposed method strictly improves efficiency. The proof is relegated to the Appendix.

Theorem 1. *Assume random sampling and assignment without replacement as per Assumption 1. Further, assume that* $\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}} = \frac{N^{P_2}_{P_1:s,P_2:t}}{N_{P_2}} = \frac{N_{P_1:s,P_2:t}}{N}$ *and that* $s^2_{P_1,g_{P_1:s,P_2:t}} \leq s^2_{P_1,N}$, $s^2_{P_2,g_{P_1:s,P_2:t}} \leq s^2_{P_2,N}$ *for all* s,t *with* $s \neq t$, *and* $s^2_{g_{P_1:s,P_2:s}} \leq s^2_{P_1,N}$ *for all* s. *Then for the evaluation of policy* P_1, *the estimated variance of the estimator under the proposed approach is strictly less than the estimated variance of the estimator under the standard approach. Furthermore, for the comparison of policies* P_1 *and* P_2, *the estimated variance of the estimator of the difference under the proposed approach is strictly less than the estimated variance of the estimator of the difference under the standard approach.*

Comparing n Policies

We note that the extension from the comparison of two policies to the comparison of $n > 2$ policies, with $n \leq T$, is straightforward. With multiple policies, the intersection of their recommended actions provides an even more fine-grained segmentation of customers. For example, with two possible actions, a single targeting policy maps customers into two segments (assuming the policy is not degenerate). The intersection of n targeting policies maps customers into 2^n segments (although it is possible that some segments may be empty). With n targeting policies, we define segments of customers in the super-population in terms of what action each of the n policies would recommend. The same efficiency advantages as described before are in effect. First, segmentation reduces variance introduced by between-segment differences. Second, in a segment in which k of the n policies recommend the same action, the proposed approach uses the outcome from all customers within the segment that were assigned to any of the k policies to evaluate these policies; the sample size thus increases. Third, the proposed approach recognizes that in a segment in which k of the n policies recommend the same action, the difference in the outcomes of the k policies is zero by construction; random error does not get introduced.

Limitations of the Proposed Approach

We have highlighted how segmenting by the actions recommended by candidate policies can improve the efficiency of experiments designed to evaluate different targeting policies. Theorem 1 establishes that, under assumptions, the estimated variance under the proposed approach will be smaller than the estimated variance under the standard approach. However, we recognize that there are limitations to this approach when the assumptions of Theorem 1 about balanced experimental design and relatively small within-segment variances are violated. The most prominent limitation is that, if the between-segment variation is small, this approach could theoretically lower precision due to small sample effects.

As argued by Imbens and Rubin (2015) and Athey and Imbens (2017), in expectation the variance under stratification (i.e., segmentation) cannot be larger than the variance without stratification, despite the small samples. This means that there is no cost to segmentation in terms of the variance itself; nevertheless, there is a cost in terms of estimation of the variance. The variance with segmentation is less than or equal to the variance without segmentation, and, assuming unbiased estimators for the variance, the expectation of the estimated variance with segmentation will therefore be less than or equal to the expectation of the estimated variance without segmentation. However, it is the variance of the estimator of the variance that can be larger with segmentation than without segmentation. The reduction of the degrees of freedom due to smaller sample sizes may result in cases where the estimated variance with segmentation is larger than the estimated variance without segmentation.

In practice, because the segments are constructed using the recommendation actions, the adjustment in degrees of freedom due to small samples will generally be outweighed by the benefits of comparing outcomes within segments and then aggregating within-segment differences to remove between-segment variation. However, if there is a risk of large outliers, or there is a large number of possible actions, then these small sample effects could be important. We briefly discuss both possibilities.

An example of large outliers could include promotions sent to attract new customers. Typically fewer than 1% of customers respond to prospecting promotions.[3] If most of the customers who respond spend an average of $500 a year, but an occasional customer spends over $10,000 (perhaps because they are re-selling merchandise), then there is a risk of large outliers. This risk is not a problem if the sample size is large enough. However, comparing outcomes within segments reduces the sample size of the comparisons. This risk may be mitigated by truncating outlying observations (such as capping annual purchases at $1,000).

[3] A low response rate is generally optimal. If the average response rate is above 1%, then it is often profitable for the firm to mail to more prospective customers (including those who have a lower probability of responding).

The small sample limitation also becomes more relevant if the number of possible actions is very large. Segmenting by recommended actions may divide the sample into many segments. This limitation is particularly relevant for dynamic policies that involve a sequence of actions. For example, Simester et al. (2006) tested their dynamic catalog targeting policy using a sequence of 12 catalog mailing opportunities. This represents 212 possible mailing sequences, which correspond to 4,096 different actions. Similarly, Toubia et al. (2013) tested their dynamic conjoint design algorithm using a sequence of 16 paired comparison conjoint questions. An action in this setting represents a sequence of 16 paired comparison conjoint questions, and each question is chosen from a wide range of potential questions. As a result, the entire action space includes millions of potential question combinations. In this setting it would be impractical to segment customers according to the actions recommended by alternative policies.

4 An Example Using Actual Data

In this section we illustrate the efficiency advantages offered by the proposed approach using data from an actual experiment designed to validate different targeting policies. We use this data to calculate the estimates and standard errors under both the standard and proposed approaches.

In a recent study, Simester, Timoshenko, and Zoumpoulis (2017) (hereafter "STZ") investigate how a retailer should target prospective customers with promotions. They consider three different actions, including two different types of promotions sent by mail and a third no mail (control) treatment. We will label these actions as actions a, b and c. They compare seven different targeting policies, which we label $P_1, \ldots, P_7$.

In their study, STZ randomly assign approximately four million prospective households to ten experimental conditions. These ten experimental conditions correspond to the seven candidate policies and the three actions. If a household were in one of the seven experimental conditions associated with a targeting policy, the household received the action recommended by that targeting policy. If the household were in one of the three experimental conditions associated with an action, the household received that action.

We will initially focus on the two experimental conditions associated with policies P_1 and P_2 (in the next section we extend the focus to the conditions associated with the three actions). We illustrate the standard approach and proposed approach for comparing the outcomes of these two policies. Table 2 reports the sizes of the samples that are assigned to each of the two policies P_1 and P_2 and the average outcomes for these samples, where the households are segmented according to the actions recommended by each policy. To preserve confidentiality, the average profits are all multiplied by a common random number.

When evaluating a single policy, the standard approach aggregates the outcomes within each experimental condition. This yields standard errors of \$0.043 and \$0.038 for policies P_1 and P_2, respectively, as shown in Table 3.

Table 2: STZ study: Outcomes for the Experimental Conditions Associated with Policies P_1 and P_2

	Recommended Action		**Sample Size**		**Average Profit**			
	Policy P_1	Policy P_2	Policy P_1	Policy P_2	**Standard Approach**		**Proposed Approach**	
Row	s	t	$N^{P_1}_{P_1:s,P_2:t}$	$N^{P_2}_{P_1:s,P_2:t}$	P_1	P_2	$\hat{y}_{\underline{P_1}:s,P_2:t}$	$\hat{y}_{P_1:s,\underline{P_2}:t}$
1	a	a	492	924	\$0.253	−\$0.375	−\$0.157	−\$0.157
2	a	b	17,342	12,285	\$0.316	\$0.048	\$0.316	\$0.048
3	a	c	21,685	20,430	−\$0.058	\$0.191	−\$0.058	\$0.191
4	b	a	28,182	32,236	\$2.136	\$1.757	\$2.136	\$1.757
5	b	b	214,693	209,792	\$1.886	\$1.786	\$1.837	\$1.837
6	b	c	67,405	69,165	\$1.067	\$0.800	\$1.067	\$0.800
7	c	a	2,831	1,563	\$0.043	−\$0.056	\$0.043	−\$0.056
8	c	b	19,946	14,772	\$0.173	−\$0.228	\$0.173	−\$0.228
9	c	c	52,413	46,671	\$0.086	\$0.133	\$0.108	\$0.108
	Total		424,989	407,838	\$1.293	\$1.210	\$1.278	\$1.192

The table reports outcomes from an experiment reported by Simester, Timoshenko, and Zoumpoulis (2017). It reports the sample size and average profit in the two experimental conditions associated with policies P_1 and P_2. The outcomes are reported for each customer segment, where the customer segments are identified by the actions recommended by P_1 and P_2. The shading identifies the segments in which the two policies recommend the same action. The profits are all multiplied by a common random number.

The proposed approach introduces two changes when evaluating a single policy. First, standard errors are calculated separately within each segment, and these standard errors are then aggregated across segments. This helps reduce variation introduced by between-segment differences. Second, for the three segments in which the two policies recommend the same action (rows 1, 5 and 9 in Table 2), the proposed approach uses the outcome from all customers, irrespective of which of the two policies they are assigned to. Pooling across the two conditions in these segments increases the sample size, yielding an additional efficiency improvement. The standard errors under the proposed approach are \$0.033 and \$0.032 for policies P_1 and P_2, respectively. As expected, these are both smaller than the corresponding standard errors under the standard approach.

This pooling is only possible because we segmented the customers by the recommended actions of the candidate policies. If we had segmented by a covariate, the recommended actions would have almost certainly varied between the two policies within every segment. Recall also that the STZ study actually implemented seven different targeting policies. Although not reported here, these additional experimental conditions offer additional opportunities to pool across conditions where the policies recommend the same actions.

The efficiency improvements when evaluating a single policy are relatively modest compared to the efficiency improvements when comparing two policies. The standard approach to comparing the outcome of two targeting policies is to first calculate the average outcome for each policy, and then to calculate the difference in these means. This approach yields an estimated difference of \$0.083

Table 3: Evaluating and Comparing Policies P_1 and P_2 in the STZ study: Estimates and Standard Errors

		Standard Approach	**Proposed Approach**
Policy P_1	Estimate	\$1.293	\$1.278
	Standard Error	\$0.043	\$0.033
Policy P_2	Estimate	\$1.210	\$1.192
	Standard Error	\$0.038	\$0.032
Policy P_1 **- Policy** P_2	Estimate	\$0.083	\$0.086
	Standard Error	\$0.057	\$0.026

The table reports the estimates of the performances of two policies and their comparison, along with standard errors, using data from Simester, Timoshenko, and Zoumpoulis (2017). The profits are all multiplied by a common random number.

with a standard error of \$0.057.

The proposed approach again offers two efficiency advantages when comparing policies. The first advantage is that it reduces variance introduced by between-segment differences. Differencing outcomes within each segment and then aggregating these differences across segments is an effective way of controlling for segment-specific effects. These segment-specific effects are a source of variation in the standard approach, but are essentially removed under the proposed approach.

The second efficiency improvement when comparing policies occurs in segments for which the candidate policies recommend the same action. By pooling customers across the two conditions in these segments, the proposed approach recognizes that the difference in the outcomes of the two policies is zero by construction. For example, in the first row of Table 2, policies P_1 and P_2 both recommend action a. Because the two policies both recommend the same action, there is no difference in the *true* outcome for this segment. However, even though we know the outcomes should be identical for this segment, under the standard approach the *observed* outcomes attributed to each policy are different. The firm earned \$0.253 on average from the 492 households in the experimental condition associated with policy P_1 and lost \$0.375 on average from the 924 households in the experimental condition associated with policy P_2. By including these results in the aggregate outcomes, the standard approach introduces an additional source of noise to its estimates. In contrast, under the proposed approach, the estimates for both policies are the same for the segment in the first row (and equal to $-\$0.157$).

These changes reduce the standard error when comparing the two policies, from \$0.057 under the standard approach, to \$0.026 under the proposed approach. This is a substantial efficiency improvement, which also has substantive importance in our example. By cutting the standard error in half, the difference in outcomes between the two policies becomes statistically significant ($p < 0.05$), whereas the difference was not significant when using the standard approach.

If we were only interested in the relative performance of the two policies, and were not interested in the absolute performance of either policy, then we could omit from the study customers in the three segments in which the two policies recommend the same action. This represents 524,985 customers, or 63% of the sample of customers that were assigned either to P_1 or to P_2. This 63% of the sample provides no information about the relative performance of the two policies. The costs associated with mailing to this 63% of the sample could either be saved, or re-allocated to increasing the number of customers assigned to either P_1 or P_2 from the other six segments. If we were to re-allocate this 63% of the sample to the other six segments using the weights reported in Table 2, the expected average difference in the outcomes from the two policies would not change. However, in those six segments the increase in sample size will reduce the standard errors of the differences quite drastically.[4]

5 An Alternative Experimental Design: Randomizing by Action

Our primary focus is to illustrate how we can design segments that allow us to improve efficiency when comparing targeting policies. These efficiency improvements require that we segment customers by the actions recommended by the candidate targeting policies. At least some of the benefits of this segmentation approach survive when evaluating targeting policies using an alternative experimental design.

Recall that the traditional experimental design for evaluating targeting policies constructs a separate experimental condition for each candidate policy. Customers are randomly assigned to policies, and so all of the customers in an experimental condition are targeted with the actions recommended by a single policy. A disadvantage of this traditional design is that it only allows comparison of the policies that were implemented. If a new policy is subsequently identified, it is generally not possible to compare the performance of the new policy without implementing it. An alternative experimental design allows evaluation of any targeting policy, and so any new policy can be evaluated using the existing data.

Instead of randomly assigning customers to policies, the alternative design randomly assigns customers to actions. Instead of one experimental condition for each candidate policy, there is one experimental condition for each possible action. The number of experimental conditions matches the number of possible actions (i.e., the size of the action space). In contrast, under the traditional design, the number of experimental conditions matches the number of candidate targeting policies, which could be very large.

For example, consider a retailer choosing whether to mail promotion a, b or c. One randomly

[4] A simple simulation estimates the standard error of the difference in the two policies would be reduced from \$0.026 to \$0.016. To run the simulation, we re-allocate each observation in the 63% of the sample (i.e., in the three segments in which the two policies recommend the same action) to one of the remaining six segments, selected at random, with probabilities proportional to the sizes of the six segments. Having chosen a segment for an observation, we then simulate its profit value to be equal to one of the realized profit values within that segment, selected uniformly at random. We repeat ten times and average the results.

selected sample of customers receives promotion a, a second randomly selected sample receives promotion b, and a third randomly selected sample promotion c. As long as the samples are randomly selected, the sample sizes across these three experimental conditions need not be the same. For example, 100,000 could receive promotion a, 50,000 could receive promotion b, and the retailer's remaining 20 million customers could receive promotion c (which might be no promotion at all).[5]

Randomizing by action guarantees that, within each segment that is large enough, each possible action is received by a random sample of customers. The key idea behind the alternative experimental design is the following: to evaluate a policy on a segment of customers for which the policy recommends a specific action, use the customers within the segment that were hit with that specific action.

We can illustrate this point by returning to the example that we discussed at the start of Section 2. Consider the segment of (female) customers for which policy P_1 recommends promotion b and policy P_2 recommends promotion c. Consider an experimental design that randomly assigns a third of all customers to each promotion (a, b, and c). Because the assignment is random, the three sub-samples of female customers are equivalent. This allows us to directly evaluate the outcomes from the female customers that received promotion b and the female customers that received promotion c. If we also use the outcomes from the male customers who received promotion a, we can then evaluate the outcome of each of the policies P_1 and P_2.

This is only possible if we segment customers using the recommended actions from each policy. If instead of segmenting customers by the recommended actions, we segment by age, then within each age segment the policies recommend different actions for male and female customers. As a result, there is no longer a single action for each policy associated with a segment, and so we cannot use the alternative experimental design to evaluate the outcome of a policy. An experimental design in which we randomize by action allows us to evaluate the outcome of a policy on a segment for which the policy recommends a single action, but does not allow us to evaluate the outcome of multiple actions.[6]

It is obviously important that to evaluate every alternative policy, the experiment randomly assigns every type of customer to every possible action. For example, if the experiment just included experimental conditions for promotions a and b, we could not evaluate policies that recommend policy c. This limitation never arises when customers are randomly assigned to actions. The random assignment to actions ensures that for every large enough segment, every action is implemented for a

[5]This is a relatively simple experiment to implement. Within each experimental condition all participants receive the same action. In practice, this is often easier to implement than targeting different participants within the same experimental condition with different actions. It is also likely to be less susceptible to implementation errors.

[6]A rare exception could arise if a targeting policy recommended randomizing between actions within a segment. In this case we could use a weighted average of the outcomes for each action. However, this would be a very unusual targeting policy. In general targeting policies target customers with different actions because the customers are different. It is these customer differences that prevent us from using a weighted average of the outcomes for each action.

random sample of customers. However, this condition may not be satisfied when randomly assigning customers to policies instead of actions. Consider a segment of customers for which two policies both recommend promotion a. Based on the outcomes of the two policies on the segment, we cannot evaluate an alternative policy that recommends promotion b or c on that segment.[7]

Limitations of Randomizing by Action

While randomly assigning customers by actions has clear advantages, there are also disadvantages. One potential disadvantage is cost. It is sometimes obvious that an action is optimal for only a small segment of the population, and so randomly assigning customers to receive this action may lead to an opportunity cost. For example, if mailing a catalog to customers is profitable for most customers, then deciding to withhold these catalogs from a randomly selected sample of customers will result in foregone profit. This cost can be minimized by under-sampling the actions that are not optimal for most customers.

In a related point, it may be unethical or unacceptable to randomly assign some customers to some conditions. For example, there is an extensive literature studying the impact of interventions designed to reduce poverty. An important research question in this literature is the design of targeting policies to ensure that the interventions only target the truly poor households and not the rich (Hanna and Karlan, 2017). Randomly assigning actions to households could result in some of the rich households receiving the interventions, which may be politically unacceptable. Alternatively, in the medical field, it may be unethical to withhold some treatments from some patients. This limitation can be easily addressed by designing the randomization procedures to prevent experimental conditions that are unacceptable or unethical. Although this may prevent evaluation of every possible policy, it does not prevent evaluation of any policy that is acceptable and ethical.

Applying our Proposed Analysis to Randomization by Action

Randomization by action can be thought of as an example of cross validation. In order for cross validation to be feasible, customers must be segmented according to the recommended actions from each policy. In this respect, our insight about the benefits of segmenting customers using recommended actions are critical to evaluation of policies under a randomized by action design. Without this segmentation, we cannot evaluate a targeting policy when using this experimental design.

[7] This is a second reason for which we may not want to omit segments of customers in which alternative policies recommend the same action. Recall from our earlier discussion that omitting these customers results in no loss of information about the relative performance of the two policies. However, omitting these customers means that it is no longer possible to evaluate the absolute performance of either policy. It also makes it impossible to test policies that recommend a different action for this segment of customers. Recall that evaluating every alternative policy requires that we have randomly assigned every type of customer to every action. For both reasons, instead of omitting customer segments for which the policies make the same recommendations, it will often be preferable to simply under-sample these segments.

This has an important implication when comparing the "standard" and "proposed" approaches for estimating the outcome of a single policy or the difference in two policies. Because estimating outcomes *requires* segmentation of the customers according to the recommended actions, a randomized by action design cannot use the "standard" approach that is available when randomizing by policy. There is nothing equivalent to the standard approach when randomizing by action.

However, one could use the following standard approach to estimate the variance: identify all of the observations that are used to evaluate a policy and estimate the variance in the outcome using the sample variance across all of these observations. Such approach is neither mathematically justified nor efficient. Instead, the proposed approach to estimating the variance discussed in the previous sections can also be extended to a randomized by action design. In particular, the variance of the outcomes can be estimated within each segment, and these within-segment variances can then be aggregated across segments.

We make two additional comments. Our first comment focuses on the evaluation of a single policy. In the previous sections we noted that when evaluating a single policy under a randomized by policy design, there are two sources of efficiency improvement; (a) improved treatment of between-segment differences, and (b) pooling of observations across experimental conditions when two policies recommend the same action. While the first benefit survives, the second benefit is not relevant when analyzing a randomized by action experimental design. Because we do not have separate experimental conditions associated with each policy, we do not have observations across separate experimental conditions to pool over.

The second comment makes a similar point when comparing two policies. Recall that under a randomized by policy design the two benefits of using the proposed approach were: (a) improved treatment of between-segment differences, and (b) recognition that the difference in performance is truly zero in segments in which the policies recommend the same actions. The benefit from the treatment of between-segment differences again survives under a randomized by action design. The second benefit is embedded to the randomized by action design by construction: under this design, any estimate of the outcome relies upon segmentation of customers using the recommended actions, and so an estimator will recognize there is no difference in outcomes when policies recommend the same actions. In a segment where both policies recommend the same action, the evaluation of each policy uses an identical set of observations, and so there can be no difference in the policies' estimated outcomes.

In the Appendix we demonstrate how to extend our estimate and variance calculations under the proposed approach to an experiment in which customers are randomly assigned to actions (instead of policies). We formally describe how to evaluate a single policy and how to compare two policies under the proposed approach, and we derive expressions for the respective variance estimators.

In the remainder of this section we return to the STZ study and illustrate how our proposed approach improves efficiency when using a randomized by action design.

Application to the STZ Study

Recall that the STZ study implemented both a randomized by policy and a randomized by action design. In seven of the experimental conditions customers were assigned a targeting policy, and thus received the action that the policy would recommend for them. In three of the experimental conditions customers were assigned an action. This provides an ideal dataset to illustrate the application of the alternative experimental design.

In Table 2 we previously summarized the outcomes in the two experimental conditions associated with two of the seven targeting policies (we labelled them policies P_1 and P_2). We again focus on these two targeting policies and illustrate how to evaluate them using only the experimental conditions associated with the three actions. In Table 4 we report the sample sizes and average outcomes in each of these three experimental conditions. These findings are reported separately for each segment, where the segments are defined by the actions recommended by the two policies.

Table 4: STZ Study: Outcomes for the Experimental Conditions Associated with the Three Actions

	Recommended Action		**Sample Size**			**Average Profit**		
	Policy P_1	Policy P_2	Action a	Action b	Action c	Action a	Action b	Action c
Row	s	t	$N^{P_a}_{P_1:s,P_2:t}$	$N^{P_b}_{P_1:s,P_2:t}$	$N^{P_c}_{P_1:s,P_2:t}$			
1	a	a	1,579	3,963	869	−\$0.347	−\$0.157	\$0.952
2	a	b	13,057	11,564	16,744	\$0.096	\$0.013	\$0.243
3	a	c	21,789	10,712	18,480	−\$0.089	−\$0.009	\$0.273
4	b	a	21,688	17,076	26,424	\$1.799	\$1.467	\$1.159
5	b	b	204,833	225,996	202,136	\$1.499	\$1.750	\$1.321
6	b	c	64,403	51,847	73,883	\$1.206	\$1.194	\$0.799
7	c	a	3,714	3,165	3,251	−\$0.392	−\$0.535	\$0.306
8	c	b	22,281	21,448	18,675	−\$0.263	−\$0.081	\$0.146
9	c	c	50,838	51,153	52,333	−\$0.268	−\$0.308	\$0.118
	Total		404,182	396,924	412,795	\$0.993	\$1.166	\$0.912

The table reports outcomes from an experiment reported by Simester, Timoshenko, and Zoumpoulis (2017). It reports the sample size and average profit in the three experimental conditions associated with actions a, b, and c. The outcomes are reported for each customer segment, where the customer segments are constructed using the actions recommended by policies P_1 and P_2. The shading identifies the outcomes used to evaluate policy P_1. The profits are all multiplied by a common random number.

We use shading to highlight the outcomes that can be used to evaluate policy P_1: to evaluate a policy on a segment of customers to which the policy recommends a specific action, we use the customers within the segment that were hit with that specific action. For example, for the segment of customers in which P_1 recommended action a and P_2 recommended action b (row 2), we observe an average profit of \$0.096 as the outcome for action a, \$0.013 for action b, and \$0.243 for action c. Only the outcome for action a is relevant for evaluating policy P_1 for the segment of customers in that row. Notice that randomization ensures that within this segment, a sub-segment of customers were assigned to each of the three actions.

In Table 5 we report the estimates and standard errors of the profits earned for each policy and the difference in these profits between the two policies. As we discussed, in a randomized by action design there is nothing equivalent to the standard approach for the estimator of the outcome. Any estimator relies upon segmentation of the customers according to the recommended actions. In particular, to evaluate policy P_1 we first segment the 405,603 observations into nine segments (represented by the nine shaded cells in Table 4). We calculate the average outcome in each segment, and then aggregate these segment averages using their corresponding weights.

To estimate the variance of this estimator, we estimate the variance within each segment and aggregate this variance across segments. In contrast, a standard (yet unjustified) approach would be to use the sample variance across all 405,603 observations. These calculations are detailed in Table 5.

To evaluate the difference in the performance of two policies, we calculate the difference in the estimates of each policy's performance. To estimate the variance in this estimate of the difference, we estimate the variance of the difference within each of the nine segments, and aggregate this variance across segments. In contrast, a standard (yet unjustified) approach would be to use the overall variance estimates for policies P_1 and P_2. These calculations are detailed in Table 5. The findings confirm that the proposed approach is more efficient as it produces smaller standard errors.

Table 5: Evaluating and Comparing Policies P_1 and P_2 in the STZ Study Using the Alternative Experimental Design

		Standard Approach	**Proposed Approach**
Policy P_1	Estimate	—	\$1.206
	Standard Error	\$0.047	\$0.046
Policy P_2	Estimate	—	\$1.157
	Standard Error	\$0.044	\$0.043
Policy P_1 **- Policy** P_2	Estimate	—	\$0.048
	Standard Error	\$0.065	\$0.027

The table reports the estimates of the performances of two policies and their comparison, along with standard errors, when evaluating the policies using the alternative experimental design, using data from Simester, Timoshenko, and Zoumpoulis (2017). The profits are all multiplied by a common random number.

Recall that the advantage of randomizing by action is that it allows evaluation of any targeting policy. To evaluate every possible targeting policy we require that for any segment of customers there is a random sample of customers that received action a, another random sample that received action b, and a third random sample that received action c. Randomizing by action guarantees that this is always satisfied, assuming the segments are large enough. In contrast, randomizing by policy does not guarantee this, as we mentioned above. However, it is possible that variation in recommended actions across the seven targeting policies may mean that randomizing by policy also satisfies that, for some segments, every action is implemented for a sample of customers. We can investigate this possibility by asking: given a segment, are all three actions represented in the actions recommended by the seven policies for the households in the segment? For example, if policy P_1 recommends action a, policy P_4 recommends action c and policy P_7 recommends action b, then all three actions would be represented. A simple comparison reveals that in the STZ study, all three actions are represented by the seven policies for just 30.92% of the approximately four million households. For the remaining 69.08% of households, there is at least one action that is not recommended by any of the seven candidate policies. For these households it is not possible to evaluate policies that recommend the omitted action(s).

This example also highlights one of the limitations of the randomized-by-action experimental design. Recall that randomly assigning actions to customers may introduce an opportunity cost if some actions are optimal for only a small number of customers. In the STZ study, action b is more profitable than the other two actions; the average profit from action b is \$1.166 compared to \$0.993 and \$0.912 for the other two actions (see Table 4). The seven optimized policies $P_1, \ldots, P_7$ recognize the profitability of action b, and so they recommend this action for most of the households. As a result, the average profit across the approximately 2.8 million households in the seven experimental conditions pertaining to the seven policies is \$1.208. This is significantly higher than the \$1.022 average profit earned from the approximately 1.2 million in the conditions associated with the three actions. Randomization by action on these approximately 1.2 million customers resulted in an opportunity cost to the firm of approximately \$225,000. This cost could be reduced by underweighting actions a and c when randomly allocating customers to actions.

6 Conclusions

The gold standard for evaluating targeting policies is to evaluate them using an experiment. We have presented an approach to analyzing these experiments that improves their efficiency. The key insight is that there are important advantages of segmenting the experimental data according to the actions recommended by the target policies. This segmentation not only controls for between-segment differences, it also yields segments in which candidate targeting policies recommend the same actions. Within the segments in which the policies recommend the same actions, we know there is no difference in the true performance of the policies. This insight helps improve efficiency both when evaluating a single policy and when comparing policies.

We provide expressions for the estimates and variance calculations under both the standard approach and proposed approach. These expressions allow us to formally compare the efficiency of the two approaches and establish conditions under which efficiency is guaranteed to improve under the proposed approach. We also illustrate how to apply these calculations using data from an actual experiment.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2014), "Finite population causal standard errors." Working Paper 20325, National Bureau of Economic Research, URL http://www.nber.org/papers/w20325.

Athey, S. and G.W. Imbens (2017), "The econometrics of randomized experiments." *Handbook of Economic Field Experiments*, 1, 73 – 140, URL http://www.sciencedirect.com/science/article/pii/S2214658X16300174. Handbook of Field Experiments.

Barajas, Joel, Ram Akella, Marius Holtan, and Aaron Flores (2016), "Experimental designs and estimation for online display advertising attribution in marketplaces." *Marketing Science*, 35, 465–483, URL https://doi.org/10.1287/mksc.2016.0982.

Belloni, Alexandre, Mitchell J. Lovett, William Boulding, and Richard Staelin (2012), "Optimal admission and scholarship decisions: Choosing customized marketing offers to attract a desirable mix of customers." *Marketing Science*, 31, 621–636.

Congdon, William J., Jeffrey R. Kling, Jens Ludwig, and Sendhil Mullainathan (2017), "Social policy: Mechanism experiments and policy evaluations." In *Handbook of Field Experiments* (A. Banerjee and E. Duflo, eds.), 1, chapter 15, Elsevier.

Hanna, R. and D. Karlan (2017), "Designing social protection programs." *Handbook of Economic Field Experiments*, 2, 515 – 553, URL http://www.sciencedirect.com/science/article/pii/S2214658X16300022. Handbook of Economic Field Experiments.

Imbens, Guido W. and Donald B. Rubin (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, NY, USA.

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, URL +http://dx.doi.org/10.1093/qje/qjv023.

Li, Jimmy Q., Paat Rusmevichientong, Duncan Simester, John N. Tsitsiklis, and Spyros I. Zoumpoulis (2015), "The value of field experiments." *Management Science*, 61, 1722–1740, URL https://doi.org/10.1287/mnsc.2014.2066.

Louviere, Jordan, Deborah J. Street, and Leonie Burgess (2004), *A 20+ Years' Retrospective on Choice Experiments*, 201–214. Springer US, Boston, MA, URL http://dx.doi.org/10.1007/978-0-387-28692-1_9.

Lu, Shasha, Li Xiao, and Min Ding (2016), "A video-based automated recommender (var) system for garments." *Marketing Science*, 35, 484–510, URL `https://doi.org/10.1287/mksc.2016.0984`.

Mantrala, Murali K., P. B. Seetharaman, Rajeeve Kaul, Srinath Gopalakrishna, and Antonie Stam (2006), "Optimal pricing strategies for an automotive aftermarket retailer." *Journal of Marketing Research*, 43, 588–604.

Mersereau, A. J., P. Rusmevichientong, and J. N. Tsitsiklis (2009), "A structured multiarmed bandit problem and the greedy policy." *IEEE Transactions on Automatic Control*, 54, 2787–2802.

Neslin, S.A., T.P. Novak, K.R. Baker, and D.L. Hoffman (2009), "An optimal contact model for maximizing online panel response rates." *Management Science*, 55, 727–737.

Neyman, Jerzy (1934), "On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection." *Journal of the Royal Statistical Society*, 97, 558–625, URL `http://www.jstor.org/stable/2342192`.

Simester, D., A. Timoshenko, and S. I. Zoumpoulis (2017), "Customizing marketing decisions using field experiments." *Working Paper*.

Simester, D. I., P. Sun, and J. N. Tsitsiklis (2006), "Dynamic catalog mailing policies." *Management Science*, 52, 683–696.

Skiera, Bernd and Nadia Abou Nabout (2013), "Practice prize paper—prosad: A bidding decision support system for profit optimizing search engine advertising." *Marketing Science*, 32, 213–220, URL `https://doi.org/10.1287/mksc.1120.0735`.

Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié (2013), "Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters." *Management Science*, 59, 613–640, URL `https://doi.org/10.1287/mnsc.1120.1570`.

Urban, Glen L., Gui Liberali, Erin MacDonald, Robert Bordley, and John R. Hauser (2014), "Morphing banner advertising." *Marketing Science*, 33, 27–46.

A Evaluating a Single Policy Using the Standard Approach - the Variance

We have $\text{Var}_W(\hat{y}_{P_1}) = \mathbb{E}_W[\hat{y}_{P_1}^2] - (\mathbb{E}_W[\hat{y}_{P_1}])^2$. For the first term, we write

$$\mathbb{E}_W[\hat{y}_{P_1}^2] = \mathbb{E}_W\left[\left(\frac{1}{N_{P_1}}\sum_{i=1}^{N}\mathbb{1}_{W_i=P_1}Y_i^{obs}\right)^2\right].$$

The squared terms of the sum are of the form

$$\mathbb{E}_W\left[\left(\mathbb{1}_{W_i=P_1}Y_i^{obs}\right)^2\right] = \frac{N_{P_1}}{N}\cdot Y_i^2\left(P_1(i)\right),$$

while the cross terms are of the form

$$\begin{aligned}
\mathbb{E}_W\left[\left(\mathbb{1}_{W_i=P_1}Y_i^{obs}\right)\left(\mathbb{1}_{W_j=P_1}Y_j^{obs}\right)\right] &= \Pr\left(W_i = W_j = P_1\right)\cdot Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right)\\
&= \frac{\binom{N-2}{N_{P_1}-2}}{\binom{N}{N_{P_1}}}\cdot Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right)\\
&= \frac{N_{P_1}(N_{P_1}-1)}{N(N-1)}\cdot Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right),
\end{aligned}$$

for $i \neq j$. Overall, we can write

$$\mathbb{E}_W[\hat{y}_{P_1}^2] = \frac{1}{N_{P_1}^2}\left(\frac{N_{P_1}}{N}\sum_{i=1}^{N}Y_i^2\left(P_1(i)\right) + \frac{N_{P_1}(N_{P_1}-1)}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right)\right)$$

and

$$(\mathbb{E}_W[\hat{y}_{P_1}])^2 = \left(\frac{1}{N}\sum_{i=1}^{N}Y_i\left(P_1(i)\right)\right)^2 = \frac{1}{N^2}\left(\sum_{i=1}^{N}Y_i^2\left(P_1(i)\right) + \sum_{i=1}^{N}\sum_{j\neq i}Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right)\right).$$

Combining the two, we have that

$$\text{Var}_W(\hat{y}_{P_1}) = \left(\frac{1}{N_{P_1}\cdot N} - \frac{1}{N^2}\right)\sum_{i=1}^{N}Y_i^2\left(P_1(i)\right) - \left(\frac{1}{N^2} - \frac{N_{P_1}-1}{N_{P_1}\cdot N\cdot(N-1)}\right)\sum_{i=1}^{N}\sum_{j\neq i}Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right) \tag{7}$$

Starting from the definition of $S^2_{P_1,N}$, we write

$$\begin{aligned}
S^2_{P_1,N} &= \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i\left(P_1(i)\right)-\frac{1}{N}\sum_{j=1}^{N}Y_j\left(P_1(j)\right)\right)^2 \\
&= \frac{1}{N-1}\left(\left(\sum_{i=1}^{N}Y_i^2\left(P_1(i)\right)\right)-N\left(\frac{\sum_{i=1}^{N}Y_i(P_1(i))}{N}\right)^2\right) \\
&= \left(\frac{1}{N-1}-\frac{1}{(N-1)N}\right)\sum_{i=1}^{N}Y_i^2\left(P_1(i)\right)-\frac{1}{(N-1)N}\sum_{i=1}^{N}\sum_{j\neq i}Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right). \quad (8)
\end{aligned}$$

We compare expressions (7) and (8). We show that the coefficients for the terms $\sum_{i=1}^{N} Y_i^2\left(P_1(i)\right)$, $\sum_{i=1}^{N}\sum_{j\neq i} Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right)$ in Equation (7) equal the coefficients in Equation (8), scaled by $\frac{1}{N_{P_1}}\left(1-\frac{N_{P_1}}{N}\right)$.

Indeed, we have

$$\begin{aligned}
\frac{1}{N_{P_1}}\left(1-\frac{N_{P_1}}{N}\right)\left(\frac{1}{N-1}-\frac{1}{(N-1)N}\right) &= \left(\frac{1}{N_{P_1}}-\frac{1}{N}\right)\left(\frac{1}{N-1}-\frac{1}{(N-1)N}\right) \\
&= \frac{N-N_{P_1}}{N_{P_1}\cdot N^2} \\
&= \frac{1}{N_{P_1}\cdot N}-\frac{1}{N^2}.
\end{aligned}$$

We also have that

$$\frac{1}{N_{P_1}}\left(1-\frac{N_{P_1}}{N}\right)\frac{1}{(N-1)N}=\frac{N-N_{P_1}}{N_{P_1}\cdot N^2\cdot(N-1)},$$

which is equal to the respective coefficient of Equation (7), because

$$\frac{1}{N^2}-\frac{N_{P_1}-1}{N_{P_1}\cdot N\cdot(N-1)}=\frac{N_{P_1}(N-1)-(N_{P_1}-1)N}{N_{P_1}\cdot N^2\cdot(N-1)}=\frac{N-N_{P_1}}{N_{P_1}\cdot N^2\cdot(N-1)}.$$

We have thus shown that $\mathrm{Var}_W(\hat{y}_{P_1})=\frac{S^2_{P_1,N}}{N_{P_1}}\left(1-\frac{N_{P_1}}{N}\right)$.

B Evaluating a Single Policy Using the Proposed Approach - Unbiasedness

For $s \neq t$, we have

$$\begin{aligned}
\mathbb{E}_W\left[\hat{y}_{\underline{P_1}:s,P_2:t}\right] &= \frac{1}{N^{P_1}_{P_1:s,P_2:t}} \sum_{i \in g_{P_1:s,P_2:t}} \mathbb{E}_W[\mathbb{1}_{W_i=P_1}] \cdot Y_i^{obs} \\
&= \frac{1}{N^{P_1}_{P_1:s,P_2:t}} \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}} \sum_{i \in g_{P_1:s,P_2:t}} Y_i(s) \\
&= \frac{1}{N_{P_1:s,P_2:t}} \sum_{i \in g_{P_1:s,P_2:t}} Y_i(s) \\
&= y_{\underline{P_1}:s,P_2:t}.
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{E}_W\left[\hat{y}_{P_1:s,P_2:s}\right] &= \frac{1}{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s}} \sum_{i \in g_{P_1:s,P_2:s}} \mathbb{E}_W\left[\mathbb{1}_{W_i=P_1} + \mathbb{1}_{W_i=P_2}\right] \cdot Y_i^{obs} \\
&= \frac{1}{N^{P_1}_{P_1:s,P_2:s} + N^{P_2}_{P_1:s,P_2:s}} \left(\frac{N^{P_1}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}} + \frac{N^{P_2}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}} \right) \sum_{i \in g_{P_1:s,P_2:s}} Y_i(s) \\
&= \frac{1}{N_{P_1:s,P_2:s}} \sum_{i \in g_{P_1:s,P_2:s}} Y_i(s) \\
&= y_{P_1:s,P_2:s}.
\end{aligned}$$

We can therefore write

$$\mathbb{E}_W[\hat{y}_{P_1}] = \frac{\sum_{s,t} N_{P_1:s,P_2:t} \cdot \mathbb{E}_W\left[\hat{y}_{\underline{P_1}:s,P_2:t}\right]}{N} = \frac{\sum_{s,t} N_{P_1:s,P_2:t} \cdot y_{\underline{P_1}:s,P_2:t}}{N} = y_{P_1}.$$

C Evaluating a Single Policy Using the Proposed Approach - the Variance

Showing that, for $s \neq t$,

$$\mathrm{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right) = \frac{S^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} \left(1 - \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}}\right),$$

follows the exact same steps as the derivation in Appendix A, restricted to the segment $g_{P_1:s,P_2:t}$.

The derivation for the case of a segment in which the two policies recommend the same action is similar: a calculation shows that

$$\begin{aligned}
\mathrm{Var}_W\left(\hat{y}_{P_1:s,P_2:s}\right) &= \left(\frac{1}{\left(N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}\right)N_{P_1:s,P_2:s}}-\frac{1}{N^2_{P_1:s,P_2:s}}\right)\sum_{i\in g_{P_1:s,P_2:s}}Y_i^2(s) \\
&\quad -\left(\frac{1}{N^2_{P_1:s,P_2:s}}-\frac{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}-1}{\left(N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}\right)N_{P_1:s,P_2:s}\left(N_{P_1:s,P_2:s}-1\right)}\right) \\
&\qquad \cdot\sum_{i\in g_{P_1:s,P_2:s}}\sum_{j\neq i}Y_i(s)\cdot Y_j(s) \\
&= \frac{S^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}}\left(1-\frac{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}}\right).
\end{aligned}$$

D Comparing Two Policies Using the Standard Approach - the Variance

We start by writing

$$\mathrm{Var}_W\left(\hat{y}_{P_1-P_2}\right)=\mathrm{Var}_W\left(\hat{y}_{P_1}-\hat{y}_{P_2}\right)=\mathrm{Var}_W\left(\hat{y}_{P_1}\right)+\mathrm{Var}_W\left(\hat{y}_{P_2}\right)-2\cdot\mathrm{Cov}_W\left(\hat{y}_{P_1},\hat{y}_{P_2}\right).$$

We calculate the covariance. We have

$$\mathbb{E}_W\left[\hat{y}_{P_1}\cdot\hat{y}_{P_2}\right] = \mathbb{E}_W\left[\frac{1}{N_{P_1}}\sum_{i=1}^N\mathbb{1}_{W_i=P_1}Y_i^{obs}\cdot\frac{1}{N_{P_2}}\sum_{i=1}^N\mathbb{1}_{W_i=P_2}Y_i^{obs}\right].$$

We first calculate, for $i\neq j$,

$$\begin{aligned}
\mathbb{E}_W\left[\left(\mathbb{1}_{W_i=P_1}Y_i^{obs}\right)\cdot\left(\mathbb{1}_{W_j=P_2}Y_i^{obs}\right)\right] &= \Pr\left(W_i=P_1,W_j=P_2\right)\cdot Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right) \\
&= \frac{\binom{N-2}{N_{P_1}-1,N_{P_2}-1,N-N_{P_1}-N_{P_2}}}{\binom{N}{N_{P_1},N_{P_2},N-N_{P_1}-N_{P_2}}}\cdot Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right) \\
&= \frac{N_{P_1}N_{P_2}}{N(N-1)}\cdot Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right),
\end{aligned}$$

therefore

$$\mathbb{E}_W\left[\hat{y}_{P_1}\cdot\hat{y}_{P_2}\right]=\frac{1}{N(N-1)}\sum_{i=1}^N\sum_{j\neq 1}Y_i\left(P_1(i)\right)\cdot Y_j\left(P_1(j)\right).$$

We also calculate

$$\begin{aligned}
\mathbb{E}_W\left[\hat{y}_{P_1}\right] \cdot \mathbb{E}_W\left[\hat{y}_{P_2}\right] &= y_{P_1} \cdot y_{P_2} \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i\left(P_1(i)\right) \cdot \frac{1}{N}\sum_{i=1}^{N} Y_i\left(P_2(i)\right) \\
&= \frac{1}{N^2}\left[\sum_{i=1}^{N} Y_i\left(P_1(i)\right) \cdot Y_i\left(P_2(i)\right) + \sum_{i=1}^{N}\sum_{j\neq i} Y_i\left(P_1(i)\right) \cdot Y_j\left(P_2(j)\right)\right].
\end{aligned}$$

Therefore the covariance is

$$\begin{aligned}
\mathrm{Cov}_W\left(\hat{y}_{P_1}, \hat{y}_{P_2}\right) &= \mathbb{E}_W\left[\hat{y}_{P_1} \cdot \hat{y}_{P_2}\right] - \mathbb{E}_W\left[\hat{y}_{P_1}\right] \cdot \mathbb{E}_W\left[\hat{y}_{P_2}\right] \\
&= -\frac{1}{N^2}\sum_{i=1}^{N} Y_i\left(P_1(i)\right) \cdot Y_i\left(P_2(i)\right) + \frac{1}{N^2(N-1)}\sum_{i=1}^{N}\sum_{j\neq i} Y_i\left(P_1(i)\right) \cdot Y_j\left(P_2(j)\right). \quad (9)
\end{aligned}$$

Overall, we have

$$\begin{aligned}
\mathrm{Var}_W\left(\hat{y}_{P_1-P_2}\right) &= \mathrm{Var}_W\left(\hat{y}_{P_1}\right) + \mathrm{Var}_W\left(\hat{y}_{P_2}\right) - 2 \cdot \mathrm{Cov}_W\left(\hat{y}_{P_1}, \hat{y}_{P_2}\right) \\
&= \left(\frac{1}{N_{P_1}N} - \frac{1}{N^2}\right)\sum_{i=1}^{N} Y_i^2(P_1(i)) + \left(\frac{1}{N_{P_2}N} - \frac{1}{N^2}\right)\sum_{i=1}^{N} Y_i^2(P_2(i)) \\
&\quad - \left(\frac{1}{N^2} - \frac{N_{P_1}-1}{N_{P_1}N(N-1)}\right)\sum_{i=1}^{N}\sum_{j\neq i} Y_i\left(P_1(i)\right) \cdot Y_j\left(P_1(j)\right) \\
&\quad - \left(\frac{1}{N^2} - \frac{N_{P_2}-1}{N_{P_2}N(N-1)}\right)\sum_{i=1}^{N}\sum_{j\neq i} Y_i\left(P_2(i)\right) \cdot Y_j\left(P_2(j)\right) \\
&\quad + \frac{2}{N^2}\sum_{i=1}^{N} Y_i\left(P_1(i)\right) \cdot Y_i\left(P_2(i)\right) - \frac{2}{N^2(N-1)}\sum_{i=1}^{N}\sum_{j\neq i} Y_i\left(P_1(i)\right) \cdot Y_j\left(P_2(j)\right), \quad (10)
\end{aligned}$$

where the second equality follows from Equations (7) and (9).

We write

$$\begin{aligned}
\frac{S^2_{P_1,N}}{N_{P_1}} &= \frac{1}{N_{P_1}(N-1)}\sum_{i=1}^{N}\left(Y_i(P_1(i)) - \frac{1}{N}\sum_{j=1}^{N} Y_j(P_1(j))\right)^2 \\
&= \frac{1}{N_{P_1}(N-1)}\left[\left(1-\frac{1}{N}\right)\sum_{i=1}^{N} Y_i^2\left(P_1(i)\right) - \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} Y_i\left(P_1(i)\right) \cdot Y_j\left(P_1(j)\right)\right], \quad (11)
\end{aligned}$$

$$
\begin{aligned}
\frac{S^2_{P_2,N}}{N_{P_2}} &= \frac{1}{N_{P_2}(N-1)}\sum_{i=1}^{N}\left(Y_i(P_2(i)) - \frac{1}{N}\sum_{j=1}^{N}Y_j(P_2(j))\right)^2 \\
&= \frac{1}{N_{P_2}(N-1)}\left[\left(1-\frac{1}{N}\right)\sum_{i=1}^{N}Y_i^2\left(P_2(i)\right) - \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}Y_i\left(P_2(i)\right)\cdot Y_j\left(P_2(j)\right)\right], \qquad (12)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{S^2_{P_1,P_2,N}}{N} &= \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(Y_i(P_1(i)) - Y_i(P_2(i)) - \frac{1}{N}\sum_{j=1}^{N}\Big(Y_j(P_1(j)) - Y_j(P_2(j))\Big)\right)^2 \\
&= \frac{1}{N(N-1)}\Bigg[\sum_{i=1}^{N}Y_i^2(P_1(i)) + \sum_{i=1}^{N}Y_i^2(P_2(i)) + \frac{1}{N}\left(\sum_{i=1}^{N}Y_i(P_1(i))\right)^2 + \frac{1}{N}\left(\sum_{i=1}^{N}Y_i(P_2(i))\right)^2 \\
&\quad -2\cdot\sum_{i=1}^{N}Y_i(P_1(i))\cdot Y_i(P_2(i)) - 2\cdot\sum_{i=1}^{N}Y_i(P_1(i))\frac{\sum_{j=1}^{N}Y_j(P_1(j))}{N} \\
&\quad +2\cdot\sum_{i=1}^{N}Y_i(P_1(i))\frac{\sum_{j=1}^{N}Y_j(P_2(j))}{N} + 2\cdot\sum_{i=1}^{N}Y_i(P_2(i))\frac{\sum_{j=1}^{N}Y_j(P_1(j))}{N} \\
&\quad -2\cdot\sum_{i=1}^{N}Y_i(P_2(i))\frac{\sum_{j=1}^{N}Y_j(P_2(j))}{N} - \frac{2}{N}\left(\sum_{i=1}^{N}Y_i(P_1(i))\right)\left(\sum_{i=1}^{N}Y_i(P_2(i))\right)\Bigg] \\
&= \frac{1}{N(N-1)}\Bigg[\left(1-\frac{1}{N}\right)\left(\sum_{i=1}^{N}Y_i^2(P_1(i)) + \sum_{i=1}^{N}Y_i^2(P_2(i))\right) \\
&\quad -\frac{1}{N}\left(\sum_{i=1}^{N}\sum_{j\neq i}Y_i(P_1(i))\cdot Y_j(P_1(j)) + \sum_{i=1}^{N}\sum_{j\neq i}Y_i(P_2(i))\cdot Y_j(P_2(j))\right) \\
&\quad -\left(2-\frac{2}{N}\right)\sum_{i=1}^{N}Y_i(P_1(i))\cdot Y_i(P_2(i)) + \frac{2}{N}\sum_{i=1}^{N}\sum_{j\neq i}Y_i(P_1(i))\cdot Y_j(P_2(j))\Bigg], \qquad (13)
\end{aligned}
$$

We compare the terms of Equation (10) with the terms of

$$
\frac{S^2_{P_1,N}}{N_{P_1}} + \frac{S^2_{P_2,N}}{N_{P_2}} - \frac{S^2_{P_1,P_2,N}}{N}, \qquad (14)
$$

using Equations (11), (12), and (13).

We first compare the terms with $\sum_{i=1}^{N}Y_i^2(P_1(i))$. Expression (14) has this term multiplied by $\frac{1}{N_{P_1}(N-1)}\left(1-\frac{1}{N}\right) - \frac{1}{N(N-1)}\left(1-\frac{1}{N}\right)$. This is equal to $\frac{1}{N_{P_1}N} - \frac{1}{N^2}$, which is the multiplier of term $\sum_{i=1}^{N}Y_i^2(P_1(i))$ in expression (10).

The comparison for the terms with $\sum_{i=1}^{N}Y_i^2(P_2(i))$ is similar.

We now compare the terms with $\sum_{i=1}^{N}\sum_{j\neq i}Y_i(P_1(i))\cdot Y_j(P_1(j))$. Expression (10) has this term multiplied by $\frac{N_{P_1}-1}{N_{P_1}N(N-1)} - \frac{1}{N^2}$. This is equal to $-\frac{1}{N_{P_1}(N-1)}\frac{1}{N} + \frac{1}{N(N-1)}\frac{1}{N}$, which is the multiplier of term $\sum_{i=1}^{N}\sum_{j\neq i}Y_i(P_1(i))\cdot Y_j(P_1(j))$ in expression (14).

The comparison for the terms with $\sum_{i=1}^{N}\sum_{j\neq i} Y_i(P_2(i)) \cdot Y_j(P_2(j))$ is similar.

We now compare the terms with $\sum_{i=1}^{N} Y_i(P_1(i)) \cdot Y_i(P_2(i))$. Expression (14) has this term multiplied by $\frac{1}{N(N-1)}\left(2-\frac{2}{N}\right)$. This is equal to $\frac{2}{N^2}$, which is the multiplier of term $\sum_{i=1}^{N} Y_i(P_1(i)) \cdot Y_i(P_2(i))$ in expression (10).

We finally compare the terms with $\sum_{i=1}^{N}\sum_{j\neq i} Y_i(P_1(i)) \cdot Y_j(P_2(j))$. In both expressions (10) and (14), term $\sum_{i=1}^{N}\sum_{j\neq i} Y_i(P_1(i)) \cdot Y_j(P_2(j))$ is multiplied with $-\frac{2}{N^2(N-1)}$.

We have thus shown that

$$\mathrm{Var}_W\left(\hat{y}_{P_1-P_2}\right) = \frac{S^2_{P_1,N}}{N_{P_1}} + \frac{S^2_{P_2,N}}{N_{P_2}} - \frac{S^2_{P_1,P_2,N}}{N}.$$

E Comparing the Efficiency of the Standard and the Proposed Approaches

We first compare the estimated variances under each approach for single policy evaluation.

We start with the estimated variance under the proposed approach:

$$\begin{aligned}
&\sum_{\substack{s,t\\ s\neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} \left(1-\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}}\right) \\
&\quad + \sum_{s} \left(\frac{N_{P_1:s,P_2:s}}{N}\right)^2 \frac{s^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}} \cdot \left(1-\frac{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}}\right),
\end{aligned} \tag{15}$$

which we want to compare to the estimated variance under the standard approach,

$$\frac{s^2_{P_1,N}}{N_{P_1}}\left(1-\frac{N_{P_1}}{N}\right). \tag{16}$$

We assume a balanced experimental design, such that the proportion of customers assigned to a policy within a segment is the same as the proportion of customers assigned to that policy in any other segment, i.e.,

$$\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}} = \frac{N_{P_1}}{N}$$

for all s,t.

For $s\neq t$, we can write

$$\begin{aligned}
\left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} &= \left(\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}}\right)^2 \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} \\
&= \frac{1}{N_{P_1}} \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}} s^2_{P_1,g_{P_1:s,P_2:t}},
\end{aligned}$$

and for the segments where the two policies recommend the same action, we can write

$$
\begin{aligned}
\left(\frac{N_{P_1:s,P_2:s}}{N}\right)^2 \frac{s^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}} &= \left(\frac{N^{P_1}_{P_1:s,P_2:s}}{N_{P_1}}\right)^2 \frac{s^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}} \\
&< \left(\frac{N^{P_1}_{P_1:s,P_2:s}}{N_{P_1}}\right)^2 \frac{s^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s}} \\
&= \frac{1}{N_{P_1}} \frac{N^{P_1}_{P_1:s,P_2:s}}{N_{P_1}} s^2_{g_{P_1:s,P_2:s}}
\end{aligned}
$$

We have $1 - \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}} = 1 - \frac{N_{P_1}}{N}$, and $1 - \frac{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}}{N_{P_1:s,P_2:s}} = 1 - \frac{N_{P_1}+N_{P_2}}{N} < 1 - \frac{N_{P_1}}{N}$. We can now compare the expressions (15) and (16), after eliminating the adjustment for the size of the super-population.

$$
\begin{aligned}
&\sum_{\substack{s,t \\ s \neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} + \sum_{s} \left(\frac{N_{P_1:s,P_2:s}}{N}\right)^2 \frac{s^2_{g_{P_1:s,P_2:s}}}{N^{P_1}_{P_1:s,P_2:s}+N^{P_2}_{P_1:s,P_2:s}} \\
&\quad < \frac{1}{N_{P_1}} \left(\sum_{\substack{s,t \\ s \neq t}} \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}} s^2_{P_1,g_{P_1:s,P_2:t}} + \sum_{s} \frac{N^{P_1}_{P_1:s,P_2:s}}{N_{P_1}} s^2_{g_{P_1:s,P_2:s}} \right).
\end{aligned}
$$

If it holds that $s^2_{P_1,g_{P_1:s,P_2:t}} \leq s^2_{P_1,N}$ for all s,t with $s \neq t$, and $s^2_{g_{P_1:s,P_2:s}} \leq s^2_{P_1,N}$ for all s, then the previous expression is bounded above by

$$
\frac{1}{N_{P_1}} \sum_{s,t} \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}} s^2_{P_1,N} = \frac{s^2_{P_1,N}}{N_{P_1}},
$$

showing that the proposed method strictly improves the variance over the standard method.

We now compare the estimated variances under the standard and the proposed approach for comparing two policies. We start with the estimated variance under the proposed approach

$$
\sum_{\substack{s,t \\ s \neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \left(\frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} + \frac{s^2_{P_2,g_{P_1:s,P_2:t}}}{N^{P_2}_{P_1:s,P_2:t}} \right),
$$

which can be written as

$$
\begin{aligned}
&\textstyle\sum_{\substack{s,t \\ s \neq t}} \left(\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}}\right)^2 \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_1}_{P_1:s,P_2:t}} + \sum_{\substack{s,t \\ s \neq t}} \left(\frac{N^{P_2}_{P_1:s,P_2:t}}{N_{P_2}}\right)^2 \frac{s^2_{P_2,g_{P_1:s,P_2:t}}}{N^{P_2}_{P_1:s,P_2:t}} \\
&\quad = \textstyle\frac{1}{N_{P_1}} \sum_{\substack{s,t \\ s \neq t}} \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}} s^2_{P_1,g_{P_1:s,P_2:t}} + \frac{1}{N_{P_2}} \sum_{\substack{s,t \\ s \neq t}} \frac{N^{P_2}_{P_1:s,P_2:t}}{N_{P_2}} s^2_{P_2,g_{P_1:s,P_2:t}},
\end{aligned}
$$

where we have used the assumption of a balanced experimental design:

$$\frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}} = \frac{N^{P_2}_{P_1:s,P_2:t}}{N_{P_2}} = \frac{N_{P_1:s,P_2:t}}{N}.$$

If it holds that $s^2_{P_1,g_{P_1:s,P_2:t}} \leq s^2_{P_1,N}$ and $s^2_{P_2,g_{P_1:s,P_2:t}} \leq s^2_{P_2,N}$ for all s,t with $s \neq t$, then the expression for the variance is strictly less than

$$\frac{1}{N_{P_1}} \sum_{s,t} \frac{N^{P_1}_{P_1:s,P_2:t}}{N_{P_1}} s^2_{P_1,N} + \frac{1}{N_{P_2}} \sum_{s,t} \frac{N^{P_2}_{P_1:s,P_2:t}}{N_{P_2}} s^2_{P_2,N} = \frac{s^2_{P_1,N}}{N_{P_1}} + \frac{s^2_{P_2,N}}{N_{P_2}},$$

which is the estimated variance under the standard approach.

F Randomizing by Action — Evaluating a Single Policy Using the Proposed Approach

We use P_s to denote the policy that assigns action s, regardless of the customers' covariates. We want to evaluate a targeting policy P_1 without implementing it, using the outcomes from the experimental conditions pertaining to the actions.

Our goal (the estimand) is the population-level measure

$$y_{P_1} = \frac{1}{N} \sum_{i=1}^{N} Y_i \left(P_1(i)\right),$$

which can also be written as

$$y_{P_1} = \frac{\sum_{s,t} \sum_{i \in g_{P_1:s,P_2:t}} Y_i(s)}{\sum_{s,t} N_{P_1:s,P_2:t}} = \frac{\sum_{s,t} N_{P_1:s,P_2:t} \cdot y_{\underline{P_1}:s,P_2:t}}{N},$$

where $y_{\underline{P_1}:s,P_2:t} = \frac{\sum_{i \in g_{P_1:s,P_2:t}} Y_i(s)}{N_{P_1:s,P_2:t}}$.

We estimate y_{P_1} by calculating the outcome in each segment and then aggregating across segments:

$$\hat{y}_{P_1} = \frac{\sum_{s,t} N_{P_1:s,P_2:t} \cdot \hat{y}_{\underline{P_1}:s,P_2:t}}{N},$$

with

$$\hat{y}_{\underline{P_1}:s,P_2:t} = \frac{\sum_{i \in g_{P_1:s,P_2:t}} \mathbb{1}_{W_i = P_s} Y^{obs}_i}{N^{P_s}_{P_1:s,P_2:t}},$$

where $N^{P_s}_{P_1:s,P_2:t}$ is the number of customers in segment $g_{P_1:s,P_2:t}$ that are randomly assigned to receive policy P_s. This estimator is at the heart of the proposed alternative experimental design: to evaluate a policy on a segment of customers to which the policy recommends a specific action, we look at customers within the segment that were hit with that specific action.

It is straightforward to establish that $\hat{y}_{P_1}$ is an unbiased estimator of y_{P_1}. We can also write expressions for the variance of the estimator $\hat{y}_{P_1}$ within each segment:

$$\mathrm{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right) = \frac{S^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_s}_{P_1:s,P_2:t}}\left(1 - \frac{N^{P_s}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}}\right),$$

where $S^2_{P_1,g_{P_1:s,P_2:t}}$ is given by Equation (3). We estimate $\mathrm{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right)$ by

$$\widehat{\mathrm{Var}}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right) = \frac{s^2_{P_1,g_{P_1:s,P_2:t}}}{N^{P_s}_{P_1:s,P_2:t}}\left(1 - \frac{N^{P_s}_{P_1:s,P_2:t}}{N_{P_1:s,P_2:t}}\right),$$

where

$$s^2_{P_1,g_{P_1:s,P_2:t}} = \frac{1}{N^{P_s}_{P_1:s,P_2:t} - 1} \sum_{i \in g_{P_1:s,P_2:t}, W_i = P_s} \left(Y_i^{obs} - \hat{y}_{\underline{P_1}:s,P_2:t}\right)^2.$$

Overall, the variance of estimator $\hat{y}_{P_1}$ across all segments is

$$\mathrm{Var}_W(\hat{y}_{P_1}) = \sum_{s,t}\left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \mathrm{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right),$$

which we estimate by

$$\widehat{\mathrm{Var}}_W(\hat{y}_{P_1}) = \sum_{s,t}\left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \widehat{\mathrm{Var}}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t}\right).$$

Next we describe an estimator for estimating the difference between two policies.

G Randomizing by Action — Comparing Two Policies Using the Proposed Approach

Our target estimand is

$$y_{P_1-P_2} = y_{P_1} - y_{P_2} = \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i\left(P_1(i)\right) - Y_i\left(P_2(i)\right)\Big),$$

which, accounting for the segmentation of the customers according to the recommended actions, can also be written as

$$\frac{\sum_{s,t}\sum_{i\in g_{P_1:s,P_2:t}}\Big(Y_i(s) - Y_i(t)\Big)}{\sum_{s,t} N_{P_1:s,P_2:t}} = \frac{\sum_{\substack{s,t\\ s\neq t}}\sum_{i\in g_{P_1:s,P_2:t}}\Big(Y_i(s) - Y_i(t)\Big)}{\sum_{s,t} N_{P_1:s,P_2:t}} = \frac{\sum_{\substack{s,t\\ s\neq t}} N_{P_1:s,P_2:t}\left(y_{\underline{P_1}:s,P_2:t} - y_{P_1:s,\underline{P_2}:t}\right)}{N},$$

where, for $s \neq t$,

$$y_{\underline{P_1}:s,P_2:t} = \frac{\sum_{i \in g_{P_1:s,P_2:t}} Y_i(s)}{N_{P_1:s,P_2:t}}, \qquad y_{P_1:s,\underline{P_2}:t} = \frac{\sum_{i \in g_{P_1:s,P_2:t}} Y_i(t)}{N_{P_1:s,P_2:t}}.$$

We estimate $y_{P_1 - P_2}$ with the unbiased estimator

$$\hat{y}_{P_1 - P_2} = \frac{\sum_{\substack{s,t \\ s \neq t}} N_{P_1:s,P_2:t} \left(\hat{y}_{\underline{P_1}:s,P_2:t} - \hat{y}_{P_1:s,\underline{P_2}:t} \right)}{N},$$

where, for $s \neq t$,

$$\hat{y}_{\underline{P_1}:s,P_2:t} = \frac{\sum_{i \in g_{P_1:s,P_2:t}} \mathbb{1}_{W_i = P_s} Y_i^{obs}}{N^{P_s}_{P_1:s,P_2:t}}, \qquad \hat{y}_{P_1:s,\underline{P_2}:t} = \frac{\sum_{i \in g_{P_1:s,P_2:t}} \mathbb{1}_{W_i = P_t} Y_i^{obs}}{N^{P_t}_{P_1:s,P_2:t}}.$$

In segments for which the two policies recommend the same action, the true difference in the outcome between the two policies is zero. In segments for which the two policies recommend different actions ($s \neq t$), the variance of the difference between the two policies is given by

$$\mathrm{Var}_W \left(\hat{y}_{\underline{P_1}:s,P_2:t} - \hat{y}_{P_1:s,\underline{P_2}:t} \right) = \frac{S^2_{P_1, g_{P_1:s,P_2:t}}}{N^{P_s}_{P_1:s,P_2:t}} + \frac{S^2_{P_2, g_{P_1:s,P_2:t}}}{N^{P_t}_{P_1:s,P_2:t}} - \frac{S^2_{P_1,P_2, g_{P_1:s,P_2:t}}}{N_{P_1:s,P_2:t}},$$

where variances $S^2_{P_1, g_{P_1:s,P_2:t}}, S^2_{P_2, g_{P_1:s,P_2:t}}, S^2_{P_1,P_2, g_{P_1:s,P_2:t}}$ are given by Equations (3) and (6). We estimate $S^2_{P_1, g_{P_1:s,P_2:t}}$ and $S^2_{P_2, g_{P_1:s,P_2:t}}$ with

$$s^2_{P_1, g_{P_1:s,P_2:t}} = \frac{1}{N^{P_s}_{P_1:s,P_2:t} - 1} \sum_{i \in g_{P_1:s,P_2:t}, W_i = P_s} \left(Y_i^{obs} - \hat{y}_{\underline{P_1}:s,P_2:t} \right)^2$$

and

$$s^2_{P_2, g_{P_1:s,P_2:t}} = \frac{1}{N^{P_t}_{P_1:s,P_2:t} - 1} \sum_{i \in g_{P_1:s,P_2:t}, W_i = P_t} \left(Y_i^{obs} - \hat{y}_{P_1:s,\underline{P_2}:t} \right)^2,$$

respectively.

The term $S^2_{P_1,P_2, g_{P_1:s,P_2:t}}$ is in general impossible to estimate empirically because we never observe the outcome of both actions s, t for the same customer. We use the Neyman variance estimator

$$\mathrm{Var}^{\mathrm{Neyman}}_{g_{P_1:s,P_2:t}} = \frac{s^2_{P_1, g_{P_1:s,P_2:t}}}{N^{P_s}_{P_1:s,P_2:t}} + \frac{s^2_{P_2, g_{P_1:s,P_2:t}}}{N^{P_t}_{P_1:s,P_2:t}},$$

which is upwardly biased.

Overall, the variance of estimator $\hat{y}_{P_1-P_2}$ across all segments is

$$\mathrm{Var}_W\left(\hat{y}_{P_1-P_2}\right) = \sum_{\substack{s,t\\ s\neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \mathrm{Var}_W\left(\hat{y}_{\underline{P_1}:s,P_2:t} - \hat{y}_{P_1:s,\underline{P_2}:t}\right),$$

which we estimate with

$$\mathrm{Var}^{\mathrm{Neyman}}_{P_1-P_2} = \sum_{\substack{s,t\\ s\neq t}} \left(\frac{N_{P_1:s,P_2:t}}{N}\right)^2 \mathrm{Var}^{\mathrm{Neyman}}_{g_{P_1:s,P_2:t}}.$$

This is upwardly biased and, as a result, confidence intervals will be conservative.