Targeting for long-term outcomes*

Jeremy Yang[1], Dean Eckles[1], Paramveer Dhillon[2], and Sinan Aral[1]

[1]Massachusetts Institute of Technology

[2]University of Michigan

Working Paper

October 2020

Abstract

Decision-makers often want to target interventions (e.g., marketing campaigns) so as to maximize an outcome that is observed only in the long-term. This typically requires delaying decisions until the outcome is observed or relying on simple short-term proxies for the long-term outcome. Here we build on the statistical surrogacy and off-policy learning literature to impute the missing long-term outcomes and then approximate the optimal targeting policy on the imputed outcomes via a doubly-robust approach. We apply our approach in large-scale proactive churn management experiments at *The Boston Globe* by targeting optimal discounts to its digital subscribers to maximize their long-term revenue. We first show that conditions for validity of average treatment effect estimation with imputed outcomes are also sufficient for valid policy evaluation and optimization; furthermore, these conditions can be somewhat relaxed for policy optimization. We then validate this approach empirically by comparing it with a policy learned on the ground truth long-term outcomes and show that they are statistically indistinguishable. Our approach also outperforms a policy learned on short-term proxies for the long-term outcome. In a second field experiment, we implement the optimal targeting policy with additional randomized exploration, which allows us to update the optimal policy for each new cohort of customers to account for potential non-stationarity. Over three years, our approach had a net-positive revenue impact in the range of $4-5 million compared to *The Boston Globe*'s current policies.

*Correspondence should be addressed to Jeremy Yang (yangzhen@mit.edu) and Dean Eckles (eckles@mit.edu). This research was supported in part by a grant from Boston Globe Media. We thank Boston Globe Media and particularly Jessica Bielkiewicz, Thomas Brown, Ryan McVeigh, and Shannon Rose for their partnership in conducting the field experiments. This work benefited from comments by Susan Athey, John Hauser, Günter Hitsch, Duncan Simester, participants in seminars at MIT, the Harvard Business School Digital Doctoral Workshop, and the NeurIPS CausalML Workshop.

1 Introduction

Advertising revenues have been stagnating for newspapers in recent years.[1] As a consequence, newspapers are looking for ways to strengthen their subscription-based business model. Take the New York Times as an example: in 2019, their total subscription revenue was twice their total advertising revenue (Figure A.1, A.2). Their CEO recently said: "... we still regard advertising as an important revenue stream, but we believe that our focus on establishing close and enduring relationships with paying, deeply engaged subscribers, and the long-range revenues which flow from those relationships, is the best way of building a successful and sustainable news business".[2] Hence, to succeed in a subscription-based business model, news publishers must retain their existing subscribers and maximize their long-term values. However, they also work to acquire new subscribers, often at highly-discounted prices, who then may be a great risk of churning after that introductory period (cf. Datta et al., 2015). A common approach to achieving this goal is to target existing subscribers with marketing interventions, such as price discounts or other personalized offers. Publishers are hardly the only firms that care about optimizing long-term customer outcomes. Most firms that monetize through subscription models fall into this category. Even more generally, decision-makers in business, medicine and public health, and government typically care about outcomes that are only observed over the long-term.

"Long-term" and "short-term" outcomes are fruitfully understood as defined relative to the targeting cycle. For example, if a firm runs a campaign every year, then all outcomes that are observed within a year, such as their 1-year revenue, are considered "short-term" because these outcomes are observed before the firm takes action (decides whom to target with what) in their next campaign. Hence, future policies can be optimized on these observed outcomes. In contrast, "long-term" outcomes materialize over time horizons longer than the window of opportunity for action, for example, three-year or five-year revenue, rendering the firm incapable of optimizing their next campaign based on them. So, a natural question arises: How can firms learn and implement an optimal targeting policy when the primary outcome of interest is "long-term"?

A straightforward solution to this problem is to wait until the long-term outcome materializes and choose a policy based on the realized long-term outcome. But this implies that the firm can not learn anything in the meantime, and therefore is unable to implement the optimal policy until years later. Another solution is to find a short-term proxy (e.g., short-term revenue) for the long-term outcome and optimize for it instead. However, this could be problematic as the proxy and the long-term outcome might not be well aligned. Hence, a policy that performs well on the proxy might not perform well in the long-run. To further complicate things, firms almost always operate in a changing environment. Subscribers may differ in their characteristics from one cohort to the next and in the way they respond to the marketing interventions over time. Therefore, a static

[1] The print advertising revenue is declining with a compound annual growth rate (CAGR) of -12.6% from 2016-2021, while digital ads revenue is still growing at a CAGR of 2.2%, it's not enough to compensate for the loss in print. Source: US Online and Traditional Media Advertising Outlook

[2] Source: https://www.nytimes.com/2018/02/08/business/new-york-times-company-earnings.html

targeting policy optimized on one customer cohort may not remain optimal for future cohorts due to this non-stationarity in the environment.

In this paper we propose to use surrogates (Prentice, 1989; VanderWeele, 2013) to impute the missing long-term outcomes and use the imputed long-term outcomes to optimize a targeting policy. We estimate the missing long-term outcome as the expectation of the long-term outcome conditional on surrogates of that outcome in a historical dataset in which the long-term outcome was observed. Surrogate index estimators combine multiple surrogates in the estimation (Xu and Zeger, 2001; Athey et al., 2019). Once we have the imputed long-term outcomes, we optimize the targeting policy efficiently by using a doubly-robust approach (Dudík et al., 2014; Athey and Wager, 2020; Zhou et al., 2018) on the imputed long-term outcomes. We prove analytically that this approach recovers the optimal policy learned on true long-term outcomes under certain assumptions. We implement the optimal policy via bootstrapped Thompson sampling (Eckles and Kaptein, 2014; Osband et al., 2016) to maintain exploration so we can update and re-optimize the policy for every new cohort of customers to allow for potential non-stationarity.

We evaluate the efficacy of our approach empirically by running two large-scale field experiments that target discounts to the digital subscribers of *The Boston Globe*, a regional leader in news media. Boston Globe Media, which operates *The Boston Globe* newspaper and associated websites, is facing a similar problem to many other publishers. Our goal is to learn an optimal targeting policy that treats some subscribers with certain discounts to maximize their retention and long-term revenue. Here a policy is a mapping from subscriber characteristics to a specific price or discount (or a distribution over them when the policy is stochastic). In this subscriber retention context, this is also known as proactive churn management.[3] To construct the surrogate index, we use the observed revenue and content consumption 1-6 months after treatment as our surrogates. We compare how well the policies learned using surrogate index perform against policies determined directly on short-term proxies or surrogates (benchmark) or realized long-term outcomes (the ground truth), we also consider alternative selections of surrogates for the construction of surrogate index. Our approach increases the firm's total projected digital subscription revenue by \$4-5 million over a three-year period relative to the status quo in the two experiments.

The rest of the paper is organized as follows. In Section 2 we review related work. The empirical context is described in Section 3. We explain the imputation of the long-term outcome using the surrogate index and prove sufficient conditions for it to be valid for policy evaluation and optimization in Section 4. Then we describe the policy learning framework and how it is implemented in Section 5. Section 6 discusses the empirical validation of our approach and experimental results are reported in Section 7. We conclude in Section 8.

[3]Proactive simply means that the intervention (discount) happens before a churn intention is observed, by contrast, reactive churn management means that the company first waits for customers to request to cancel their subscription then offers some discount or other benefits in reaction to this in the hope of retaining them. One analogy is that the proactive approach is like diagnosing and preventing illness before the patient shows strong symptoms, and the reactive approach is like treating patients who are already ill.

2 Related Work

Our paper builds on a large body of literature in biostatistics and medicine on surrogate outcomes (i.e., endpoints, biomarkers); see, e.g., Joffe and Greene (2009) and Weir and Walley (2006) for reviews. In clinical trials the goal is often to study the efficacy of an intervention on outcomes such as the long-term health or survival rate of patients. However, the primary outcome of interest might be very rare or only observed after years of delay (e.g., a 5 or 10-year survival rate). It is common to use the effect of an intervention on surrogate outcomes as a proxy for its effect on long-term outcomes. In a seminal paper, Prentice (1989) argued that to be a valid surrogate, treatment and outcome have to be independent conditional on the surrogate. One intuitive way for this and, critically, stronger conditions to be satisfied is if the surrogate fully mediates the treatment effect.[4] In practice it is hard to find a single variable that plausibly satisfies the condition (Freedman et al., 1992), but Xu and Zeger (2001) showed that combining multiple surrogates to predict the outcome can be preferable to using a single surrogate because the treatment effect may operate through multiple pathways and, even when there is a single pathway, using multiple surrogates can reduce measurement error. This idea is further developed in a recent paper in econometrics (Athey et al., 2019), where the combination is referred to as a surrogate index. This literature focuses on using surrogates to identify treatment effects on long-term outcomes and, in this paper, we extend this to the optimization of targeting policies.

Another popular approach to modeling long-term outcomes is to posit a particular parametric generative model for the long-term outcomes. In the context of marketing, this is typically a model of customer lifetime value (CLV or LTV). CLV models are widely used in marketing for customer segmentation and targeting; see Gupta et al. (2006) and Fader and Hardie (2007) for surveys. CLV is defined as the sum of discounted future revenues or profits from a customer. To calculate CLV we typically need to posit a parametric survival function and extrapolate the survival or retention probability into the future. A recent example in the context of churn management is Godinho de Matos et al. (2018), where a parametric survival function is used. One advantage of this approach is that we can apply it even when the long-term outcomes are never observed[5] because the prediction is based on functional form assumptions, unlike the surrogate index approach which needs access to long-term outcomes in a historical dataset; on the other hand, standard parametric CLV approachs may suffer from model misspecification. Furthermore, building a CLV model may require substantial work to formalize business logic in anything but the simplest subscription businesses. A synthesis of these approaches is also possible in that a CLV prediction, if already available, can also be used as one of the surrogates in the construction

[4] This can also be described as an exclusion restriction, as in instrumental variables. Like that case this assumption has both testable and untestable implications. It might be tempting to regress the outcome on surrogate and treatment and test if the coefficient of treatment is zero. This naive test is not valid when there are unobserved confounders for the surrogate and outcome, conditioning on the surrogate or a "collider" in such a case will generate spurious correlation between treatment and confounder, and hence between treatment and outcome. See Joffe and Greene (2009) for a more detailed discussion.

[5] The model is often assumed to be infinite horizon.

of a surrogate index.

This paper is also related to the literature on targeting policy evaluation and optimization, which has recently developed within marketing research. Hitsch and Misra (2018) proposed a direct estimation method for conditional average treatment effect (CATE) based on k-nearest neighbors (kNN) and used it for policy optimization. Simester et al. (2019a) showed that we can compare targeting policies more efficiently if we only compare the outcome of subscribers on whom the policies prescribe different actions. Simester et al. (2019b) documented non-stationarity such as covariate and concept shifts between two experiments and evaluated how robust different machine learning models used to optimize policies are to these changes in the environment. Yoganarasimhan et al. (2020) used different machine learning models to estimate CATE and evaluated how targeting policies constructed using these models perform against each other. In another recent work, Lemmens and Gupta (2020) examine using a CLV model combined with field experimentation to optimize targeting in the policy learning framework.

Our work complements this literature by addressing an orthogonal problem and is novel in a few ways. First, we focus directly on targeting for long-term outcomes; outcomes used in these other works are short-term (in the sense that they are observable when we optimize and implement the policy) or extrapolation is done using a parametric CLV model.[6] Second, we systematically add randomized exploration around the learned policy, which allows us to evaluate and update the policy for future cohorts in case the environment changes. Hitsch and Misra (2018) and Yoganarasimhan et al. (2020) studied the problem in a static setting. Simester et al. (2019b) did look at changes in the environment but they focused on evaluating the robustness of different machine learning models. Third, we use a doubly-robust (DR) approach (Dudík et al., 2014) for both policy evaluation and learning in contrast to Hitsch and Misra (2018) and Yoganarasimhan et al. (2020) who used an inverse probability weighting (IPW) estimator for policy evaluation. Lemmens and Gupta (2020) introduce a specialized incremental-profit-based loss function that performs well in their empirical evaluation, but lacks the asymptotic efficiency results available for doubly-robust policy learning; it is also unclear how to combine this with known probabilities of treatment (i.e., design-based propensity scores) that arise in sophisticated experiments. In particular, even when probabilities of treatment are known exactly (as in our setting), DR estimators have advantages in statistical efficiency compared with IPW estimators (Athey and Wager, 2020; Zhou et al., 2018).

Substantively, our study adds to the literature on proactive churn management. Earlier work focused on developing better prediction algorithms to more accurately identify potential churners. Neslin et al. (2006) provides a detailed comparison of different churn prediction models. Recently, the literature started to look into the causal effect of targeting interventions on churn using field experiments. For example, Ascarza (2018) and Lemmens and Gupta (2020) note that firms should not target customers based on their outcome level (churn risk) but should target based on treatment effects. Ascarza et al. (2016) showed evidence from a field experiment with a telecom-

[6] Yoganarasimhan et al. (2020) showed in their particular case the policy learned on short-term outcome also does well on long-term outcomes, but the policy is not directly optimized on long-term outcome.

munication company that proactive churn interventions can backfire and increase the churn rate in practice. They argued that this is because proactive intervention lowers customers' inertia to switch plans and increases the salience of past-usage patterns among potential churners. Our paper contributes to this literature by proposing an experimental framework that can be applied to directly optimize targeting policies for long-term customer retention and revenues.

3 Empirical Context

Founded in 1872, *The Boston Globe* is the oldest and largest daily newspaper in the greater Boston area. It has won a total of 26 Pulitzer Prizes and is widely regarded as one of the most prestigious papers in the US. We ran targeting experiments on all digital only[7] subscribers of *The Boston Globe* in two cohorts. Our analysis is of a random sample of about 45K digital subscribers in the first cohort and 95K in the second. For each subscriber we observed the short-term outcome (e.g., monthly churn and revenue) and three sets of features: demographics (e.g., zip code), account activities (e.g., billing address change, credit card expiration date, complaints), and content consumption (e.g., when and what articles they read). There was only one intervention in the first cohort, which lowered the price for treated subscribers from \$6.93 per week to \$4.99 per week for 8 weeks. Approximately 1,000 subscribers were treated in the first cohort. An email (Figure B.1a) was sent to all treated subscribers in August 2018 telling them that a discount had been automatically applied to their accounts. We implemented 6 interventions in the second cohort: a thank you email, a \$20 gift card, a discount to \$5.99 for 8 weeks, a discount to \$5.99 for 4 weeks, a discount to \$4.99 for 8 weeks (the same as the intervention in the first cohort), and a discount to \$3.99 for 8 weeks. About 6,000 subscribers were treated in the second cohort, with about 1,000 subscribers assigned to each of these conditions uniformly at random conditional on being assigned to treatment. A similar email (Figure B.1b) was sent to all treated subscribers in July 2019 with the corresponding message, and a treated subscriber had to click on a button at the bottom of the email to redeem the benefit. There was no overlap of treated subscribers between the two cohorts. All results in the paper are from intent-to-treat (ITT) analyses that do not condition on potentially endogenous post-treatment behaviors, such as opening the email or redeeming the benefit.

4 Imputing a Long-term Outcome with a Surrogate Index

We first introduce the notation that we use throughout the paper: let $\pi \in \Pi$ be a targeting policy that maps from the space of subscriber (or, more generally, unit) characteristics $\mathbb{X}$ to a space of distributions (simplex) over a set of discrete actions $\mathbb{A}$ (we index actions by $\{0, 1, 2, .., K-1\}$, where 0 is control and others are different interventions). When the policy is non-degenerate, it defines a probability distribution over possible actions conditional on covariates $\pi(a|x) := \mathbb{P}(A = a|X = x), \forall a \in \mathbb{A}, x \in \mathbb{X}$. When it is degenerate, it maps to a fixed action with probability 1. The

[7] *The Globe* also has a combined print and digital subscription. All subscribers are paying customers.

goal is to learn a policy that maximizes some average long-term outcome Y over a population of n units.[8]

Definition 1. *A Policy and its Value*

$$\pi : \mathbb{X} \to \Delta(\mathbb{A}) \tag{1}$$

$$V(\pi) := \mathbb{E}[Y_i(x_i, \pi(x_i))] \tag{2}$$

Definition 2. *Optimal Policy*

$$\pi^* := \underset{\pi}{\operatorname{argmax}}\, V(\pi) \tag{3}$$

In our application, the primary outcome of interest is long-term subscriber retention or revenue[9], but we do not observe these outcomes in the short-term, i.e., after the intervention in the first cohort and before we implemented the learned policy for the second cohort of customers. Hence, we use a surrogate index to address this problem. This entails using intermediate outcomes that are observed over the short-term period following the intervention, such as a subscriber's content consumption on the newspaper's website and short-term revenue.[10] These surrogate variables are then combined with the long-term outcomes in the historical data to impute missing long-term outcomes for subscribers in the experiment. Assume we have two datasets, one from the experiment labeled E and one based on historical (observational) data labeled H. We observe draws of the tuple (X, A, S) in the experiment where $X \in \mathbb{X}$ represents the subscriber characteristics, $A \in \mathbb{A}$ is the action (i.e., treatments, interventions) and $S \in \mathbb{S}$ is the potentially vector valued set of intermediate outcomes or surrogates. Note that we don't observe the long-term outcome Y in the experiment. In the historical dataset, we observe draws of the tuple (X, S, Y); note that there was no intervention in this dataset (i.e., it is observational), but the long-term outcome Y is observed. We can define a surrogate index $\tilde{Y}$ for the long-term outcome Y as the expectation of the long-term outcome conditional on subscriber characteristics and surrogates in the historical dataset H:[11]

Definition 3. *Surrogate Index*

$$\tilde{Y}_i := \mathbb{E}_H[Y_i | S_i, X_i] \tag{4}$$

Under Assumption 1-3 listed below, a central result in Athey et al. (2019) is that the average treatment effect (ATE) on $\tilde{Y}$ recovers the ATE on long-term outcome Y. That is, by constructing the surrogate index we can identify and feasibly estimate the ATE on some long-term outcomes without having to wait until they are observed.

[8] This definition can be modified to be interpretable with finite populations if $\mathbb{E}$ is understood as $\frac{1}{n}\sum_{i=1}^{n}$.

[9] Being a digital service, marginal costs are negligible compared with subscription revenue.

[10] These intermediate outcomes are known as surrogates or proxies for their instrumental value in predicting the long-term outcome of interest.

[11] One advantage of this approach is that the estimation of the conditional expectation can be treated as a supervised learning problem and can be performed using flexible non-parametric machine learning methods like XGBoost (Chen et al., 2015).

Assumption 1. *Regular treatment assignment mechanism (Ignorability and Positivity): The treatment assignment is conditionally independent of potential long-term outcomes (Ignorabillity) and all units have positive probability of being assigned to each action (Positivity) in the experimental dataset.*

$$A_i \perp\!\!\!\perp (Y_i(a), S_i(a)) | X_i \; \forall a \in \mathbb{A}, i \in E \tag{5}$$

$$0 < \pi(a|x) < 1 \; \forall a \in \mathbb{A}, x \in \mathbb{X} \tag{6}$$

Assumption 2. *Surrogacy: The treatment assignment is independent of long-term outcomes conditional on the surrogates in the experimental dataset.*

$$A_i \perp\!\!\!\perp Y_i \mid S_i, X_i, i \in E \tag{7}$$

Surrogacy is implied by a generative model in which the set of surrogates fully mediate the casual effects from treatment to the long-term outcome (cf. Lauritzen, 2004). In our context, it means the effect of price discounts on retention and revenue should occur via some intermediate outcomes we observe, e.g., content consumption and short-term revenue.

Assumption 3. *Comparability: The distribution of the long-term outcome conditional on the covariates and surrogates is the same across the experimental and historical datasets.*

$$Y_i \mid S_i, X_i, i \in E \sim Y_i \mid S_i, X_i, i \in H \tag{8}$$

In our case, this assumption implies that the distribution of long-term retention and revenue (conditional on content consumption and short-term retention and revenue) should be the same between the experimental and historical datasets. Note that under comparability assumption we have:

$$\tilde{Y_i} = \mathbb{E}_H[Y_i|S_i, X_i] = \mathbb{E}_E[Y_i|S_i, X_i] \tag{9}$$

Assumption 1 is satisfied because the price discounts are randomly assigned conditional on subscriber characteristics according to the design policy. Assumption 2 is the key assumption and, while it may have some testable implications, is not directly testable. It is more plausible if we have a rich set of surrogates, something that is more likely in our setting as publishers now observe how their content is being consumed. To make Assumption 3 more plausible, we use the most recent historical data to do the estimation; that is, for the experiment run in 2018 we used the observed revenue data from 2015–2018 to estimate the 3-year revenue for subscribers in the experiment.[12]

Given these assumptions, we prove that the surrogate index is valid for policy evaluation and optimization. Policy evaluation is the estimation of $V(\pi)$ for a given policy π. Policy optimization is finding a π that maximizes $V(\pi)$. See Section 5 for more details about doing so in finite samples; here we simply consider the optimal policy defined on the population. We show that the value of

[12]We can also directly test for this after the long-term outcomes in the experiment are realized, but not before.

a policy with respect to surrogate index is identical to its value on the long-term outcome; this in turn implies that the optimal policy with respect to the surrogate index coincides with that optimal policy with respect to long-term outcomes. We state the main results here and the proofs are in Appendix C.

Let $\tilde{V}(\pi)$ denote the value of π with respect to $\tilde{Y}$ rather than Y.

Proposition 1. *Under Assumption 1-3, policy evaluation conducted on surrogate index identifies the true policy value defined on long-term outcomes.*

$$\tilde{V}(\pi) = V(\pi) \; \forall \pi \in \Pi \tag{10}$$

Since the function being maximized is identical at all points, it is also identical at its maximum.

Proposition 2. *Under Assumption 1-3, policy optimization conducted on surrogate index recovers the true optimal policy.*

$$\underset{\pi}{\operatorname{argmax}} \, \tilde{V}(\pi) = \underset{\pi}{\operatorname{argmax}} \, V(\pi) \tag{11}$$

Proposition 1 and 2 are analytical results that could justify our empirical application. However, somewhat weaker assumptions are in fact sufficient for Proposition 2 than for results for estimation of the ATE or CATEs. Let $\tau_a(x) = \mathbb{E}_E[Y(a) - Y(0)|X = x])$ and $\tilde{\tau}_a(x) = \mathbb{E}_E[\tilde{Y}(a) - \tilde{Y}(0)|X = x])$. When Assumption 2 (Surrogacy) is violated (the set of surrogates doesn't fully mediate the treatment effect on long-term outcomes), the CATE estimated using surrogate index can be biased (even with infinite data). That is, $\tau_a(x) \neq \tilde{\tau}_a(x)$ for some $x \in \mathbb{X}$. Here our aim is not estimating CATEs, but simply optimizing the policy. Bias in CATEs (i.e., non-zero $\tau_a(x) - \tilde{\tau}_a(x)$) doesn't result in a loss in the value of the optimized policy unless the bias changes the sign of that CATE.[13]

Thus, we can introduce a somewhat weaker version of Assumption 2 that is sufficient for policy optimization.

Assumption 4. *Sign Preservation: The sign of conditional average treatment effects is the same for the surrogate index and the long-term outcome.*

$$sign(\tilde{\tau}(x)) = sign(\tau(x)) \; \forall a \in \mathbb{A}, x \in \mathbb{X} \tag{12}$$

This is an assumption directly on CATEs, and so is not as readily interpretable with respect to the data-generating process. Nonetheless, we can reason about how this assumption may be more plausible in some settings than others. For example, if we hypothesize that a treatment "works" (i.e., has a large positive effect) on some groups by not others, and this treatment has some cost, then the distribution of CATEs may be bi-modal with no mass near zero. Furthermore, one can characterize this loss in policy optimization, much as Athey et al. (2019) develop bounds on the bias for the ATE. Here we state this result, with details in Appendix C.

[13] Concern with getting the sign of the treatment effect correct using surrogates has featured prominently in the literature on the "surrogate paradox", in which various surrogacy definitions are satisfied by the effect on the surrogate and outcome have opposite signs; see, e.g., Chen et al. (2007); VanderWeele (2013); Jiang et al. (2016).

Proposition 3. *There is a loss in the value of optimal policy only when a CATE estimated on surrogate index has a different sign than the true CATE. The total loss equals to the sum of the absolute value of true CATE weighted by the fraction of subscribers with the corresponding covariates that the CATE is conditioned on.*

In summary, assumptions introduced in the surrogacy literature can be used to justify policy evaluation and optimization with a surrogate index. Furthermore, it is possible to relax these assumptions for policy optimization precisely because the optimal policy is only sensitive to the sign of treatment effects.

5 Learning Optimal Policy from the Experiments

We first assigned subscribers to treatment using a behavior or design policy π_D[14] in the first cohort that balances exploration and exploitation, we do so by assigning subscribers with higher predicted churn probability into treatment with higher probability (see Appendix D.2 for a more detailed discussion). We then optimized the targeting policy using results from the first cohort and implemented it in the second cohort.

5.1 Off-policy Evaluation

Off-policy evaluation means we want to use data collected under the behavior policy π_D to estimate the value of a counterfactual policy π_P. One popular choice of estimator is based on inverse probability weighting (IPW). The Hajek estimator, a normalized version of the Horvitz–Thompson estimator (Horvitz and Thompson, 1952), is typically used to implement IPW. The average long-term outcome under an arbitrary targeting policy π_P using data collected under a design or behavior policy π_D is:

$$\hat{V}_{\text{IPW}}(\pi_P) = \left(\sum_i \frac{\pi_P(a_i|x_i)}{\pi_D(a_i|x_i)} \right)^{-1} \cdot \sum_i \frac{\pi_P(a_i|x_i)}{\pi_D(a_i|x_i)} Y_i \tag{13}$$

where Y_i is the outcome, $a_i \in \{0, 1, 2, ..., K-1\}$ is the actual treatment received by subscriber i in the first cohort assigned by the design policy π_D. π_P is the probability of assigning subscriber i to a given condition under the counterfactual policy that we want to evaluate.[15] We will use $a_i = 0$ to denote the control and $a_i = 1$ to denote the treatment when actions are binary.[16] The first term in Equation 13 is simply a normalization term; the ratio between π_P and π_D is also known

[14] In reinforcement learning literature (e.g., Sutton and Barto, 2018) the policy used to collect training data is called a behavior policy. We also call it a design policy in our experimental setting.

[15] The corresponding unnormalized Horvitz–Thompson estimator is: $\frac{1}{n} \sum_i \frac{\pi_P(a_i|x_i)}{\pi_D(a_i|x_i)} \cdot Y_i$

[16] For example, when $a_i = 1$ it means subscriber i was in treatment and she was assigned to treatment with probability $\pi_D(1|x_i)$, and $\pi_P(1|x_i)$ is the probability that i receives treatment under counterfactual policy π_P. Similarly, when $a_i = 0$ it means subscriber i was in the control and she was assigned to control with probability $\pi_D(0|x_i)$, and $\pi_P(0|x_i)$ is the probability that i will be in control (or *not* be treated) under counterfactual policy π_P.

as the importance weight. We need π_D to be strictly positive for all subscriber action pairs. Note that we don't require the policy being evaluated π_P to have this property, it can be a deterministic policy. In general, the Horvitz–Thompson estimator is unbiased but has higher variance. The Hajek estimator is biased in finite samples but consistent, and it has lower variance is therefore more widely used in practice.[17] The main advantage of IPW is that it's fully non-parametric when the propensity scores are known and it doesn't require us to specify a model for the outcome process.

However, the IPW estimator has two main limitations: first, Hajek can still suffer from high variance. Second, when evaluating a deterministic policy π_P, it only uses observations for which the actions prescribed by the target policy π_P and design policy π_D agree (when they don't agree $\pi_P(a_i|x_i)$ is always zero). This reduces the effective sample size, especially when π_P and π_D are very different.[18] Following Robins et al. (1994), one way to improve upon IPW is by augmenting it with an outcome model μ to use all observations and further stabilize the estimator. This is known as the doubly-robust method (DR) (Dudík et al., 2014). Under the DR approach, the value of a policy π_P can be estimated as:

$$\hat{V}_{\text{DR}}(\pi_P) = \frac{1}{n}\sum_i \left(\hat{\mu}(x_i, \pi_P) + \frac{\pi_P(a_i|x_i)}{\pi_D(a_i|x_i)} \cdot (Y_i - \hat{\mu}(x_i, a_i)) \right) \tag{14}$$

where

$$\hat{\mu}(x_i, \pi_P) = \sum_{a \in A} \pi_P(a|x_i)\hat{\mu}(x_i, a) \tag{15}$$

The first term in Equation 14 $\hat{\mu}(x, a)$ is an outcome model that estimates the expectation of the outcome for a given action a and covariates profile x. The second term is the importance weight multiplied by the prediction error, it corrects the first term towards the direction of the long-term outcome by an amount that is proportional to the prediction error. For a deterministic target policy π_P it does so whenever the actions prescribed by π_D and π_P agree. Note that the high variance of IPW is from the importance weights (dividing by a small probability when π_D is very unbalanced), this term vanishes if the prediction error is small. Both IPW and DR are consistent, but DR is known to have lower variance and therefore more efficient. We use the DR estimator for policy evaluation.

[17] For more discussion about the difference please see Owen (2019).

[18] Two policies are similar if they tend to prescribe the same action for a given subscriber profile, the more often they prescribe different actions for a given subscriber, the more different they are.

5.2 Off-policy Optimization

As shown in the previous section, policy optimization builds on CATE estimation. We focus on using doubly-robust estimation.[19] We can first construct a doubly-robust score for each subscriber–action pair (which also has the interpretation of an estimate of an individual potential outcome) (Robins et al., 1994; Chernozhukov et al., 2016; Dudík et al., 2014; Athey and Wager, 2020; Zhou et al., 2018):

$$\hat{\gamma}_a(x_i) = \hat{\mu}(x_i, a) + \frac{Y_i - \hat{\mu}(x_i, a)}{\pi_D(a|x_i)} \cdot 1_{\{a_i = a\}} \tag{16}$$

These doubly-robust scores are equal to the prediction of an outcome model plus a correction term based on IPW; the correction is applied if and only if the action being evaluated is the same as the action taken. This is intuitive because the correction term depends on Y_i which is the outcome under a realized action A_i, it is informative only when the action being evaluated is the same as a, otherwise the term drops out and the doubly-robust scores reduce to the outcome model. CATEs can then be estimated as:

$$\hat{\tau}_a(x_i) = \frac{1}{n} \sum_i (\hat{\gamma}_a(x_i) - \hat{\gamma}_0(x_i)) \tag{17}$$

We can use these doubly-robust scores for policy optimization (Murphy et al., 2001; Dudík et al., 2014) by solving a cost-sensitive classification problem. This has been shown to have good efficiency properties (Athey and Wager, 2020; Zhou et al., 2018).[20] That is, we estimate the optimal policy with:

$$\hat{\pi}^*(x_i) = \text{argmax}_{\pi \in \Pi} \ \frac{1}{n} \sum_i (\hat{\gamma}_1(x_i) - \hat{\gamma}_0(x_i)) \cdot (2\pi(x_i) - 1) \tag{18}$$

or in multi-action case:

$$\hat{pi}^*(x_i) = \text{argmax}_{\pi \in \Pi} \ \frac{1}{n} \sum_i < \hat{\gamma}(x_i), \pi(x_i) > \tag{19}$$

where $\hat{\gamma}(x) = (\hat{\gamma}_0(x), \hat{\gamma}_1(x), ..., \hat{\gamma}_k(x))$ is a vector of doubly-robust scores based on Equation 16 and $\pi(x)$ is a vector of probabilities with which the policy assigns a unit to each action. $< \cdot >$ is the dot product between vector valued $\hat{\gamma}(x)$ and $\pi(x)$.

In the cost-sensitive classification problem, for each unit, the correct label is the action that corresponds to the highest doubly-robust score, and the loss for classifying a unit to action a_i, when the correct label is a_i^*, is $\hat{\gamma}_{a^*}(x_i) - \hat{\gamma}_a(x_i)$. In the multi-action case, a cost-sensitive binary classification is done on every pair of actions, and the final action is chosen by a majority vote.

[19] Estimation of CATE can also be implemented in different ways. Hitsch and Misra (2018) distinguish between what they label "indirect" approaches (which first estimate the outcome model as a function of covariates and actions and then take the difference between actions as treatment effects) and "direct" methods estimate the CATE directly without first estimating an outcome function (e.g., causal trees (Athey and Imbens, 2016)), causal forest (Wager and Athey, 2018) and causal kNN (Hitsch and Misra, 2018)). This typology may be confusing to readers familiar with contextual bandit and policy learning literatures where, at least since Dudík et al. (2014), "direct methods" are those using outcome regressions without IPW (i.e. what Hitsch and Misra (2018) label "indirect").

[20] Here efficiency means that the difference between the value of a true and estimated optimal policy, also known as regret, decays faster as sample size increases.

When policy is restricted, we can choose a specific type of classifier (e.g., logistic regression or decision trees for interpretation or transparency reasons) or not allow the classifier to use certain types of information (note that we still use all information to construct the doubly-robust scores and can exclude a subset of features at the classification stage). Another advantage of this approach is that once the doubly-robust scores or labels are constructed, we can plug them into off the shelf classifiers to learn the optimal policy.

To account for the statistical uncertainty in action selection and continue exploration we use a variant of Thompson sampling, bootstrap Thompson sampling (Eckles and Kaptein, 2014; Lu and Van Roy, 2017; Osband et al., 2016), that is readily implemented for models for which Thompson sampling might be cumbersome to implement the optimal policy; see Eckles and Kaptein (2019) and Osband et al. (2017) for reviews. We use bootstrap Thompson sampling as a heuristic approach to adding randomized uncertainty-based exploration to the estimated optimal targeting policy in the second cohort where a subscriber i is assigned to action a with probability proportional to the fraction of times an action is estimated to be optimal across all bootstrap replicates.[21]

6 Surrogate Index Validation and Comparison

To evaluate the efficacy of our approach empirically, we first look at how well the surrogate index recovers the true long-term outcome and the treatment effect estimated on the true long-term outcome. We then validate it by looking at how it performs against a benchmark policy that's learned on some short-term proxies of the long-term outcomes (e.g., 1-6 month revenue), and a policy learned on the true long-term outcome (e.g., realized 18-month revenue, from August 2018 to February 2020). We also look at how the performance changes if we chose a different subset of surrogates. The surrogates we use are: content consumption (number of articles read in each of the 20 most visited sections[22] on *The Boston Globe*'s website) and revenue over the first 6 months. Intuitively, the longer we wait, the better we can estimate the long-term revenue. But we also want to learn the optimal policy fast so we can implement it; 6 month seems to strike a good balance. All policy values here are defined relative to the status quo of treating no one. And all confidence intervals are 95% from 1,000 bootstrap draws in the testing data.

First, we look at how the average treatment effect on the treated (ATT) calculated using the surrogate index compares with ATT calculated using the true outcome. The results are shown in Figures 1a. The surrogate index based ATT estimates match the true estimate quite well after just the first month. Note that the confidence intervals of ATT estimated on true outcomes are wider than the ones estimated on surrogate index. When the surrogacy assumption holds, it is more efficient to estimate the treatment effect on surrogate index because it discards irrelevant variation in the long-term outcome. For policy learning purpose, it's more important to learn the

[21] In cases where a subscriber is always or never assigned to some conditions we need to impose a probability floor and ceiling to ensure that all subscribers have positive probability being assigned to all conditions.

[22] The sections are: metro, sports, news, lifestyle, business, opinion, arts, Sunday magazine, ideas, search, member center, south, spotlight, page not found, nation, north, magazine, circulars, politics.

treatment effects on the long-term outcome than predicting the levels of the long-term outcome. Next, we look at the value of surrogate index-based policy (Figure 1c), all results are significantly better than the status quo except when we only use information from the first month. By contrast, optimizing the policy directly on short-term proxies (1-6 month revenue) doesn't outperform the status quo (Figure 1d).

We also compare the surrogate index-based policy with policy based on the true long-term outcome (Figure 1b). Although all the point estimates of the value difference are negative, none of them is distinguishable from zero, suggesting that surrogate-based policies do not perform significantly worse than the policy based on the true outcome. Lastly, we compare surrogate outcomes constructed using only content consumption information, only short-term revenue and both. As shown in Figure 2, the three approaches are not significantly different.[23]

7 Experimental Results

7.1 First Cohort

We plot the empirical survival curves in Figure 3 using data from August 2018 – February 2020. The first thing to notice is that the survival rate is relatively high, about 80% of subscribers at the beginning of the experiment remain subscribers 1.5 years later. Second, there is a gap between treatment and control group. We summarize the treatment effect over time in Appendix D.3.

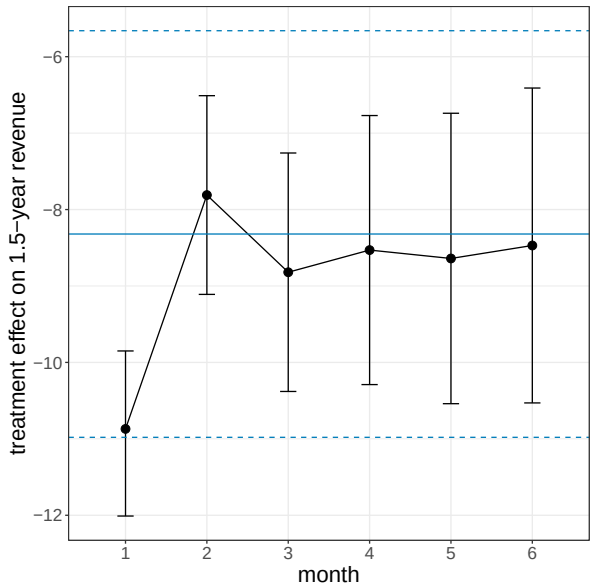
We then estimate the optimal policy via the cost-sensitive classification discussed in the previous section (Equation 18) on both observed mid-term (18-month) revenue and imputed long-term (3-year) revenue.[24] We first construct doubly-robust scores for each subscriber using Equation 16 where $\hat{\mu}$ is estimated using XGBoost via cross-fitting.[25] We then split the data into training (80%) and testing sets (20%) and use XGBoost as the classifier with hyper-parameters tuned via cross-validation.

In Appendix D.4 we compare value of the optimal policy learned on the realized 18-month revenue against (1) the value of benchmark policies including treating subscribers at random and treating subscribers with highest risk of churn; (2) the value of optimal policies learned on different subsets of features to highlight the value of information; (3) the value of optimal policies learned via different classification models, outcome regression (indirect method) and causal forest (direct method) to highlight the value of the model. In Appendix D.5 we use tools in interpretable machine learning to look at what variables are most important in determining the optimal policy, and how the optimal policy depends on these variables.

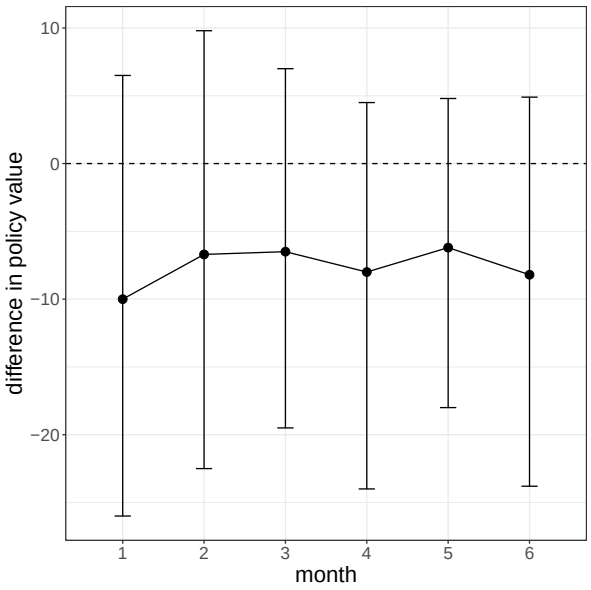
[23] Athey et al. (2019) suggests that when the surrogacy condition holds, the smallest set of surrogates has the highest precision in estimating the treatment effect.

[24] In a subscription model revenue and churn are equivalent since revenue for each subscriber = $\sum_t p_t r$ where p_t is the survival probability in period t and r is the fixed per period subscription fee.

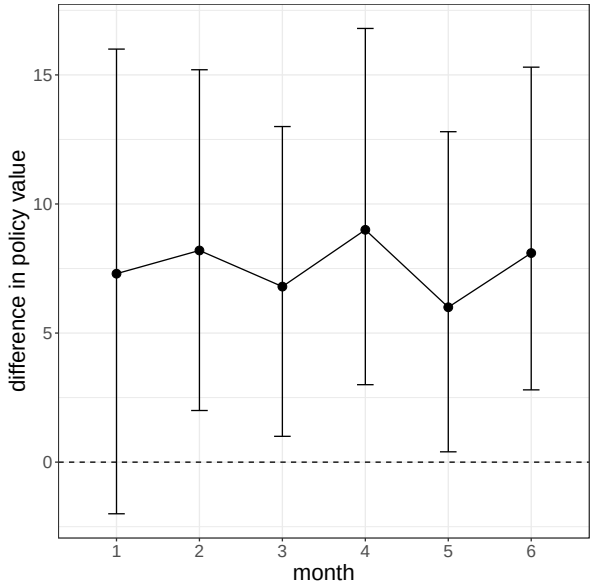
[25] Cross-fitting means that $\hat{\mu}$ for individual i is estimated *without* using i's own data in the training process. We can split data randomly into n folds, then $\hat{\mu}$ for individuals in a given fold is trained only using data from the other $n-1$ folds, it reduces over-fitting and improves efficiency (Athey and Wager, 2020; Zhou et al., 2018). We use $n = 3$ in our estimation.



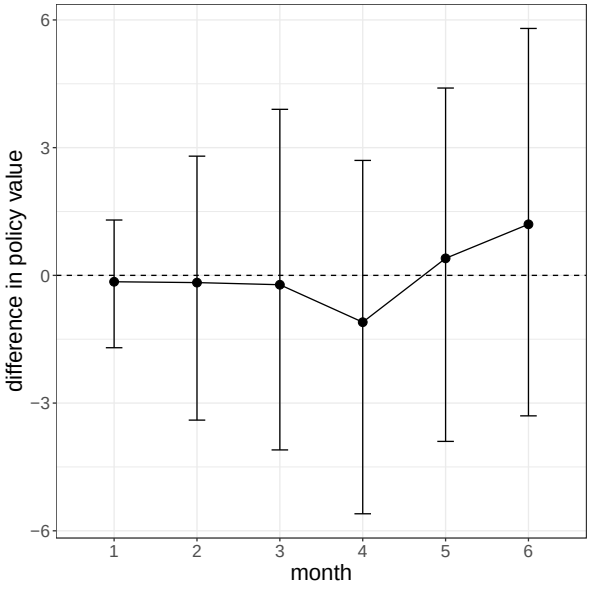
(a) Average treatment effect on the treated (ATT) on revenue using a surrogate index estimated using data from the first 1-6 month, the blue line is the ATT estimated using true 18-month revenue.



(b) The value difference between optimal policies learned on surrogate indices and true outcomes. Surrogate-index-based policies are statistically indistinguishable from the policy learned on the true outcome.



(c) The value difference between optimal policies learned on surrogate indices constructed with surrogates from 1-6 months and the current policy. Except for a single month, they outperform the status quo.



(d) The value difference between optimal policies learned with a single short-term proxy (revenue at month 1–6) and the current policy. The value is indistinguishable from the status quo.

Figure 1. The empirical validation of using surrogate index for policy learning.

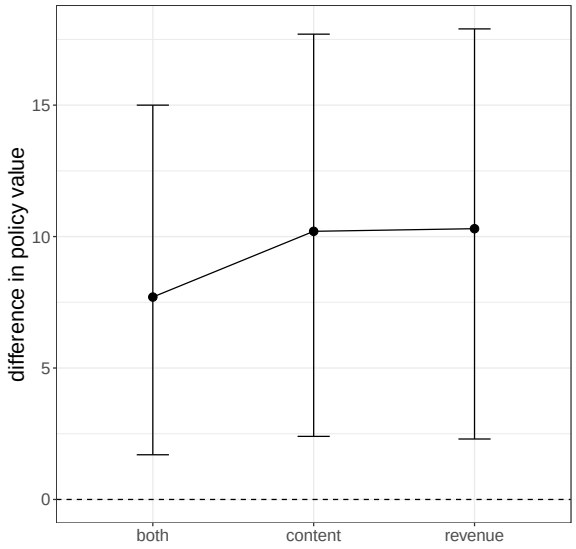


Figure 2. The value difference between policies learned using surrogate indices using content consumption variables, short-term revenue variables, or both, and the current policy. Each improves over the status quo.

The optimal policy can generate $40 per subscriber revenue increase (95% confidence interval [$10, $75]) over 3 years compared to the current policy that treats no one, which is $1.7 million dollars in total for the first cohort. We implement the optimal policy based on imputed 3-year revenue in the second cohort via bootstrap Thompson sampling. The density of treatment probability, as outputted by the bootstrap Thompson sampling using 18-month and imputed 3-year revenue as outcome, is summarized in Figure 4. We can see that both policies have the highest density near 0, but the 3-year policy assigns more subscribers to treatment than the 18-month policy. We re-scale the probabilities to make sure the total number of treated subscribers is approximately 6,000 for capacity reasons. Since we have 6 treatment conditions in the second cohort, we first used the bootstrap distribution of the optimal policy to decide who to treat, then conditional on treatment, we assigned subscribers to the 6 treatment conditions uniformly at random. We did this because all interventions except one are new. In future cohorts we can learn and implement an optimal policy over all interventions based on the results from the second cohort.

7.2 Second Cohort

We plot the survival curves in Figure 5 using data from July 2019 to February 2020. Treatment effects are reported in Appendix D.3. Surprisingly, $5.99/4 weeks and $5.99/8 weeks, which give the smallest discounts, have the biggest treatment effect on churn reduction. This, in turn, translates into the biggest effect on revenue.

We first provide some validation of estimated treatment effects by regressing churn and revenue on the interaction between treatment and treatment probability estimated. There is a significantly higher effect on subscribers that are predicted to have a bigger effect in the first cohort

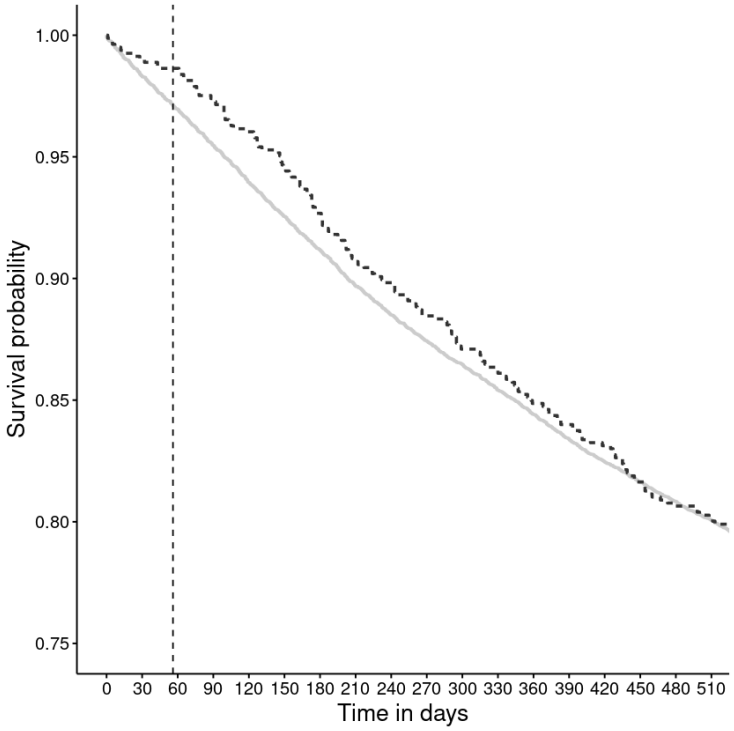


Figure 3. Empirical survival curve from the first cohort

(Table 1).[26] Then we optimize the policy, via multi-class cost-sensitive classification, using data from the second cohort. An optimal policy is summarized in Table 2 as fractions of subscribers being treated by each action. The optimal policy improves 3-year revenue by $30 per subscriber (95% confidence interval [$12, $50]) relative to the status quo that treats no one, generating $2.8 million in the second cohort.

We further compare the two cohorts to see whether there are significant changes in the environment in terms of covariate and concept shift in Appendix D.6. When the environment is stationary, it's more efficient to pool data from the two cohorts together to estimate the optimal policy for the next cohort, and when the environment is changing, it's better to down-weight observations from the first cohort using a time-decaying case weight (e.g., Russac et al., 2019). We only use data from the second cohort to estimate the optimal policy because there's some evidence for concept shift in our data and there's only one common treatment condition between the two cohorts.

8 Conclusion

Many applied problems, such as the pro-active churn management problem studied here, can be fruitfully characterized as learning a targeting policy. However, we often want to learn a policy to maximize long-term outcomes. Here we advance the practice of policy learning by

[26] We reported ATT in the table using inverse probability weights in the regression.

Table 1. Interaction between treatment and treatment probability

	Dependent variable:	
	churn	revenue
3.99/8 weeks	-0.016^{***} (0.003)	-22.032^{***} (0.279)
4.99/8 weeks	-0.005 (0.003)	-14.055^{***} (0.280)
5.99/4 weeks	-0.022^{***} (0.003)	-1.996^{***} (0.279)
5.99/8 weeks	-0.025^{***} (0.003)	-4.994^{***} (0.279)
gift card	-0.020^{***} (0.003)	-18.214^{***} (0.280)
thank you email only	-0.012^{***} (0.003)	0.905^{***} (0.278)
treatment prob	-0.005^{***} (0.001)	0.280^{***} (0.102)
3.99/8 weeks $\times$ treatment prob	-0.0002 (0.002)	-0.083 (0.144)
4.99/8 weeks $\times$ treatment prob	-0.003^{*} (0.002)	0.116 (0.146)
5.99/4 weeks $\times$ treatment prob	-0.003 (0.002)	0.530^{***} (0.145)
5.99/8 weeks $\times$ treatment prob	-0.006^{***} (0.002)	0.504^{***} (0.145)
gift card $\times$ treatment prob	-0.006^{***} (0.002)	0.152 (0.146)
thank you email only $\times$ treatment prob	0.003^{*} (0.002)	-0.332^{**} (0.145)
constant	0.105^{***} (0.002)	120.849^{***} (0.197)
Observations	95,554	95,554
Note:		*p<0.1; **p<0.05; ***p<0.01

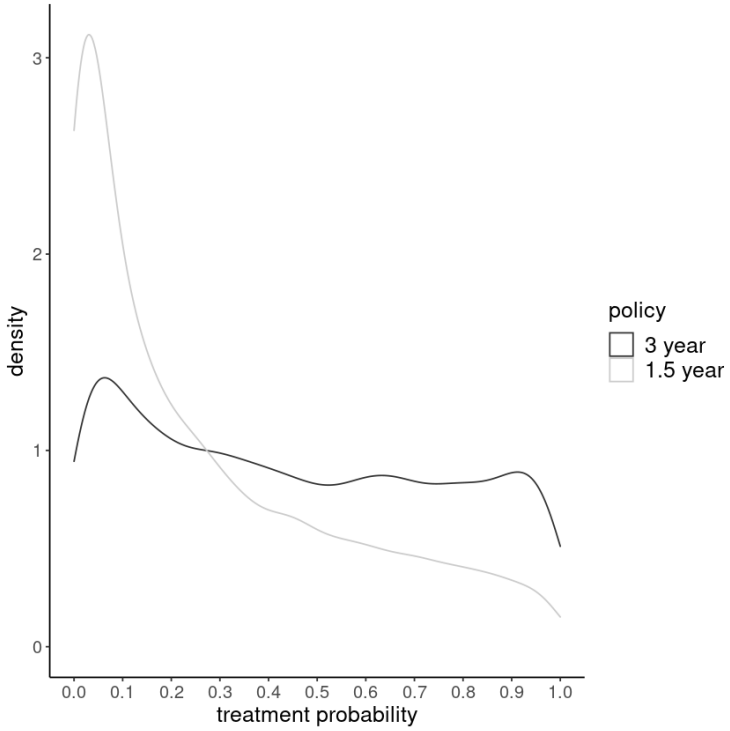


Figure 4. Distribution of treatment probabilities according to the design policy for the second cohort. These probabilities reflect uncertainty about whether a given subscriber should be treated, as computed using the bootstrap distribution. Distributions are shown for both 18-month revenue and imputed 3-year revenue.

incorporating the use of a learned surrogate index to impute the long-term outcomes. We first show analytically when a surrogate index is valid for policy evaluation and optimization in place of true unobserved long-term outcomes. Then to validate our approach empirically, we run two large-scale adaptive experiments that prescribe who should be targeted with what incentives in order to maximize long-term subscription revenue for *The Boston Globe*. We show that the policy optimized on long-term outcomes imputed by a surrogate index outperforms a policy optimized on a short-term proxy of the long-term outcomes. The surrogate index also performs similarly to the policy optimized on true long-term outcomes. We then implement the optimized policy with additional randomized exploration so that we can respond to potential non-stationarity and update the optimized policy after treating each cohort. The total 3-year revenue impact, relative to the status quo, of treating the first two cohorts with the policy optimized using the surrogate index sums to $4-5 million. Our paper adds to and complements a recent and growing literature in marketing on policy evaluation and learning (e.g., Hitsch and Misra, 2018; Simester et al., 2019a,b; Yoganarasimhan et al., 2020) and empirical work in proactive churn management (e.g., Ascarza, 2018) by focusing on optimizing targeting policies for long-term retention and revenue.

This framework can also be applied very generally to other empirical settings in business, education, or public policy where there is a need to personalize interventions to optimize some long-term outcomes and the cost of experimentation is relatively low. A natural question is how

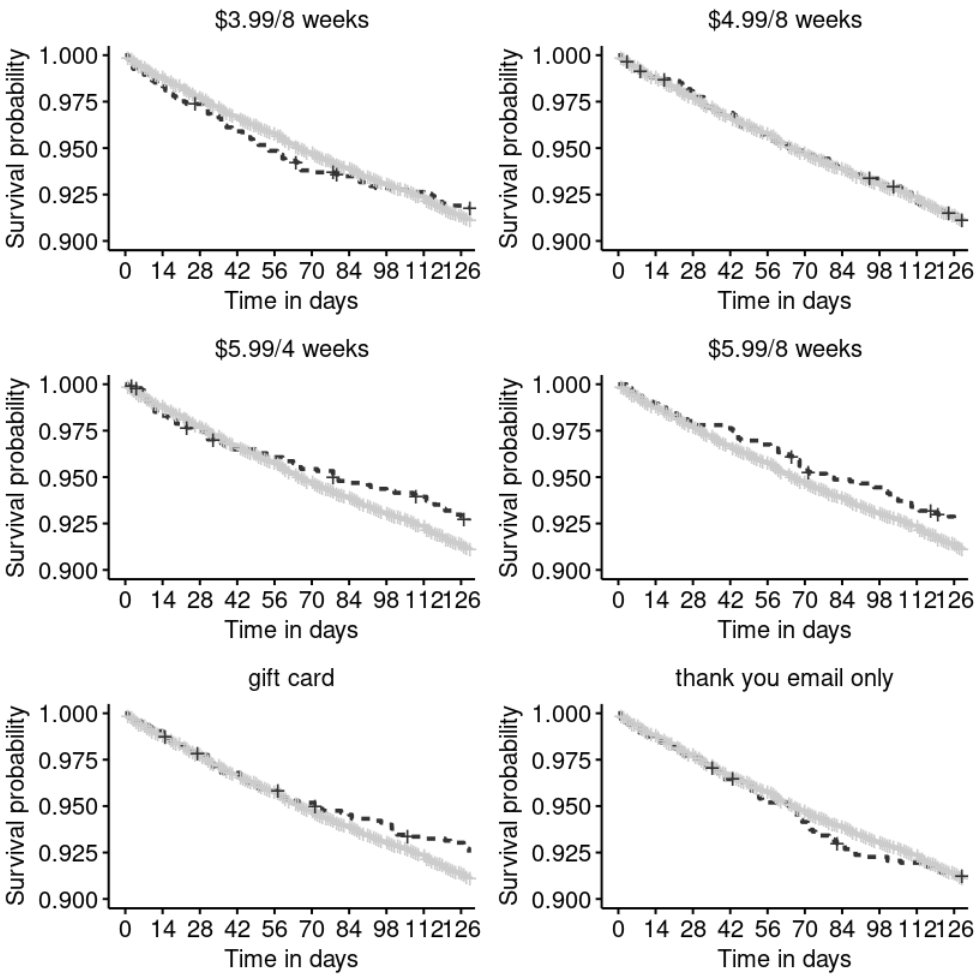


Figure 5. Empirical survival curve in the second cohort by treatment conditions. Dashed curve is the treatment group, vertical dashed line indicates when the discount ends (there's no ending date for gift card and thank you email only condition)

to choose surrogates when imputing long-term outcomes. In principle, we want to choose variables that lie on the causal chain from treatment to long-term outcomes, as suggested by domain knowledge or theory. We also want to choose surrogates that are observable shortly after the intervention so the policy can be learned quickly. If relevant experiments have been conducted in the past then the quality of surrogates can be evaluated on the realized long-term outcomes as we've shown in the paper. Surrogates that are highly predictive of the outcome are potential candidates but there's no guarantee that they will produce high policy values, as predicting the outcome level is a different task than predicting the treatment effect. Future research may examine selection of potential surrogates. Finally, since surrogacy is fundamentally a question about the underlying causal mechanism, once some surrogates have been shown to be valid for a given problem, they may be likely to remain valid for similar problems in the future. For example, we showed short-term revenues and content consumption are good surrogates for the effect of price discounts on long-term retention and subscription revenues, so the firm can tenatively rely on this assumption

Table 2. Distribution of optimal actions estimated from the second cohort

condition	percentage
control	23%
thank you email only	25%
gift card	< 1%
\$5.99/8 weeks	25%
\$5.99/4 weeks	27%
\$4.99/8 weeks	<1%
\$3.99/8 weeks	<1%

as they continue to iterate on targeting policies. We can imagine building such a knowledge base for different sets of problems and long-term outcomes as more empirical researchers work in this general framework.

References

Daniel W Apley and Jingyu Zhu. Visualizing the effects of predictor variables in black box supervised learning models. *arXiv preprint arXiv:1612.08468*, 2016.

Eva Ascarza. Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1):80–98, 2018.

Eva Ascarza, Raghuram Iyengar, and Martin Schleicher. The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, 53(1):46–60, 2016.

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Susan Athey and Stefan Wager. Policy learning with observational data. *arXiv preprint arXiv:1702.02896*, 2020.

Susan Athey, Raj Chetty, Guido W Imbens, and Hyunseung Kang. The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Technical report, National Bureau of Economic Research, 2019.

Robert C Blattberg, Byung-Do Kim, and Scott A Neslin. Why database marketing? In *Database Marketing*, pages 13–46. Springer, 2008.

Hua Chen, Zhi Geng, and Jinzhu Jia. Criteria for surrogate end points. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):919–932, 2007.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

Tianqi Chen, Tong He, Michael Benesty, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, pages 1–4, 2015.

Victor Chernozhukov, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins. Locally robust semiparametric estimation. *arXiv preprint arXiv:1608.00033*, 2016.

Hannes Datta, Bram Foubert, and Harald J Van Heerde. The challenge of retaining customers acquired with free trials. *Journal of Marketing Research*, 52(2):217–234, 2015.

Miroslav Dudík, Dumitru Erhan, John Langford, and Lihong Li. Doubly robust policy evaluation and optimization. *Statistical Science*, 29(4):485–511, 2014.

Dean Eckles and Maurits Kaptein. Thompson sampling with the online bootstrap. *arXiv preprint arXiv:1410.4009*, 2014.

Dean Eckles and Maurits Kaptein. Bootstrap Thompson sampling and sequential decision problems in the behavioral sciences. *SAGE Open*, 9(2):2158244019851675, 2019.

Peter S Fader and Bruce GS Hardie. How to project customer retention. *Journal of Interactive Marketing*, 21(1):76–90, 2007.

Laurence S Freedman, Barry I Graubard, and Arthur Schatzkin. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, 11(2):167–178, 1992.

Miguel Godinho de Matos, Ferreira Pedro, and Belo Rodrigo. Target the ego or target the group: Evidence from a randomized experiment in proactive churn management. *Marketing Science*, 2018.

Sunil Gupta, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V Kumar, Nathaniel Lin, Nalini Ravishanker, and S Sriram. Modeling customer lifetime value. *Journal of service research*, 9(2): 139–155, 2006.

Günter J Hitsch and Sanjog Misra. Heterogeneous treatment effects and optimal targeting policy evaluation. Working paper, 2018.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Zhichao Jiang, Peng Ding, and Zhi Geng. Principal causal effect identification and surrogate end point evaluation by multiple trials. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(4):829–848, 2016.

Marshall M Joffe and Tom Greene. Related causal frameworks for surrogate outcomes. *Biometrics*, 65(2):530–538, 2009.

Steffen L Lauritzen. Discussion on causality. *Scandinavian Journal of Statistics*, 31(2):189–201, 2004.

Aurélie Lemmens and Christophe Croux. Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2):276–286, 2006.

Aurélie Lemmens and Sunil Gupta. Managing churn to maximize profits. *Marketing Science*, 2020. forthcoming.

Xiuyuan Lu and Benjamin Van Roy. Ensemble sampling. In *Advances in neural information processing systems*, pages 3258–3266, 2017.

Kanishka Misra, Eric M Schwartz, and Jacob Abernethy. Dynamic online pricing with incomplete information using multiarmed bandit experiments. *Marketing Science*, 38(2):226–252, 2019.

Susan A Murphy, Mark J van der Laan, James M Robins, and Conduct Problems Prevention Research Group. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423, 2001.

Scott A Neslin, Sunil Gupta, Wagner Kamakura, Junxiang Lu, and Charlotte H Mason. Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of marketing research*, 43(2):204–211, 2006.

Ian Osband, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. Deep exploration via bootstrapped DQN. In *Advances in neural information processing systems*, pages 4026–4034, 2016.

Ian Osband, Daniel Russo, Zheng Wen, and Benjamin Van Roy. Deep exploration via randomized value functions. *Journal of Machine Learning Research*, 2017.

Art B Owen. *Monte Carlo Theory, Methods and Examples*. 2019. Book manuscript. Available at https://statweb.stanford.edu/~owen/mc/.

Ross L Prentice. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*, 8(4):431–440, 1989.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427): 846–866, 1994.

Yoan Russac, Claire Vernade, and Olivier Cappé. Weighted linear bandits for non-stationary environments. In *Advances in Neural Information Processing Systems*, pages 12017–12026, 2019.

Duncan Simester, Artem Timoshenko, and Spyros Zoumpoulis. Efficiently evaluating targeting policies: Improving upon champion vs. challenger experiments. 2019a.

Duncan Simester, Artem Timoshenko, and Spyros I Zoumpoulis. Targeting prospective customers: Robustness of machine learning methods to typical data challenges. *Management Science*, 2019b.

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

Tyler J VanderWeele. Surrogate measures and consistent surrogates. *Biometrics*, 69(3):561–565, 2013.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Christopher J Weir and Rosalind J Walley. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Statistics in Medicine*, 25(2):183–203, 2006.

Jane Xu and Scott L Zeger. The evaluation of multiple surrogate endpoints. *Biometrics*, 57(1):81–87, 2001.

Hema Yoganarasimhan, Ebrahim Barzegary, and Abhishek Pani. Design and evaluation of personalized free trials. *arXiv preprint arXiv:2006.13420*, 2020.

Zhengyuan Zhou, Susan Athey, and Stefan Wager. Offline multi-action policy learning: Generalization and optimization. *arXiv preprint arXiv:1810.04778*, 2018.

Appendix

A New York Times Example

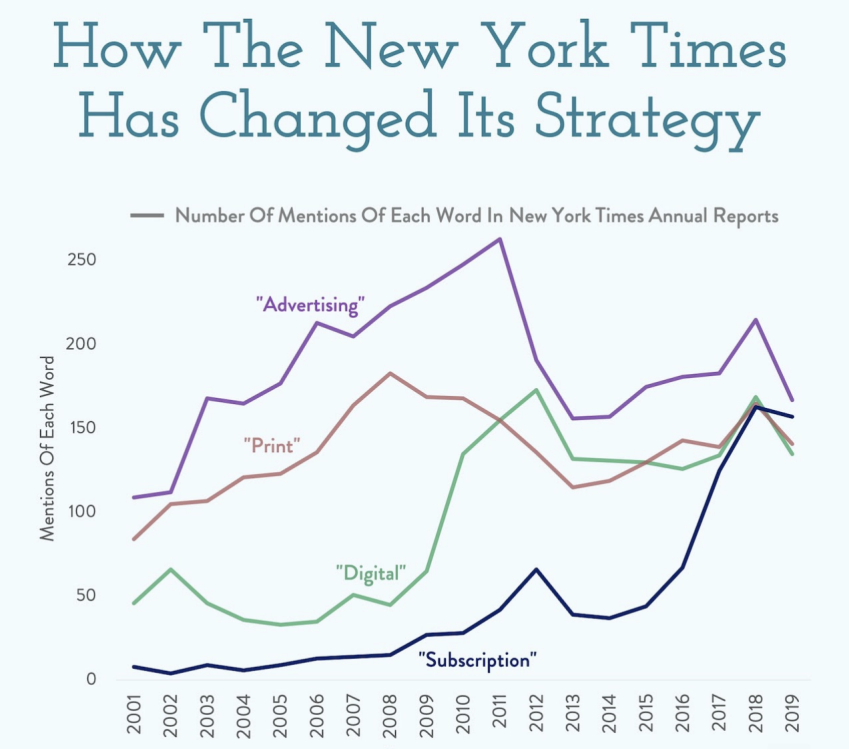


Figure A.1. Number of mentions of key works in annual report over time (Source: chartr)

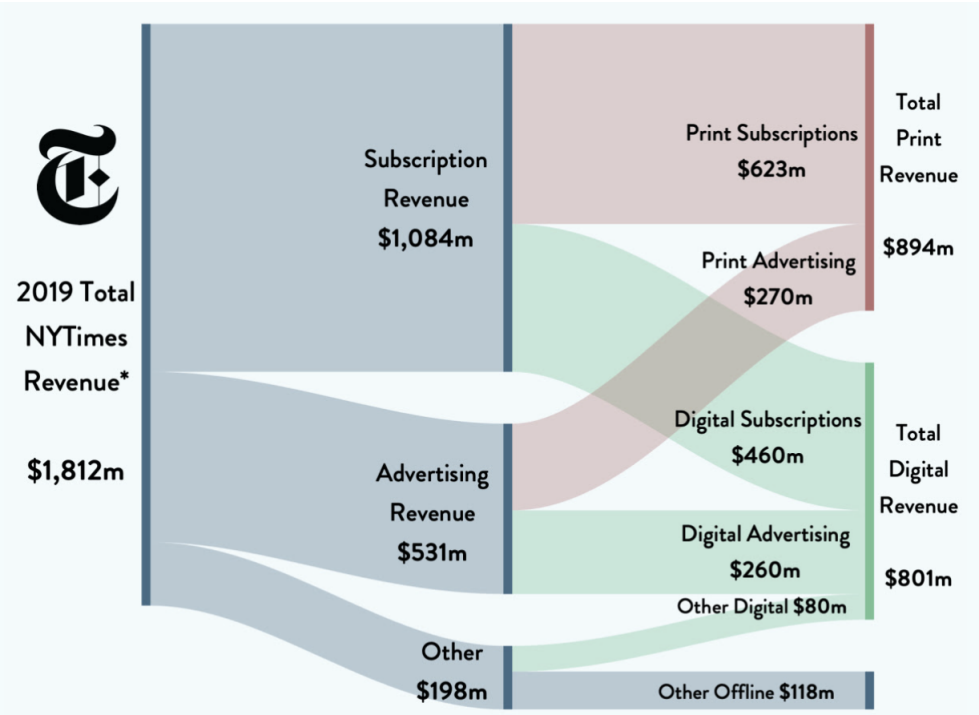
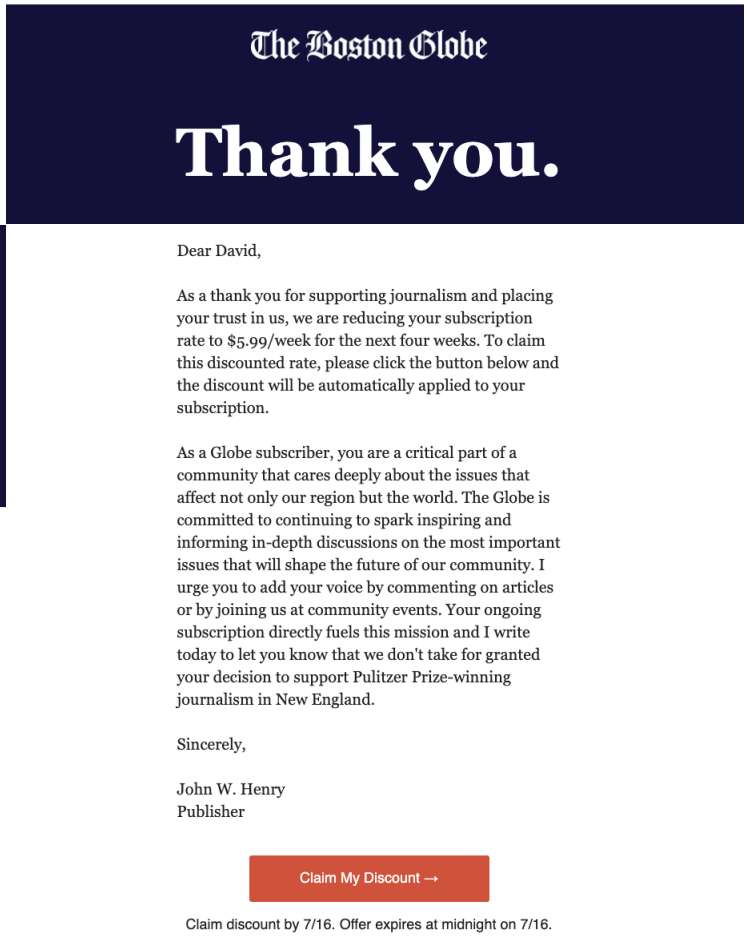


Figure A.2. Revenue breakdown in 2019 (Source: chartr)

B Targeting Emails

First cohort



Second cohort

Figure B.1. Targeting emails. First cohort (left): A sample email sent to targeted subscribers in August 2018, discounts are applied to subscribers automatically. Second cohort (right): A sample email sent to targeted subscribers in July 2019, subscribers have to redeem the offer by clicking on "claim my discount" before the expiration day which is 24 hours after the email was sent. $5.99/week for 4 weeks is one of the 6 treatment conditions

C Proof of Propositions

Proposition 1:

Proof. Consider a case with binary actions. We show that the value of a policy as defined on true long-term outcomes Y is identified using the surrogate index $\tilde{Y}$.

$$
\begin{aligned}
V(\pi) &= n^{-1} \sum_{i}^{n} \{\pi(x_i) Y_i(1) + (1 - \pi(x_i)) Y_i(0)\} \\
&= n^{-1} \sum_{i}^{n} \{\pi(x_i) \mathbb{E}[\frac{A_i Y_i}{e(x_i)}] + (1 - \pi(x_i)) \mathbb{E}[\frac{(1 - A_i) Y_i}{1 - e(x_i)}]\} \\
&= n^{-1} \sum_{i}^{n} \mathbb{E}[\pi(x_i) \frac{A_i Y_i}{e(x_i)} + (1 - \pi(x_i)) \frac{(1 - A_i) Y_i}{1 - e(x_i)}] \\
&= n^{-1} \sum_{i}^{n} \mathbb{E}[\mathbb{E}[\pi(x_i) \frac{A_i Y_i}{e(x_i)} + (1 - \pi(x_i)) \frac{(1 - A_i) Y_i}{1 - e(x_i)} | s_i, x_i]] \\
&= n^{-1} \sum_{i}^{n} \mathbb{E}[\pi(x_i) \frac{\mathbb{E}[A_i | s_i, x_i] \mathbb{E}[Y_i | s_i, x_i]}{e(x_i)} + (1 - \pi(x_i)) \frac{\mathbb{E}[1 - A_i | s_i, x_i] \mathbb{E}[Y_i | s_i, x_i]}{1 - e(x_i)}] \\
&= n^{-1} \sum_{i}^{n} \mathbb{E}[\pi(x_i) \frac{A_i \tilde{Y}_i}{e(x_i)} + (1 - \pi(x_i)) \frac{(1 - A_i) \tilde{Y}_i}{1 - e(x_i)}]
\end{aligned}
\tag{20}
$$

$e(x_i)$ is the propensity score. The first line is from the definition of the value of a policy. The second line is because under Assumption 1 (ignorability and positivity) we have

$$
\begin{aligned}
\mathbb{E}[\frac{A_i Y_i}{e(x_i)}] &= \mathbb{P}(A_i = 1 | x_i) \frac{Y_i(1)}{e(x_i)} = Y_i(1) \\
\mathbb{E}[\frac{(1 - A_i) Y_i}{1 - e(x_i)}] &= \mathbb{P}(A_i = 0 | x_i) \frac{Y_i(0)}{1 - e(x_i)} = Y_i(0)
\end{aligned}
\tag{21}
$$

The third line is because $\pi(x_i)$ is a constant. The fourth line is from the law of iterated expectation: We first condition on surrogates and covariates s_i and x_i. The fifth line is based on Assumption 2 (surrogacy) so the expectation of product can be factorized into the product of expectations. The last line is based on undoing the law of iterated expectations, the definition of surrogate index and Assumption 3 (comparability) as in 9. The same argument also goes through for multi-action cases. □

Proposition 2:

Proof. For policy optimization, consider the case of binary actions, we can see that an optimal policy π^* maximizes the average outcome by assigning a subscriber to treatment if and only if the conditional average treatment effect (CATE) for that subscriber is positive (net of the cost of

treatment if any):

$$\begin{aligned}
\text{argmax}_\pi V(\pi) &= \text{argmax}_\pi \, n^{-1} \sum_i Y(x_i, \pi(x_i)) \\
&= \text{argmax}_\pi \, n^{-1} \sum_i \{\pi(x_i) Y_i(1) + (1 - \pi(x_i)) Y_i(0)\} \\
&= \text{argmax}_\pi \, n^{-1} \sum_i \{\pi(x_i)(Y_i(1) - Y_i(0)) + Y_i(0)\} \\
&= \text{argmax}_\pi \, n^{-1} \sum_i \{\pi(x_i)\tau(x_i) + Y_i(0)\}
\end{aligned} \tag{22}$$

$$\tau(x_i) := \mathbb{E}[Y_i(1) - Y_i(0)|x_i] \tag{23}$$

$$\pi^*(x_i) = \begin{cases} 1 & \tau(x_i) \geq 0 \\ 0 & \tau(x_i) < 0 \end{cases} \tag{24}$$

where $\tau(x_i)$ is the conditional average treatment effect (CATE) for individual i. In multi-action cases the idea is similar, we can define $\tau_a(x_i)$ as the CATE of taking action $a \in A/\{0\}$ relative to the control:

$$\tau_a(x_i) := \mathbb{E}[Y_i(a) - Y_i(0)|x_i] \tag{25}$$

An unrestricted optimal policy simply assigns units to condition $\text{argmax}_{a \in A} \, \tau_a(x_i)$. Because the optimal policy depends only on the sign of CATE on the long-term outcome, the policy optimized on surrogate index is valid as long as CATE estimated on the surrogate index is of the same sign as the true CATE.

Following a similar derivation as in the proof of Proposition 1:

$$\begin{aligned}
\tau(x_i) &= \mathbb{E}[Y_i(1) - Y_i(0)|x_i] \\
&= \mathbb{E}[\frac{A_i Y_i}{e(x_i)} - \frac{(1 - A_i) Y_i}{1 - e(x_i)}|x_i] \\
&= \mathbb{E}[\mathbb{E}[\frac{A_i Y_i}{e(x_i)} - \frac{(1 - A_i) Y_i}{1 - e(x_i)}|s_i, x_i]] \\
&= \mathbb{E}[\frac{\mathbb{E}[A_i|s_i, x_i]\mathbb{E}[Y_i|s_i, x_i]}{e(x_i)} - \frac{\mathbb{E}[1 - A_i|s_i, x_i]\mathbb{E}[Y_i|s_i, x_i]}{1 - e(x_i)}|x_i] \\
&= \mathbb{E}[\frac{A_i \tilde{Y}_i}{e(x_i)} - \frac{(1 - A_i)\tilde{Y}_i}{1 - e(x_i)}|x_i]
\end{aligned} \tag{26}$$

Surrogate index can be used to construct an unbiased estimator of CATE, therefore, it can be used for policy learning. □

Proposition 3:

Proof. When Assumption 2 (surrogacy) is violated, the CATE estimated using surrogate index is biased. Athey et al. (2019) showed that the bias on ATE is bounded by $\bar{b}$:

$$|b| \leq (\frac{\text{var}(Y_i)}{\text{var}(A_i)} \cdot (1 - R^2_{Y|S}) \cdot (1 - R^2_{A|S}))^{\frac{1}{2}} := \bar{b} \tag{27}$$

where $R^2_{Y|S}$ and $R^2_{A|S}$ is the R^2 of the regression of long-term outcome on surrogates (in the historical dataset), and the regression of actions on surrogates (in the experimental dataset), respectively. Similarly, the bias on CATE is bounded by $\bar{b}_X$:

$$|b_X| \leq (\frac{\text{var}(Y_i|X_i)}{\text{var}(A_i|X_i)} \cdot (1 - R^2_{Y|S,X}) \cdot (1 - R^2_{A|S,X}))^{\frac{1}{2}} := \bar{b}_X \tag{28}$$

because an optimal policy assigns actions based on the sign of true CATE, as long as the CATE estimated on surrogate index has the same sign as the true CATE, there's no loss on the value of policy. When the signs are different, the loss is the true CATE, it follows that the total loss in policy value is bounded above by:

$$\int_X (\bar{b}_X - |\tilde{\tau}(X)|) \cdot \mathbb{1}_{\{\bar{b}_X - |\hat{\tau}(X)| > 0\}} \, dF(X) \tag{29}$$

where $\tilde{\tau}(X)$ is the CATE for the surrogate index. □

D Supplementary Analyses

D.1 Churn Prediction

The dataset we have includes demographic, transaction history and browsing history for all digital only subscribers from 2010/12/16 to now. We first pick a date and use all the information before that date to predict outcomes (whether a given subscriber churned or not) that happened within six months after that date. We picked 2018/01/30 because it gives us the most recent information before targeting the first cohort, although model performance is robust to other dates that we picked.

We select active subscribers defined by the company, it includes all subscribers who are currently active, in grace period, or in temporary stop.[27] Then we construct features from transaction history using frequency and recency by each transaction type, which are standard features in the churn prediction literature (Lemmens and Croux, 2006). Frequency is the number of times a certain transaction type occurred, and recency is the first and last time a certain transaction type occurred compared with the date we picked (in days). Then we count the number of articles read in the last week, month, 3 months and 6 month to measure the level of engagement. We also extracted how many articles a subscriber read in each section on the newspaper's website over time, although this content consumption information is not used for churn prediction, we use it for policy evaluation. We look at churn that happened between 2018/01/30 and 2018/07/18 to get the outcome labels, if a churn happened it's coded as 1, and 0 otherwise. We handle missing data in the following way: if a feature is a measure of recency, then missing means that a certain type of event has not happened yet, so we impute a large positive number 1000 (a positive number means it is in the future) and create a separate column indicating if that value is missing (1 or missing, 0 for not missing). If a feature is categorical, we create "missing" as a new category. Altogether we have 183 features. We also removed recent subscribers whose tenure is less than 60 days and who hasn't opened any emails sent by the company in the last 6 months. The reason is that recent subscribers are likely to be on an introductory rate which is already discounted, we don't want to give them more discounts.

Then we build a classification model to predict the churn risk for each subscriber by combining information from three different sources: demographics (e.g., zip code), transaction history (credit card status, credit card expire date and transaction type, including auto notice, auto renew, refund, billing change, complaint, expire, end of grace period, payment cancel, payment declined, start, stop, etc., and associated time stamp, and a source and reason code associated with each transaction), and browsing history on the newspaper's official website (number of articles read and associated time stamp, article section, article headline) from 2010/12/16 to 2018/07/18. We trained and compared a wide range of classification algorithms. Among the models we trained,

[27] Most common reason for this is traveling.

gradient boosted decision trees (XGBoost) [28] (Chen and Guestrin, 2016; Chen et al., 2015) has the best out-of-sample performance measured by AUC (area under the curve). See Table D.1 for a comparison. We have an overall out-of-sample accuracy of 97%[29], and precision of 94 %[30]. However, the recall is low at 23%[31] suggesting that we might have missed some important signals when constructing features or the information is simply unobserved.

As in many classification problems, we need to trade-off the cost of false positive and false negative. In our setting, a false positive is a non-churner classified as a churner, and a false negative is a churner classified as a non-churner. The cost of a false positive is the cost of the discount. Since the subscriber is not going to churn, the firm wasted $(\$6.93 - \$4.99) \times 8 = \$15.52$ per targeted subscriber. The cost of a false negative is harder to evaluate because it depends on how soon the churn happened and how long she would have stayed if she had been targeted with the discount. Assuming a churner churned in 2 month, the revenue collected without the discount is $\$6.93 \times 8 = \$55.4$, if the churner would stay for an extra month if she received the discount, then the revenue collected would be $\$4.99 \times 8 + \$6.93 \times 4 = \$67.6$, therefore the cost would be $\$12.2$. These numbers should be identifiable from the data.

Table D.1 shows the prediction performance of a menu of classification models, and we can see that XGBoost outperforms other models by a significant margin. Table D.2 is the confusion matrix of the performance of XGBoost on a testing sample of size of 8000. We can see that the precision is very high (we get 60 out of 64 right when we predict someone to be a churner), but the recall is low (we correctly predict 60 out of 265 real churners). Table D.3 is the top 20 features that are predictive of churn, we can see that credit card information is very important, the company also mentioned that a big number of subscribers (over 25%) churn is from an expired credit card (they do send out notification emails to tell the subscribers if their cards are about to expire). Auto renew and billing change information are also important, so is the level of engagement as measured by number of articles read last week, month, and 6 months.

Table D.1. Predictive model performance

Model	AUC
Logistic Regression	0.7557
Supporter Vector Machine	0.5824
Random Forest	0.5669
XGBoost	0.9384

[28] XGBoost tends to be the winning algorithm for many Kaggle competitions. Check this medium article for an overview of the algorithm and why it outperforms other boosting methods such as MART (multiple additive regression trees).

[29] This is not surprising given the class labels are highly imbalanced. There are about 4% churn rate in the data, the overall accuracy is most from correctly predicting non-churners.

[30] Precision is the proportion of actual churner among predicted churner. It means that when we predict a subscriber to be a churner, 94% of the time we are correct.

[31] Recall is the proportion of predicted churners among actual churners. It means that among all the actual churners we correctly identified 23% of them.

Table D.2. Confusion matrix for XGBoost on testing Data

predicted/actual	0	1
0	7731	205
1	4	60

Table D.3. Relative feature importance in XGBoost (top 20)

feature	relative importance
credit_card_statusa	100.000
credit_card_statusi	66.728
last_autorenew	39.728
cc_expire_dt	31.951
last_billingchg_reasonremovecc	23.667
first_billingchg_reasonremovecc	18.981
last_start_tenure	7.919
credit_card_typeu	6.016
original_tenure	5.786
last_billingchg	5.252
first_autorenew	4.331
last_expire	3.954
first_billingChg	3.588
last_6month	3.501
last_week	3.346
last_month	3.281
num_autorenew	2.621
num_billingChg	2.252
num_pymtdecline	1.731
first_pymtdecline	1.648

D.2 Behavior Policy and Simulation

We obtain the design policy, which is the targeting policy we implement in the first round experiment, by garbling the predicted risk score from the XGBoost with random noise generated from a normal distribution.[32] The key idea is that we are treating subscribers with higher risk of churn with higher probability, but allow everyone to be either treated or not with positive probability.

The reason that we base the behavior policy on predicted risk score is twofold. First, because churn risk is an outcome bounded between 0 and 1, it provides an upper bound on how big the beneficial treatment effect[33] can be without any further assumptions. For instance, if a subscriber has a predicted risk of 0.1, it means that the discount will *at most* lower her risk by 0.1, on the other hand, if a subscriber has a predicted risk of 0.9, the discount can lower her risk by *up to* 0.9, provided that the model is well calibrated. So it's reasonable to treat subscribers with higher risk with higher probability without any additional information. This approach can also be interpreted as treating subscribers based on an upper confidence bound (UCB) of the beneficial treatment effect with minimal assumptions. Second, if risk of churn is indeed positively correlated with treatment effect, this approach lowers regret compared with a uniform policy which is the most typical exploration policy, it also gives us more precision to learn the policy at a region that matters the most (the region where the treatment effects are the highest) because we are assigning more subscribers in this region to treatment. We conduct a simple simulation analysis to further illustrate this. The result shows that, under bounded outcome while both policies recover the true ATE well, the design policy that assigns subscribers to treatment with probability proportional to her churn risk has lower regret compared with a uniformly at random policy, and this is true under very general conditions.

The reason we added noise to predicted risk is also twofold. First, we want to explore more around the predicted risk score. Without the noise, the targeting policy would reduce to a version that is the common practice, which is to target based on predicted outcome level, which is the churn risk in this context (Blattberg et al., 2008). Some exploration allows us to learn the treatment effect at regions that our prior thinks the effect is low, that is, subscribers with medium to low predicted risk of churn. This allows us to learn an optimal policy even when our prior is wrong. Second, to use inverse propensity score for off-policy evaluation and learning, we need all subscribers to have positive probability of being in all conditions[34]. Even when this condition is satisfied in principle, the variance of the counterfactual policy evaluation is very large and unstable when some of the probabilities are very close to zero (Dudík et al., 2014). After adding noise, we essentially make the propensity scores more smooth, that is, the probability of receiving the treatment for the top churners gets lower, and the probability of receiving the treatment for

[32] It is the best performing model for churn prediction, see Appendix D.1 for more details.

[33] Beneficial means when treatment effect is in the direction that moves the outcome in a desirable direction.

[34] Note that this condition is usually not satisfied using the common practice. Suppose the targeting policy is to treat subscribers who are in the top 5% of churn risk, then 95% of subscribers have zero probability of receiving the treatment by design.

the bottom churners gets higher. It ensures that everyone has a propensity score that is bounded away from 0 and 1. See Figure D.1 for the CDF of treatment probability before and after adding the noise (it's just the raw predicted churn risk before adding noise).

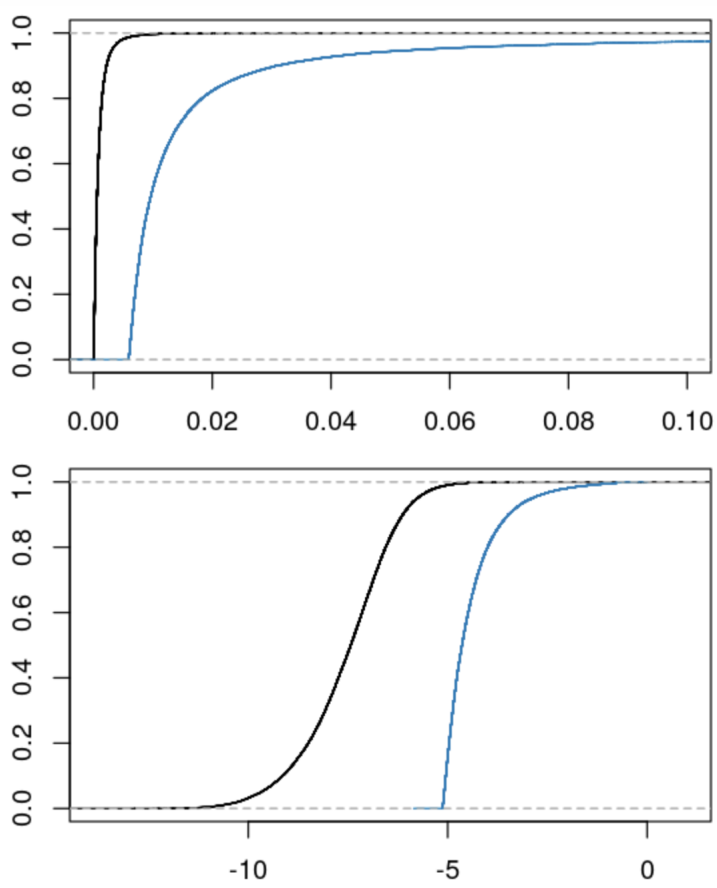


Figure D.1. CDF of risk scores (black) and treatment probability under the behavior policy (blue) on regular (top) and log scale (down): most subscribers have risk score close to zero, the behavior policy increases the treatment probability for those subscribers, but a big majority (over 80%) still has a treatment probability below 2%, the treatment probability is also capped at 50% to ensure sufficient exploration.

Let S be a vector of predicted risk scores, where $s_i \in (0, 1)$ is the predicted risk for subscriber i. A stochastic targeting policy P is define as:

$$P : X \to (0, 1)$$

it's a mapping from X, which is the covariate space of subscribers, to an open probability interval. Note that it's important for a *behavior or design policy* to be stochastic, meaning that every subscriber under the design policy has to have nonzero probability of both receiving the treatment *and* the control, so the interval should be open on both ends. In contrast, a deterministic targeting policy is $P : X \to \{0, 1\}$. The policy we want to evaluate can be both stochastic and deterministic, we only require the design policy to be stochastic. Because a policy is just the probabilities that subscribers

receive the treatment, we can think of it as a vector or propensity scores. Given the predicted risk score S, a design policy P_D is given by:

$$\begin{aligned}\mathbb{P}_D(T_i = 1) &= \mathbb{P}_D(s_i - \epsilon \geq \tau) \\ &= \mathbb{P}_D(\epsilon \leq s_i - \tau) \\ &= F_D(s_i - \tau)\end{aligned}$$

where $\epsilon \sim N(0, \sigma^2)$ is the random noise added, F is the CDF of ϵ. σ controls the amount of exploration and τ is a constant threshold that controls the total number of subscribers treated. A policy P_D is fully characterized by the choice of σ and τ. In the first round experiment the firm wants to send discount to about 1000 subscribers. In the design policy we implemented we have $\sigma = 0.003$, $\tau = 0.0068$. And we cap the probability of receiving the treatment at 50% for all subscribers.

To make the idea more concrete, we conducted a simple simulation to compare the performance of a uniformly at random policy and a design policy that assigns subscribers with higher churn risk to treatment with higher probabilities. We are particularly interested in (1) how good the outcomes are in the experiment and (2) how well the learning is.

Consider the following data generating process: let $0 < Y_i(0) < 1$ be the baseline churn risk for subscriber i without any interventions, this is essentially the output of our churn classification algorithm described in the previous section, and because of this we treat it as observables for all subscribers. Now suppose the intervention lowers churn risk, without any further assumptions we can draw $Y_i(1)$ uniformly from the interval $(0, Y_i(0))$, that is, we know the post treatment outcome $Y_i(1)$ has to be bounded above by $Y_i(0)$ and below by 0. Now we have the full schedule of potential outcomes, we can simulate two types of experiments: assigning subscribers to treatment with probability 0.01 (we call this the uniform policy), and assigning subscribers to treatment with probability proportional to churn risk (we call this the design policy) but keep the total fraction of treated subscriber fixed at 0.01. We compare (1) what's the average churn under uniform and design policy and (2) what's the estimated treatment effects under uniform and design policy (because we have the full schedule of potential outcome we can compare it with the ground truth).

Similarly, if the intervention increases churn risk, we draw $Y_i(1)$ uniformly from the interval $(Y_i(0), 1)$, that is, the post treatment outcome $Y_i(1)$ is bounded below by $Y_i(0)$ and above by 1. More generally, we can let the treatment effect for a given subscriber be negative with probability q and positive with probability $1-q$ and repeat the procedure, q captures the fraction of subscribers on whom the intervention has a negative treatment effect (we think in practice q should be quite large). We do 1000 repetitions according to policy with $q = 0, 0.5, 1$ and report the results in Figure D.2. We can see that the design policy has lower churn rate in all cases, and both design policy and uniform policy recover the true average treatment effect (ATE).

To further extend this analysis, we can allow the distributions of treatment effects to take different shapes similar to the simulation studies conducted in Misra et al. (2019). We also summarize

the simulation algorithm in pseudocode:

Data: y_0 $(n = 100,000)$ drawn from $uniform(0,1)$;
with probability $q \in [0,1]$ draws y_1 from $uniform(min = 0,\ max = y_0)$[a] and with probability $1-q$ draws y_1 from $uniform(min = y_0,\ max = 1)$[b];
initialization $i = 1$;
initialization $p = pnorm(y_0 - c), c = 2.9175$[c];
while $i \leq K = 1,000$ **do**
  draw $uniform\ policy = sample(c(0,1), n, replace = T)$;
  if $uniform\ policy == 1$ **then**
    $y_{uniform} = y_1$;
  else
    $y_{uniform} = y_0$;
  draw $design\ policy = rbern(n, p1)$;
  if $design\ policy == 1$ **then**
    $y_{design} = y_1$;
  else
    $y_{design} = y_0$;
  $\tau_{uniform} = lm(y_{uniform} \sim uniform\ policy)$;
  $\tau_{design} = lm(y_{design} \sim design\ policy, weights = 1/p\ or\ 1/(1-p))$[d];
  return$(mean(y_{uniform}), mean(y_{design}), \tau_{uniform}, \tau_{design})$

---

[a] This means y_1 is never greater than y_0 which corresponds to negative treatment effects.

[b] This means y_1 is never less than y_0 which corresponds to positive treatment effects.

[c] This number is chosen to make the number of treated subscribers under design policy close to 50,000 to make it comparable with uniform policy

[d] This is the inverse probability weight to account for the fact that under design policy subscribers are assigned to treatment and control with different probabilities, we weight each subscriber by $1/p_i$ if she is treated, and $1/(1-p_i)$ if she is in control, p is the treatment probability for subscriber i.

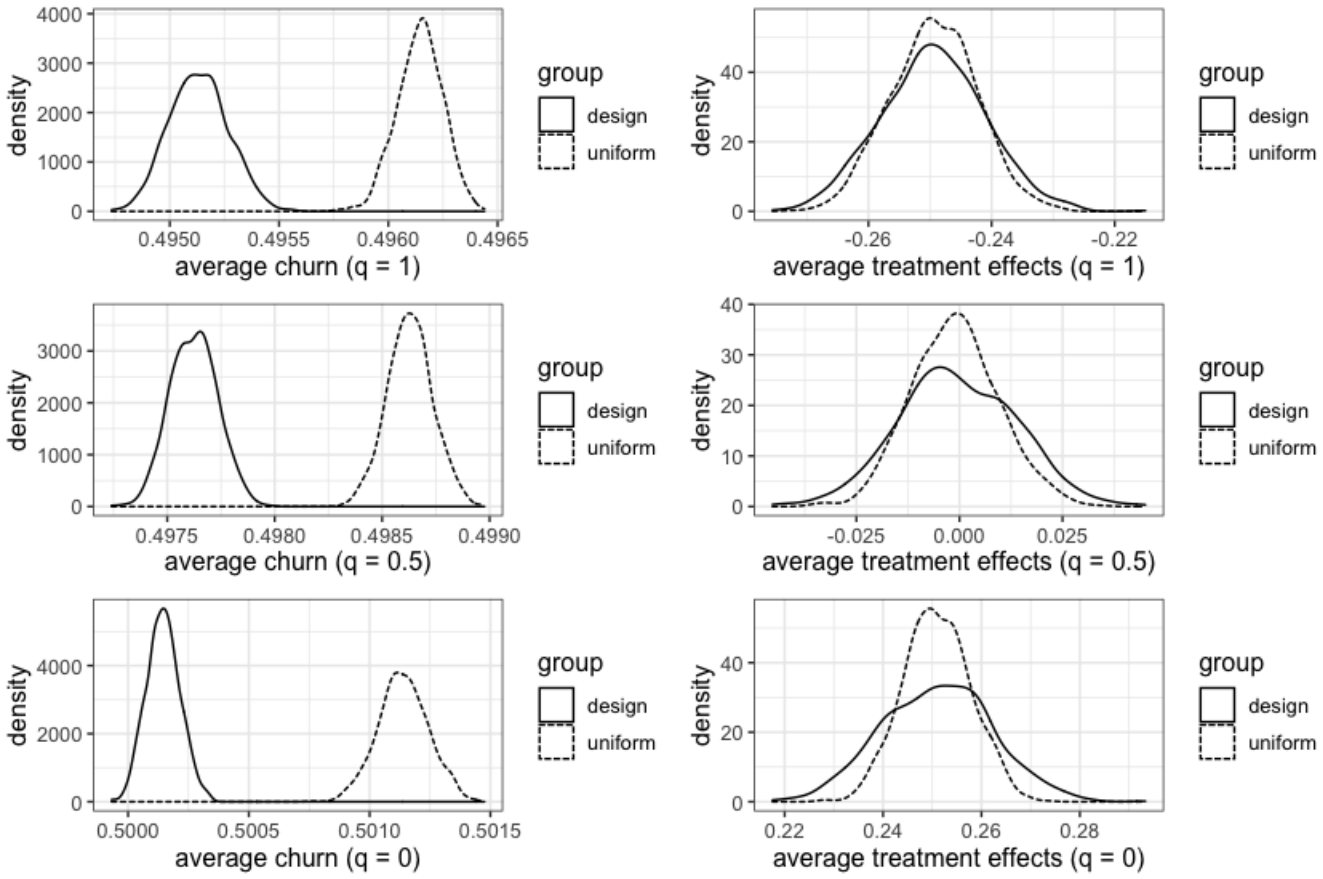


Figure D.2. Design vs. uniform policy on churn rate and ATE

D.3 Treatment Effects

First cohort: We plot the average treatment effect (ATE) and the average treatment effect on the treated (ATT) over time using churn and revenue as outcomes in Figures D.3 and D.4. When using churn as outcome, ATE has the smallest effect size and is marginally significant at month 3 (the discount ends at after month 2). The ATT stays negative but only marginally significant after month 10. That the ATT is bigger than the ATE in effect size suggests that our design policy assigns more subscribers on whom the discount tends to have a bigger effect to treatment, which is better than a uniformly random policy. The ATT on the subset of subscribers with the highest risk shows the biggest effect which provides supportive evidence to our prior and the choice of design policy. When looking at revenue, we see that the treatment effects are mostly negative 1.5 years after the experiment. This is likely due to two factors: (1) we might need to wait longer for a positive effect, which is consistent with our focus on long-term outcomes, (2) our design policy is not targeting the optimal set of subscribers, if we did so, as we will show in the policy learning section, the 18-month revenue impact will be positive.

Second Cohort: The ATT for churn and revenue in Figures D.5 and D.6 by treatment conditions. \$5.99/4 weeks and \$5.99/8 weeks, which give the smallest discounts, have the biggest treatment effect on churn reduction. This also shows up on the revenue plot. We can see that it takes much shorter for \$5.99/4 weeks to break even compared with other conditions (except for email only condition which doesn't have cost).

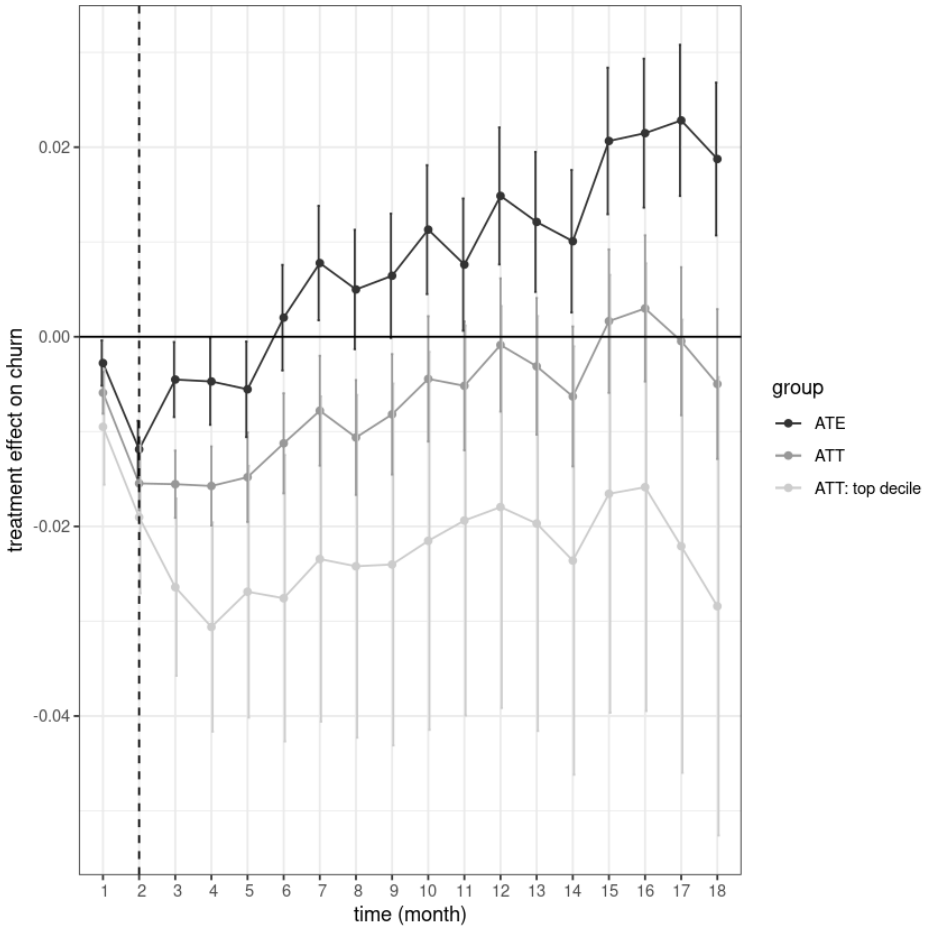


Figure D.3. Treatment effects on churn over time in the first cohort: ATE is the average treatment effect on all subscribers, ATT is the treatment effect on treated subscribers, ATT top decile is the ATT on subscribers with risk of churn in the top decile (month 2 is when the discount ends).

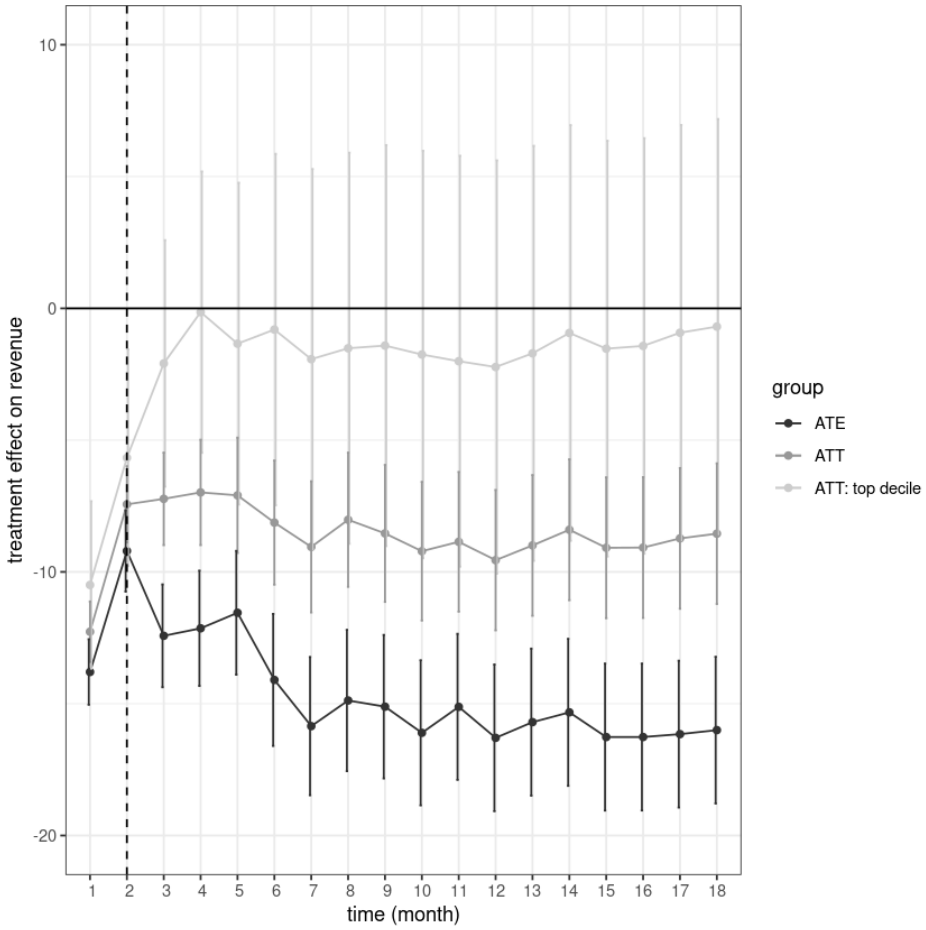


Figure D.4. Treatment effect on revenue over time in the first cohort: ATE is the average treatment effect on all subscribers, ATT is the treatment effect on treated subscribers, ATT top decile is the ATT on subscribers with risk of churn in the top decile (month 2 is when the discount ends).

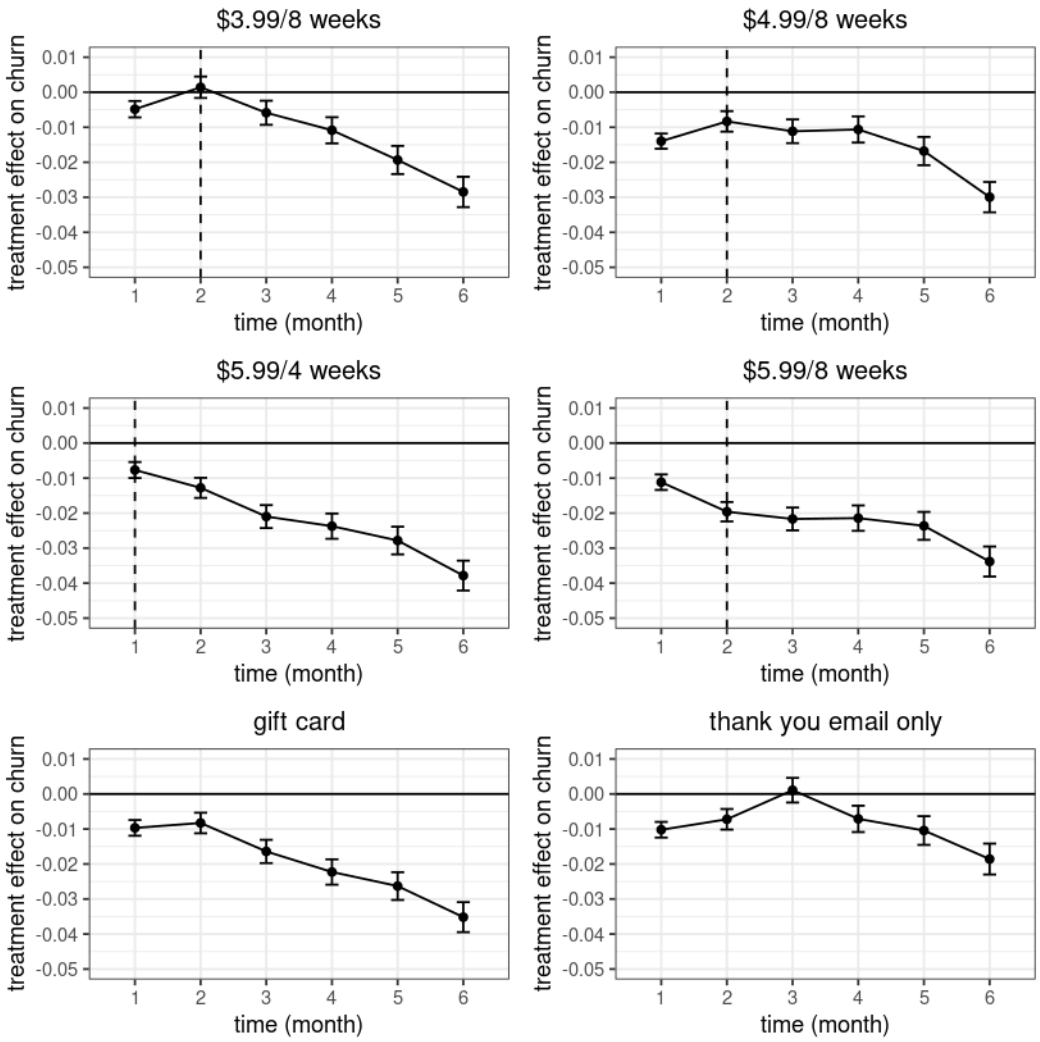


Figure D.5. ATT on churn in the second cohort by treatment conditions: the vertical dashed line indicates when the discount expires (there's no expiration date for gift card and thank you email), solid horizontal line is 0.

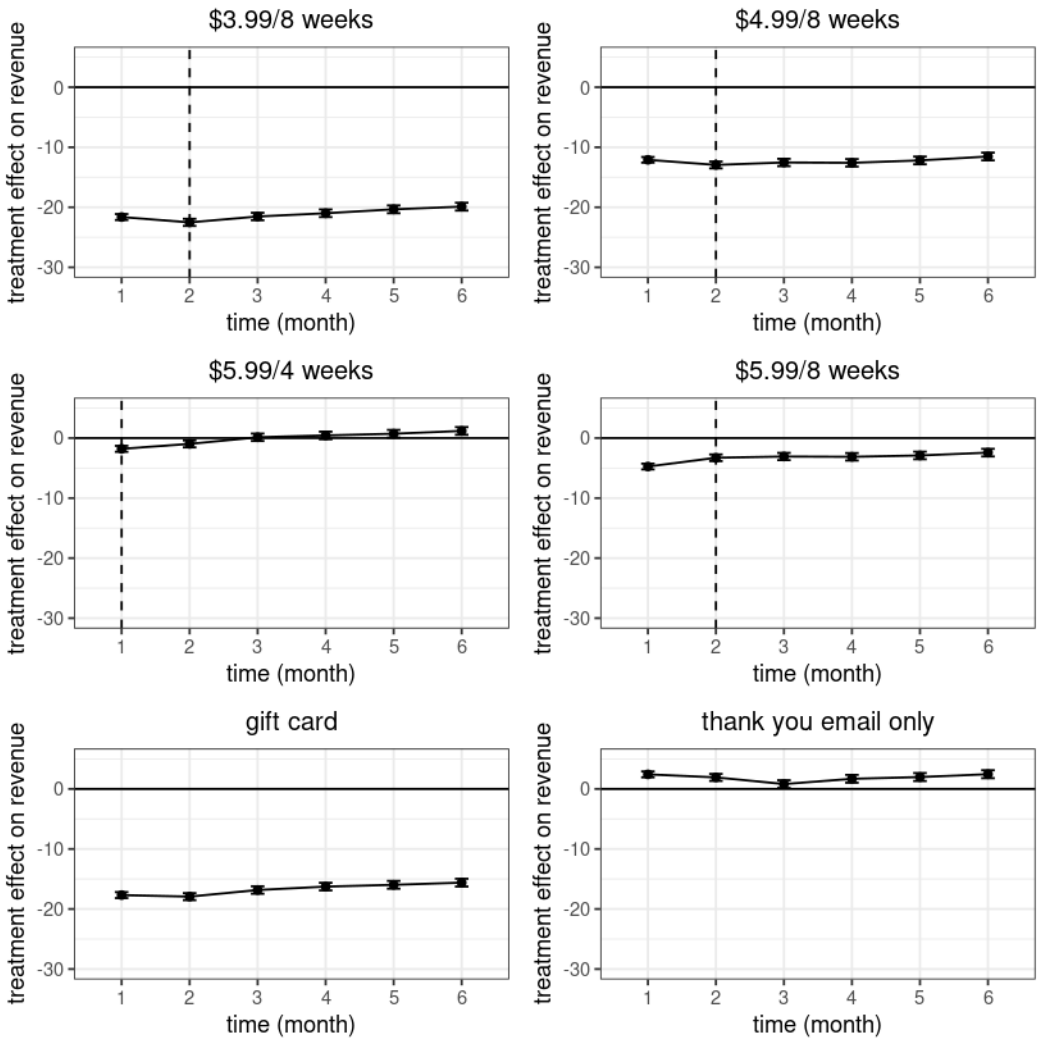


Figure D.6. ATT on revenue in the second cohort by treatment conditions: the vertical dashed line indicates when the discount expires (there's no expiration date for gift card and thank you email), solid horizontal line is 0.

D.4 Policy Evaluation

Because the optimal policy is a classifier that outputs a treatment probability for each subscriber, we can use the treatment probability to rank subscribers. When we compare policies we consider a policy constructed with different methods that sorts and treats q fraction of subscribers, $q \in 0.1, 0.2, ..., 1$ and compare the average revenue per subscriber under the policy with 95% bootstrap confidence intervals. We report results as a function of fraction of subscribers to treat because in both experiments the company has a capacity constraint (1,000 subscribers to treat in the first round experiment and 6,000 in the second). When we use the optimal policy to sort subscribers we can easily choose any fraction to target.

Using realized 18-month revenue as outcome, we first compare the value of an optimal policy with different benchmarks: a random policy, a policy that only treats subscribers with the highest risk of churn, and a policy that treats on one, which is the status quo or baseline.[35] The results are summarized in Figure D.7.[36] We can see that the optimal policy has a significant positive revenue impact relative to the baseline and also higher than random and risk score based policy. It increases 18-month revenue by up to $15 per subscriber relative to the benchmarks. A risk score based policy performs similar to a random policy, both are not significantly different from the baseline.

We then compare policies estimated with different classifiers: XGBoost, a shallow decision tree (with maximum depth 5) and logistic regression as summarized in Figure D.8. We choose decision tree and logistic regression because policy based on these two models is more transparent and easy to interpret, so else being equal, we prefer such models. The optimal policy performs the best, logistic regression also performs better than decision tree, although none of them are significant. Logistic regression and decision tree also don't perform better than the baseline.

We also look at optimal policies learned with a subset of features. Specifically, we consider a classifier trained using XGBoost on full features, with only content consumption information (e.g., how many articles a subscriber has read in the last week, month, in each section), or only with account information (e.g., credit card expiration date, all account activities). The results are in Figure D.9. The full information policy in general has higher point estimates but they are not significantly different from each other. Only the full information policy ever beats the baseline.

Last, we compare the doubly-robust score based approach for policy learning with two other popular choices as discussed in section 5.2. i.e., an indirect method that first estimates an outcome function for each action and use it to estimate CATE of each subscriber, and a direct method that estimates CATE directly. We use XGBoost and causal forest to implement these two methods, the results are in Figure D.10. doubly-robust approach works the best and causal forest works better than outcome model.

We repeat the exercise using imputed 3-year revenue as outcome and compare the optimal

[35] This is also the best performing uniform policy, that is, the action that performs the best on average is to not treat anyone.

[36] Note that the value plotted on the y-axis is the value difference between a given policy and a policy that treats no one, which we call the baseline.

policy with random and risk score based policy in Figure D.11.

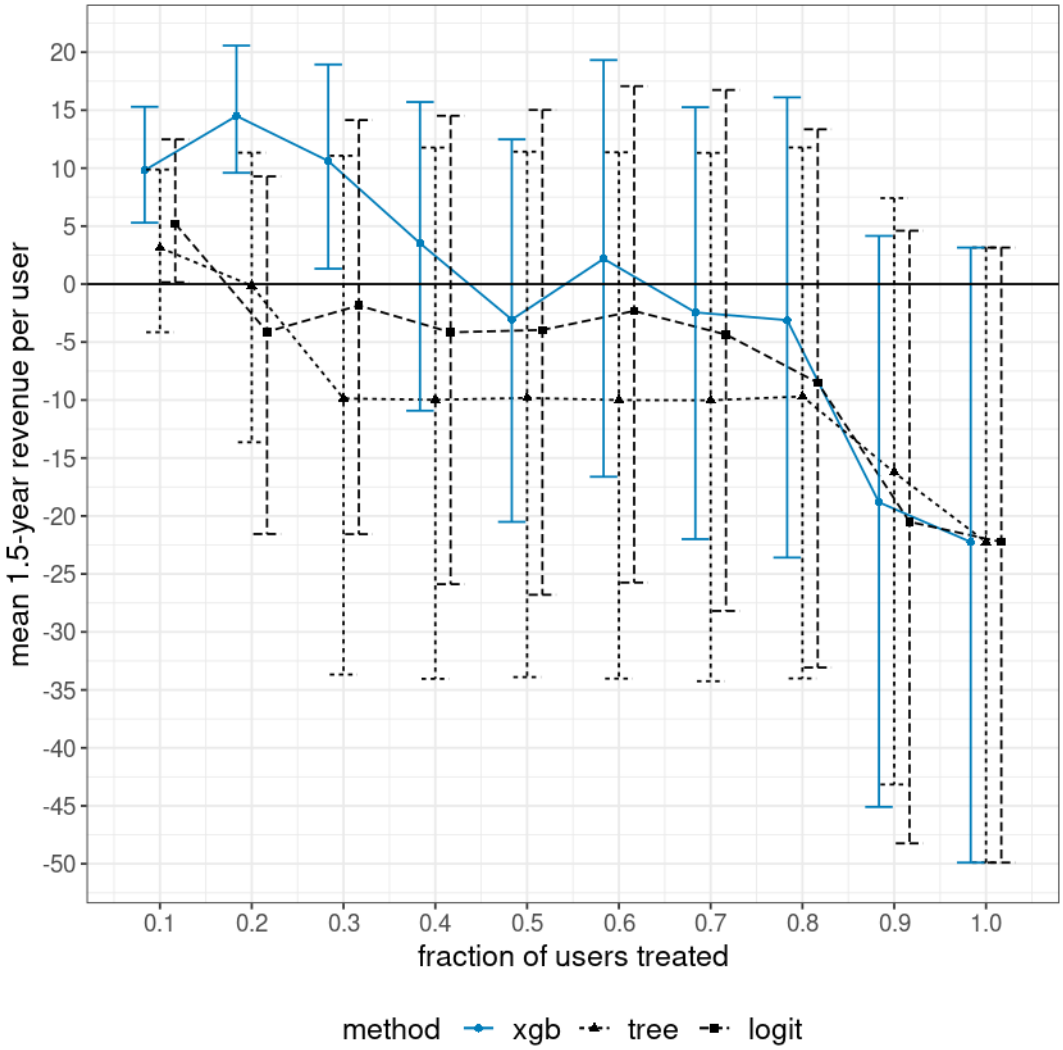


Figure D.8. Comparing an optimal policy with different models: the optimal policy is learned via XGBoost, it ranks subscribers based on estimated treatment probability and treat the top q fraction of them, decision tree and logistic regression is an optimal policy estimated using the corresponding model.

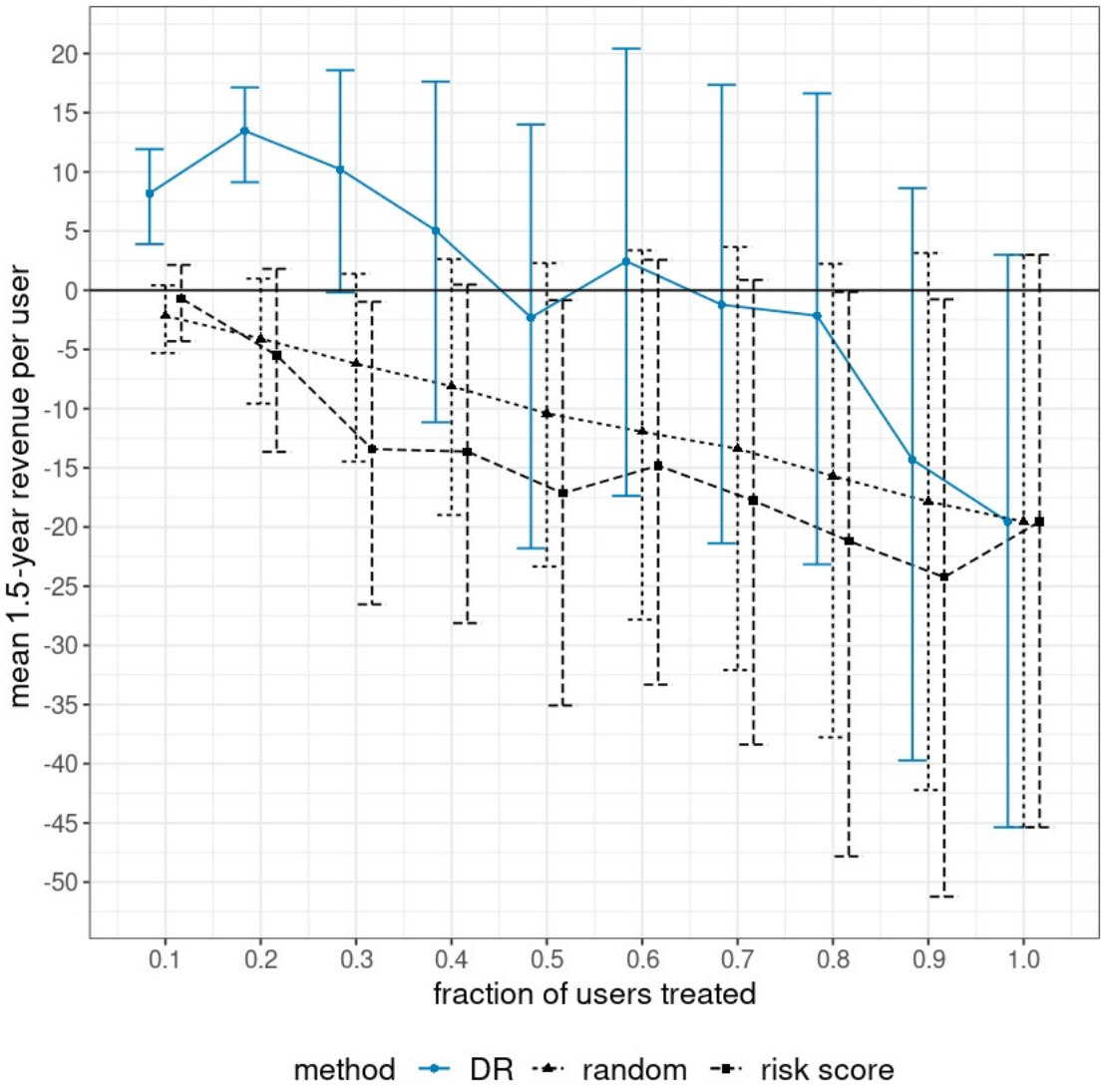


Figure D.7. Comparing an optimal policy with benchmarks: the optimal policy is learned via XGBoost, it ranks subscribers based on estimated treatment probability and treat the top q fraction of them. Random policy is to randomly treat q fraction of subscribers. Risk score policy is to rank subscribers by their estimated risk score and treat the top q fraction of them.

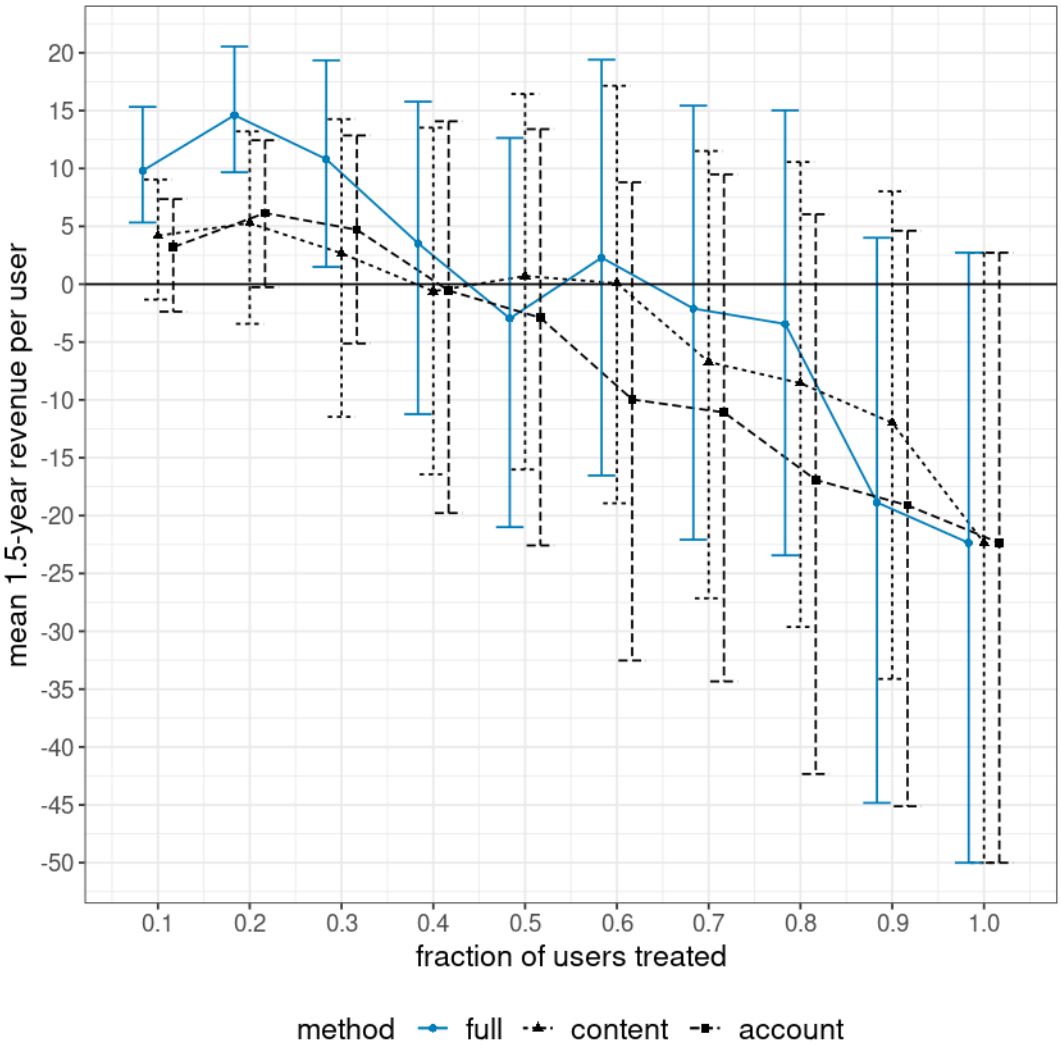


Figure D.9. Comparing an optimal policy with different information sets: the optimal policy is learned via XGBoost, it ranks subscribers based on estimated treatment probability and treat the top q fraction of them. Content and account is an optimal policy estimated with only those information.

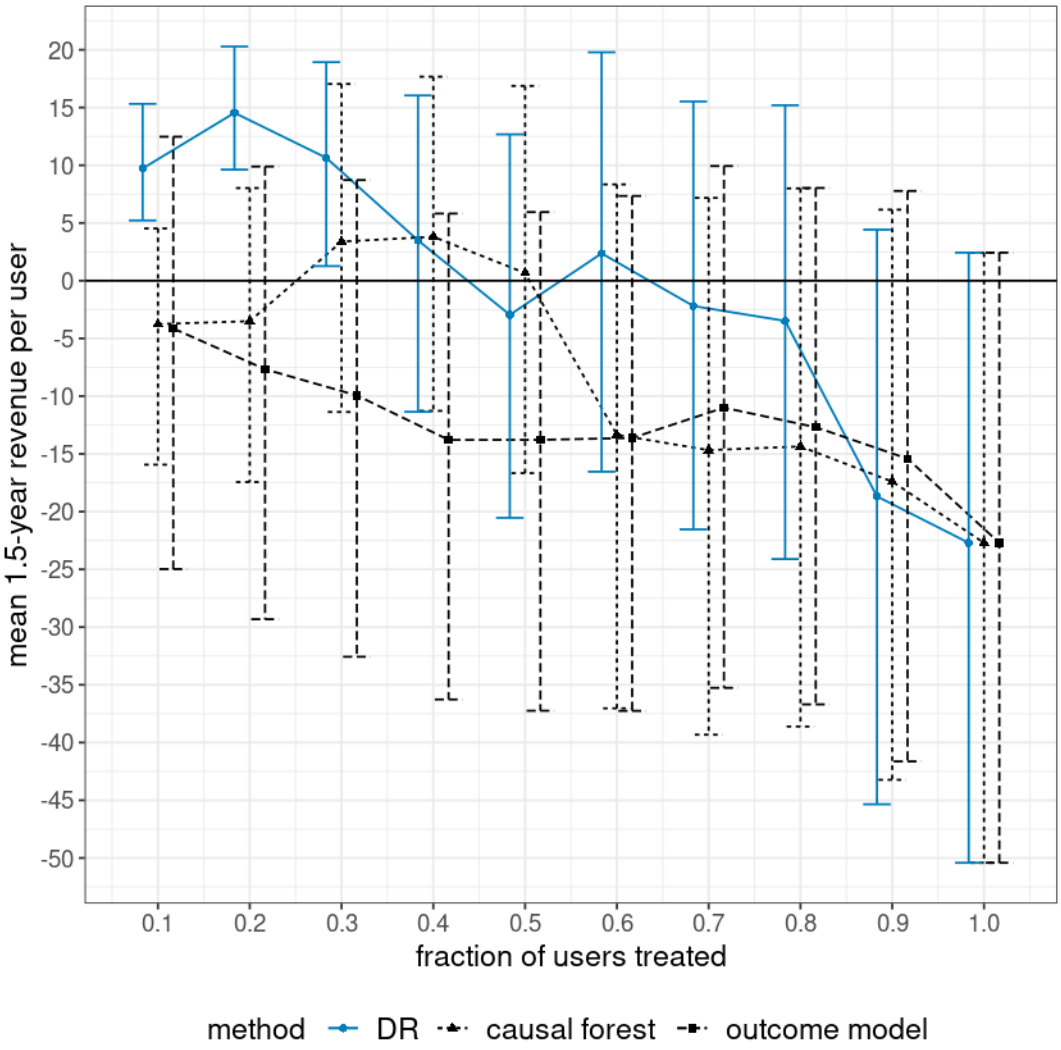


Figure D.10. Comparing an optimal policy with different methods: the optimal policy is learned via XGBoost, it ranks subscribers based on estimated treatment probability and treat the top q fraction of them. Outcome model (XGBoost) and casual forest ranks subscribers by CATE estimated with the corresponding model.

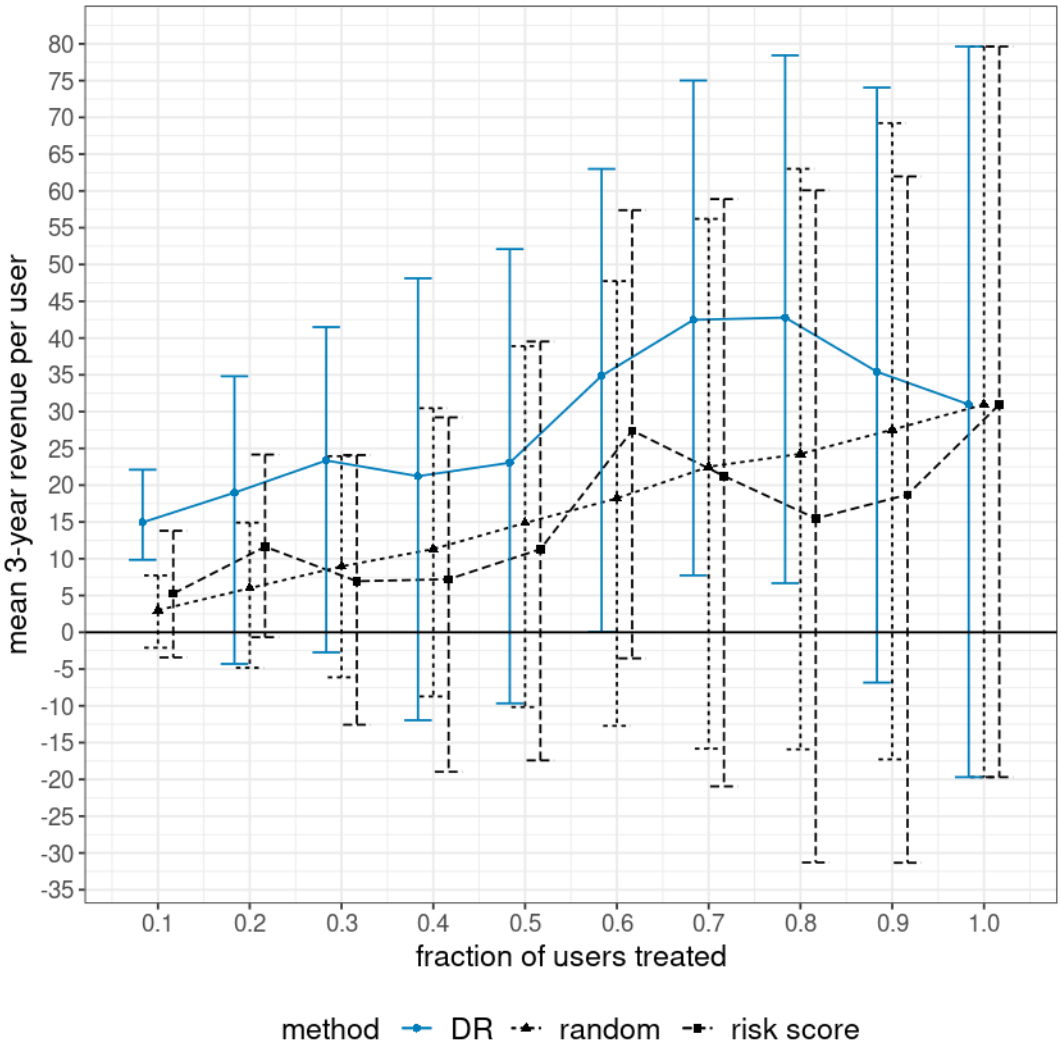


Figure D.11. Comparing an optimal policy with benchmarks: using imputed 3-year revenue as outcome.

D.5 Policy Interpretation

The feature importance plot[37] is in Figure D.12. The top 3 features are risk score (the pre-treatment risk of churn), tenure (how long a subscriber has been a subscriber) and number of sports articles read in the last 6 month (a measure of content consumption and how active a subscriber is on the website). Zip code and other content and account info also show up in the top 20 features. We plot accumulated local effects (ALE)[38] for the top three features. ALE shows how treatment probability changes when we vary the values of risk score, tenure and number of sports articles read, respectively. The optimal policy treats subscribers with shorter tenure (more recently registered subscribers) with higher probabilities. The relationship between treatment probability and number of sports articles read is not monotone: the probability is low for very inactive and active subscribers but higher for subscribers in between. The relationship with risk score is interestingly also not monotone, for subscribers with the highest risk scores the treatment probabilities are higher, this is consistent with our prior. But for some subscribers with very low risk score, the treatment probabilities are even higher. This also highlights the risk of targeting solely based on risk scores.

[37] Feature importance measure works by calculating the increase of the model prediction error after permuting the feature. A feature is more important if permuting its values increases the model error, because the model relied more on the feature for the prediction. A feature is less important if permuting its values keeps the model error unchanged, because the model ignored the feature for the prediction.

[38] ALE is similar to partial dependence but takes feature correlations into account: instead of averaging over distribution of other features in the whole dataset, ALE averages over the distribution of other features conditional on the value of a focal feature (Apley and Zhu, 2016).

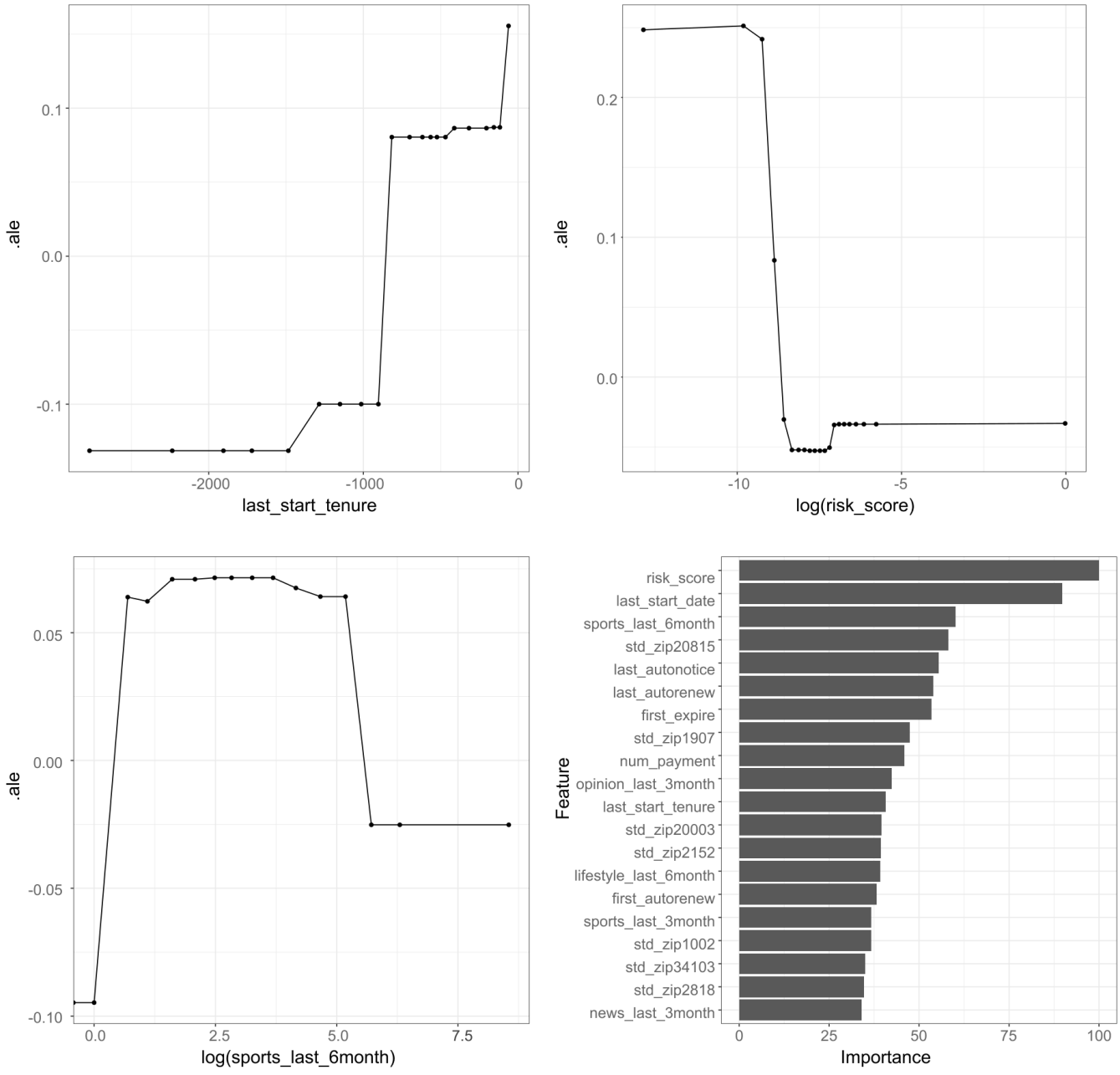


Figure D.12. ALE and featute importance plot

D.6 Non-stationarity

Covariate shift means the distribution of subscriber features are quite different between the two cohorts, when this is the case, the policy learned on the first cohort might not perform well on the second cohort because we are likely facing a different population. However this doesn't seem to be the case in our data, Figure D.13 and D.14 show the distribution of covariates in the two cohorts and we can see that they are quite similar.

Then we look at concept shift which is the change in relationship between outcome of interest, covariates and actions. We focus on the treatment effect here. Due to logistical constraints, we only have one common treatment between the two cohorts, i.e., \$4.99/8 weeks, we plot the treatment effect over time from both experiments. Because we know the two populations are comparable in terms of observed covariates, so the difference in treatment effect can be attributed to concept shift. The result is shown in Figure D.15. We can see that the treatment effect over time look somewhat different, so when learning the policy for the next cohort we only use data from the second cohort. Alternatively, we can pool data from both experiments but assign lower weights to observations in the first cohort to reflect the fact that these data is somewhat stale (Russac et al., 2019).

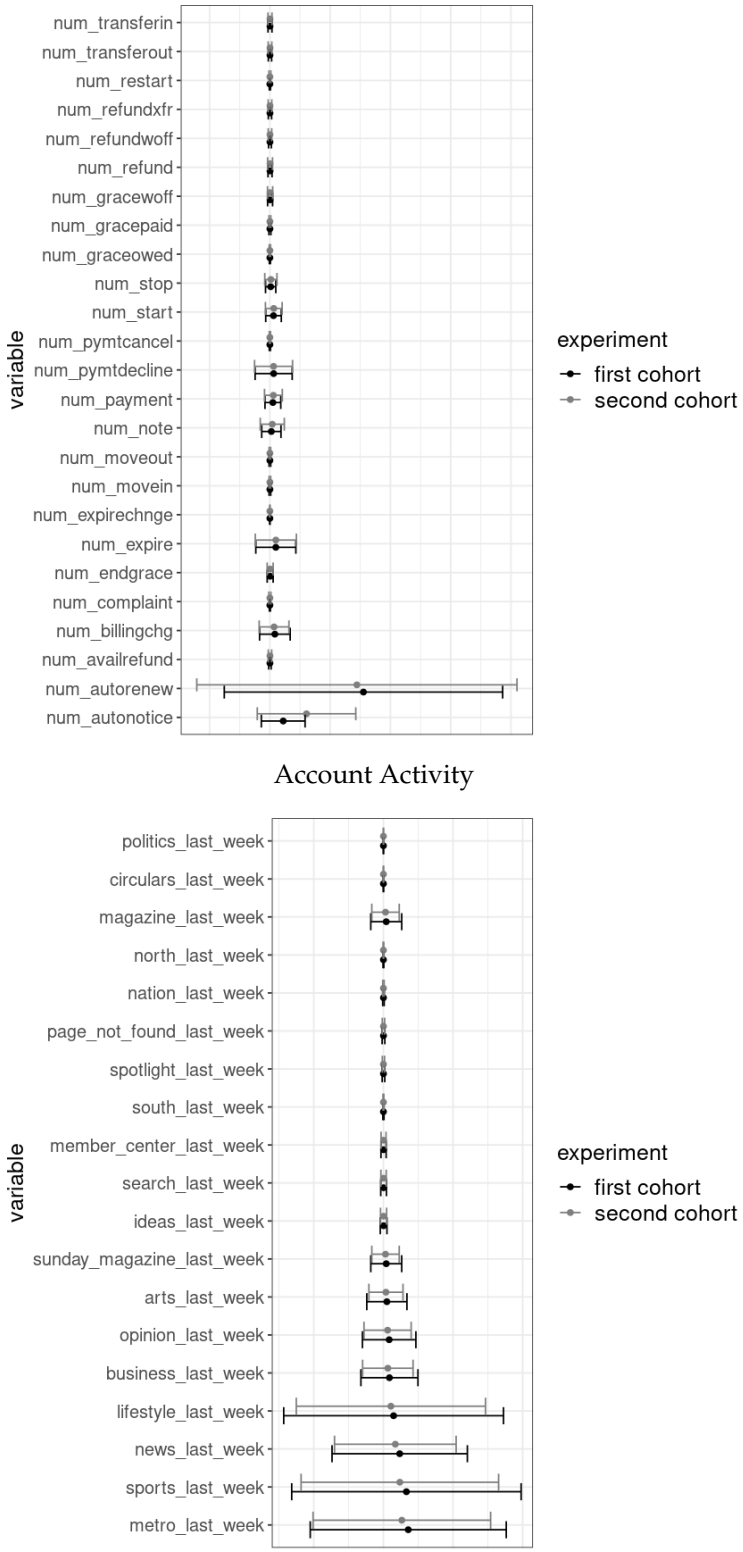
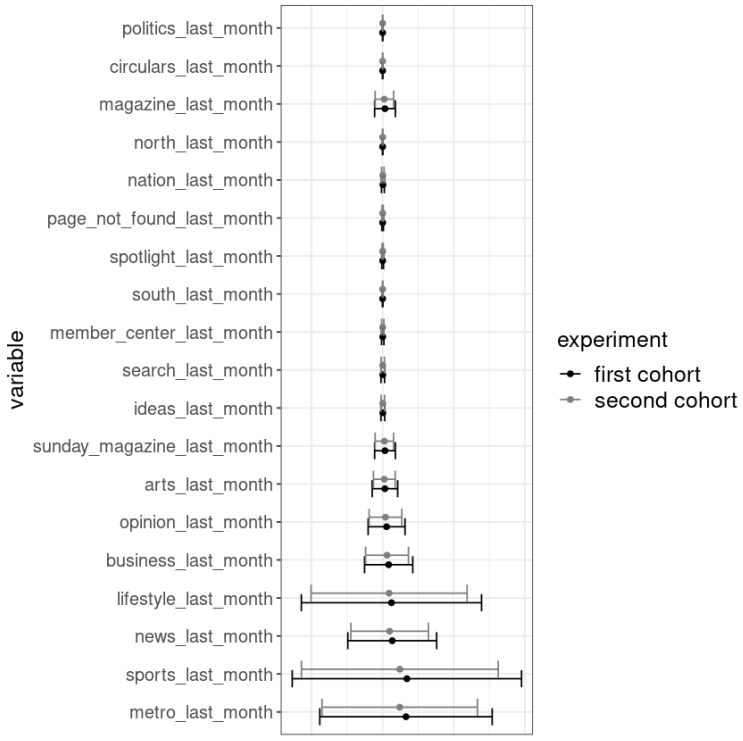
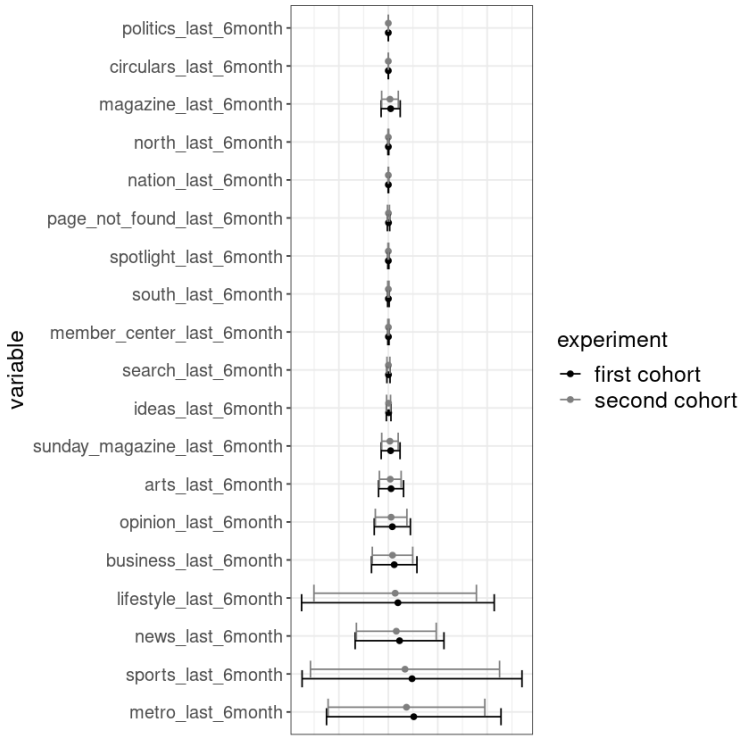


Figure D.13. Covariate shift: comparing the distribution (the two ends are 2.5 and 97.5 percentile) of continuous covariates (account activity and content consumption)



Content Consumption: Last Month



Content Consumption: Last 6 Month

Figure D.14. Covariate shift: comparing the distribution (the two ends are 2.5 and 97.5 percentile) of continuous covariates (content consumption)

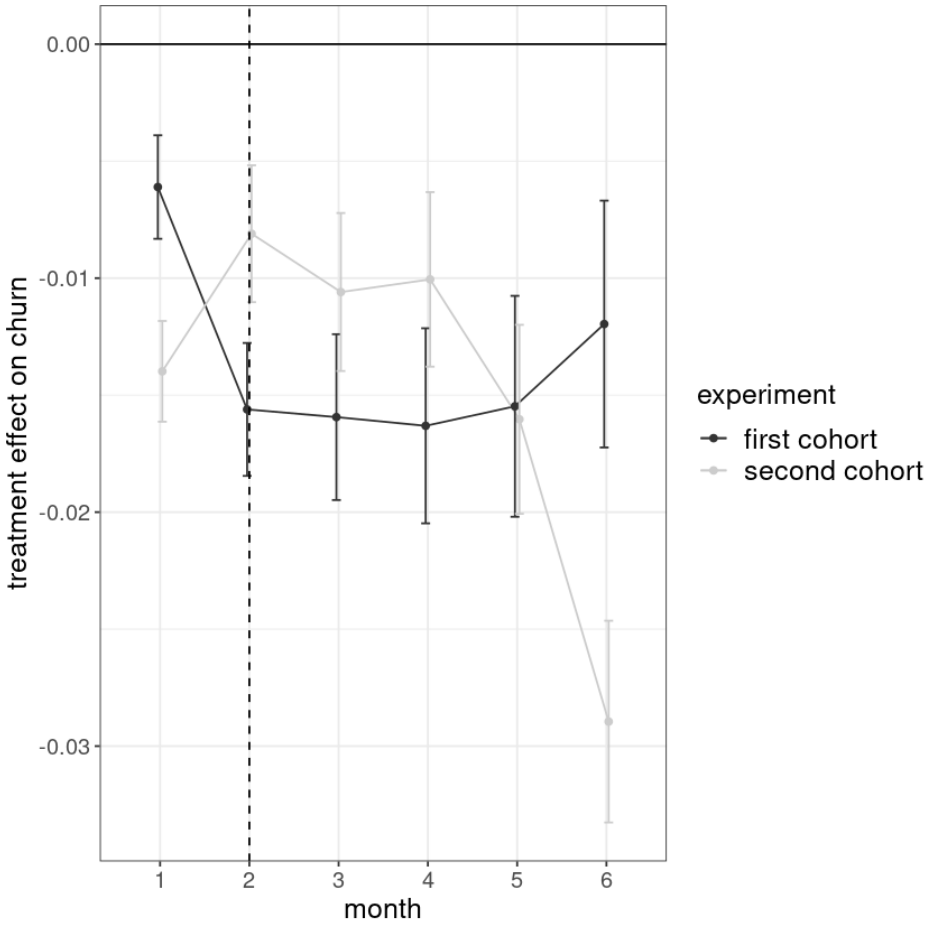


Figure D.15. Concept shift: comparing the ATT overtime for two cohorts. This is the treatment effect of the condition \$4.99/8 weeks relative to the control, we can only compare this condition because this is the only common treatment condition between the two experiments. The 95% confidence intervals overlap for most of the time periods but month 1 and 6 are quite different.

D.7 Power Calculations via Simulation

Before running the experiment on the first cohort we conducted power simulation to see if we have enough power to detect any difference between alternative targeting policies. And we suspected that given the small number of treated subscribers our experiment might be underpowered. We vary two parameters: q, the percentage of subscribers targeted under the model and τ, the effect size. For example, $q = 0.01, \tau = 0.1$ means that under the model we target the top 1% of subscribers and the discount will lower the targeted subscribers' churn risk by 10%. $Y0$, the outcomes without treatment is observed in the data, which is whether a given subscriber churned (churn = 1, not churn = 0). We simulate $Y1$ in the following way: for any subscriber whose $Y0$ is 0, we assume that the treatment won't *increase* the churn risk so her $Y1$ is also 0. For any subscriber whose $Y0$ is 1, we flip a coin, with probability $1 - \tau$ it stays 1 and with probability τ it becomes 0, τ is the effect size. After simulating the full schedule of potential outcomes we use the design policy discussed in Section 7.2 to simulate treatment assignment. The treatment assignment determines, for each individual, which potential outcome is revealed to us. This is considered one simulated experiment. Then for a fixed value of q and τ and full schedule of potential outcomes, we repeat the simulated experiment for 100 times and calculate the power (percentage of simulated experiments that have a significant result) of different estimators. We look at both churn rate and implied revenue as our outcome measure

We find that for ATT using both churn rate and implied revenue as outcome, we have over 80% of power only when the effect size is bigger than 20%. And for ATT under model based targeting, we also need the effect size to be bigger than 20% for 80% power. For ATT under random targeting we will need even a bigger effect size at 30%. We also calculated the total gain and loss for the campaign under the design policy and what it would be if we were to target using model based policy. We'd expect gains by using model based policy when effect size is moderately big (over 25%) and we don't target too many subscribers (1 or 2%). It turns out that our ATT is -28% or -44% (depending on whether we use email sent or email open as treatment), it's within the range of τ that we covered in the simulation and it's bigger than we'd expected.