Temporary Premium Benefits on a Multi-Channel Platform: Effects on Engagement and Monetization

Matthew McGranaghan* Zachary Nolan†

December 3, 2025

Abstract

In this paper, we study how temporary premium benefits affect user behavior on a multi-channel social livestreaming platform. We collect novel user-level panel data which captures viewership, chat engagement, and subscription purchases across popular creator channels on the platform. To estimate causal effects, we leverage quasi-exogenous variation in the allocation of premium benefits together with double-robust machine learning. The average benefit recipient increases time spent on the platform by more than three hours during the benefit period, and is 11% more likely to use the platform's chat feature. These behavioral effects are persistent for three months after the benefit period ends, and spill over to other creators on the platform. In the short run, temporary benefits cannibalize paid subscriptions to the channel where they apply, but generate more than offsetting increases in subscriptions to other channels. In the long run, benefit recipients have a greater propensity to subscribe to all observed channels. Subscription spillover effects are largest among heavy platform users who engage less with the channel where benefits apply, suggesting that targeting users outside their preferred content areas can generate cross-channel spillovers while mitigating cannibalization. We use our treatment effect estimates within a multi-objective optimization framework to investigate how platforms can target promotional access to temporary premium benefits to balance different, potentially competing, objectives.

Keywords: causal inference, machine learning, digital marketing, media consumption, targeting, policy evaluation

*Department of Business Administration, University of Delaware, mmcgran@udel.edu
†Department of Marketing, University of Arizona, znolan@arizona.edu

1 Introduction

Firms use a variety of promotional strategies to acquire, retain, and monetize consumers. Many promotions give consumers temporary access to products or product features that are otherwise sold at a premium. These promotions may provide information, encourage habit formation, or build brand loyalty.

In digital and service-based markets, firms increasingly allocate temporary premium benefits using automatic promotions—promotions that grant access to benefits without requiring any action or payment by the consumer. For example, a rental car company might surprise a customer with a free vehicle upgrade at check-in, an online retailer might upgrade a customer's regular shipping to overnight delivery at no charge, or a mobile game developer might award a power-up to a user predicted to be at risk of churning. These examples are in contrast to opt-in promotions, which require consumers to take action to receive promotional benefits, such as by redeeming a coupon or signing up for a free trial.

Despite having similar objectives, automatic and opt-in promotions have several important differences. First, automatic promotions may be more difficult for sophisticated consumers to predict and exploit. This may lead to a reduction in the cannibalization of full price sales. Cannibalization may occur, for example, if a free trial is seen as a substitute to a paid subscription, or a coupon is used by individuals that would have purchased without a discount (Bawa and Shoemaker, 2004). Second, by removing the need for consumer awareness or initiative, automatic promotions give firms more precise control over which users experience the offer, potentially improving targeting efficiency. These differences are consistent with broader trends in marketing toward automation and algorithmic decision-making. At the same time, opt-in promotions are becoming more limited or unavailable in some settings, as firms consider them less effective tools for building a long-term consumer base (Kan, 2020; Foubert and Gijsbrechts, 2016).

Evaluating the impact of promotions has historically posed several empirical challenges. First, with opt-in promotions, self-selection prevents firms from controlling who actually

receives the promotional benefits. Firms can control who is exposed to the promotion, but only consumers who go through the effort to use it receive the benefits. Moreover, the consumers who self-select into receiving these benefits may be systematically different from those who choose not to (Datta et al., 2015). Second, attribution problems can complicate the analysis if promotional uptake co-occurs with factors such as new content releases, platform updates, or other marketing activities (Goldfarb et al., 2022). For example, a consumer may sign up for a free trial on a streaming platform in anticipation of a new season of their favorite show. Firms can use automatic promotions, which give the firm control over who receives promotional benefits and when, to address both self-selection and attribution problems and more accurately learn about customer responsiveness.

Marketing managers also face the decisions of which outcomes to measure and over what time horizons to measure them. For example, focusing exclusively on paid conversion may overlook other valuable effects, such as changes in product usage that may build brand loyalty, changes in social engagement that may drive network effects, or spillovers into additional products or offerings (Datta et al., 2018; Chae et al., 2017; Pattabhiramaiah et al., 2019). These behavioral changes all contribute to long-term consumer value but are not captured by any single metric. Moreover, evaluating the impact of a promotion over too short of a time horizon may miss important longer-run effects (Yang et al., 2024). Navigating these decisions is essential for accurately measuring the effects of promotions and implementing effective marketing strategies.

Firms with multiple services face an additional challenge: promotions on one service can affect user behavior across their entire portfolio. These cross-service spillovers may increase engagement and purchases of other offerings, or they may cannibalize activity and revenue from those services. Firms must understand these effects to design promotional strategies that optimize value across all their services, not just the promoted one.

Despite these challenges, digital services are particularly well-positioned to benefit from automatic promotions. Digital services possess large amounts of user behavioral data that

can inform targeting, can easily implement promotions through their digital infrastructure, and maintain a high degree of control over the user experience. Additionally, many digital services offer a variety of product tiers and premium benefits, including ad-free content, exclusive content, and account customization, which can also be offered as promotional benefits.

In this paper, we study how digital platforms can effectively use automatic promotions to allocate temporary premium benefits by answering the following research questions. What are the causal effects of receiving temporary premium benefits on user behavior, and how wide-ranging and persistent are these effects? How do heterogeneous responses to automatic promotions inform targeting strategy? Which users should platforms target to achieve different, potentially competing, objectives?

To answer these questions, we analyze user behavior on a popular live-streaming platform that features a variety of content across creator-specific channels. Users on the platform can watch content, engage in chat, and purchase subscriptions to individual channels. The platform uses an algorithm that quasi-randomly allocates some users temporary (30-day) access to premium benefits. This allocation creates a natural experiment that we use to identify the causal impact of temporary premium benefits, similar to a free trial, on user behavior.

We use channel-specific audience information at the time of each algorithmic allocation to define treated users, those who were present and received the promotion, and control users, those who were present but did not receive the promotion. Using double-robust machine learning (Chernozhukov et al., 2018), we model the platform's promotion allocation algorithm and a variety of short and long-term user-level behavioral outcomes. These outcomes include watch behavior, social engagement, and subscription behaviors on both the trial channel, where the user received premium benefits, and all other observed channels on the platform. To understand the dynamic effects of these promotions, we measure each behavior over a variety of time horizons, from the day in which the promotion was received to three

months after the benefit period expired. To explain variation in the allocation model and user-level responses, we use a broad set of behavioral covariates including prior watch behavior, chat engagement, and subscription history. These variables can capture patterns that prior literature identifies as important sources of heterogeneity, including prior experience effects (Reza et al., 2021), habitual usage patterns (Shah et al., 2014), and variety-seeking behavior (McAlister and Pessemier, 1982; Kim et al., 2002). We use causal forests (Athey et al., 2019) to characterize heterogeneity in treatment effects based on pre-treatment user behaviors. Finally, we incorporate user-level treatment effects into a multi-objective framework (Rafieian et al., 2024) to measure tensions between optimizing different platform objectives and evaluate the performance of counterfactual promotion targeting policies.

Temporary access to premium benefits results in significant and sustained changes in user behavior. The average user increases viewership on both the trial channel and other channels, with effects starting immediately after the promotion is received and lasting throughout the post-trial observation period. These changes in viewership translate to managerially significant increases in platform activity. The average user watches approximately 150 minutes of additional content per month across the platform, with the largest percentage increases among historically lighter users. This promotion also increases the propensity and intensity of a user's chat behavior, both on the trial channel and other channels, suggesting deeper platform engagement. For paid subscriptions, in the short term, promotion recipients are less likely to purchase subscriptions to the channel on which promotional benefits were received. However, users are more likely to purchase subscriptions to other channels on the platform during the promotional period and in the months after temporary benefits expire. While short-term decreases in paid subscriptions to the trial channel likely reflect cannibalization, the strong positive spillovers result in a net positive effect on subscription revenue. Taken together, these results highlight the importance of considering a wide range of outcomes over an appropriate time horizon when evaluating promotional effectiveness.

We find managerially-relevant heterogeneity in the effects of temporary access to premium

benefits across user segments. Subscription spillover effects are largest among users who engage less with the trial channel relative to other channels, suggesting that cross-channel promotions can be effective when targeting users outside their preferred content areas. In contrast, retention effects are strongest among users who are less engaged with both trial and other channels, suggesting that these promotions can be used to retain users who may be at risk of churning. Notably, targeting the most loyal platform users yields low incremental returns on both outcomes.

We use a multi-objective optimization framework to quantify the tradeoffs inherent in implementing an automatic promotion targeting strategy that balances paid subscriptions on the trial channel, paid subscriptions on other channels, and user retention on the platform. We find limited overlap in the users targeted across different single-objective policies, highlighting tensions between competing objectives. Multi-objective optimization policies substantially outperform both the observed and random allocation policies, which perform similarly to one another. This reveals high opportunity costs associated with maximizing any single outcome at the expense of others.

The remainder of the paper is structured as follows. Section 2 overviews the relevant literature. Section 3 describes the empirical context, including the automatic promotion we study. Section 4 describes data sources and sample construction. Section 5 outlines our empirical strategy, including the identification approach, causal framework, and estimation. Section 6 presents the main results and explores heterogeneity across user types. Section 7 discusses multi-objective optimization, addressing how platforms can balance competing goals when targeting free trials. Section 8 concludes.

2 Literature

This research builds on several literatures, including how premium benefits affect paid conversion and user behavior, the spillover effects of promotional interventions, and strategies

for targeting promotions to achieve different outcomes.

A large body of literature explores how selective exposure to premium benefits through free trials, freemium models, and paywalls affects user behavior. This research addresses two related questions. First, what drives adoption of premium benefits? Bapna and Umyarov (2015) find that peer influence increases premium conversion in a freemium service, while Oestreicher-Singer and Zalmanson (2013) show that both social and content engagement drive premium conversion. Yoganarasimhan et al. (2023) demonstrate that shorter free trial duration can increase customer acquisition, retention, and profitability. Reza et al. (2021) find that prior usage levels are important predictors of the take-up and effects of free samples of experience goods. Second, how does access to premium benefits change user behavior? Iyengar et al. (2022) find that retail membership programs increase purchase frequency and basket size, while Datta et al. (2018) show that music streaming access increases consumption quantity and diversity. Our research contributes to both streams by studying how temporary access to premium benefits affects engagement and consumption, which in turn influences subscription adoption. In our context, access is granted via an automatic promotion. While most previous studies analyze opt-in promotions, one exception is von Wangenheim and Bayón (2007), which studies the impact of unexpected service upgrades. We extend these findings by studying a multi-service platform where users receive automatic promotions without opting in or selecting the upgraded service.

For multi-product firms, promotional interventions can generate spillover effects that require careful management through strategic design and targeting. Bawa and Shoemaker (2004) show that free samples can increase sales among prior and new customers, but may also cannibalize planned purchases. Sahni et al. (2017) examine cross-category spillovers in retail contexts, demonstrating how promotional activities in one category affect demand in related categories, while Pattabhiramaiah et al. (2019) show how firms with multiple distribution channels can internalize cross-channel effects. Chae et al. (2017) show that seeded word-of-mouth campaigns redirect attention across products, increasing target prod-

uct discussions while decreasing competitor discussions. We contribute to this literature by quantifying promotional spillovers in a digital platform context, showing that promotions on less-preferred channels generate large positive spillovers to preferred channels, a finding with direct implications for cross-channel targeting.

Prior research demonstrates the trade-offs associated with different promotional targeting strategies and the importance of measuring heterogeneous responses. Datta et al. (2015) highlight trade-offs between customer acquisition and quality, finding that trial-acquired customers often have lower lifetime value than customers from other channels. Foubert and Gijsbrechts (2016) document the "double-edged" nature of free trials–they accelerate adoption but may attract lower-quality customers. Ascarza (2018) and Lemmens and Gupta (2020) challenge conventional targeting approaches, showing that targeting high churn-risk customers can be suboptimal compared to targeting based on predicted treatment sensitivity. Yoganarasimhan et al. (2023) evaluate personalized targeting policies that assign different targeting treatments based on individual-level predictions of a single outcome of interest. We extend this targeting literature by developing and implementing a multi-objective optimization framework (Rafieian et al., 2024) that balances multiple firm objectives, demonstrating substantial performance gains over single-objective and benchmark policies.

Finally, this research contributes to a growing literature on social live-streaming platforms, where the variety of observable behaviors and contexts create unique environments for studying media consumption behaviors and the effects of promotional interventions (Lin et al., 2021; Lu et al., 2021; Förderer et al., 2023; Simonov et al., 2023; Huang and Morozov, 2025; Kim et al., 2025).

3 Empirical Context

3.1 Twitch

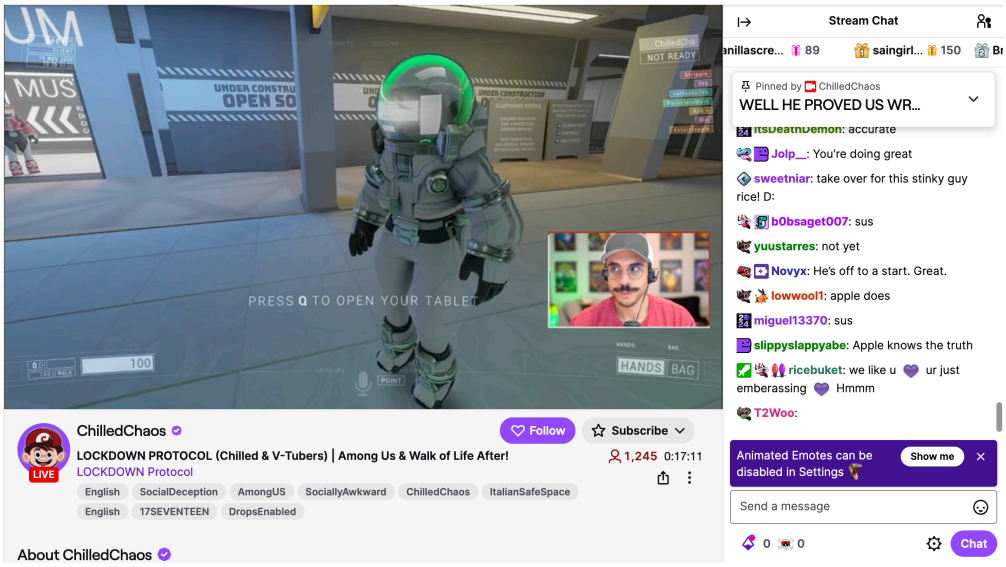
We study user behavior on Twitch,[1] a platform that connects content creators (referred to as "streamers"), users ("viewers"), and advertisers. Twitch is one of the most popular social live streaming platforms.[2] The live-streaming format creates an engaging experience where users witness events as they unfold and interact with creators and other users through chat. Twitch, like most live-streaming platforms, helps creators monetize their channel through ads and subscription revenue.

Each creator operates their own channel on the platform. Creator content often focuses on video games, but includes a wide range of topics such as music, politics, and conversations with the channel audience ("Just Chatting"). Figure 1 shows the typical user experience on a channel.

Users can subscribe to a creator's channel, enabling channel-specific benefits including ad-free viewing (e.g. eliminating pre-roll ads) and special chat privileges (e.g., badges and emojis). Users can subscribe to these benefits in one of three ways. They can self-subscribe for a fee (paid subscription) or by linking their Amazon Prime membership (Prime subscription). Users can also receive *gift* subscriptions from other users. *Gifted* subscriptions offer the same benefits to the recipient as self-subscriptions. Users gift subscriptions to other users for many reasons: to financially support creators, to strengthen a creator's community, to engage in a random act of kindness. Users can gift as few as one or up to one hundred subscriptions at a time to a creator's audience. All subscriptions last for 30 days and cost about $4.99.[3,4]

[1] https://twitch.tv
[2] https://streamscharts.com/platforms
[3] Prices vary based on device and region.
[4] https://help.twitch.tv/s/article/gift-subscriptions?language=en_US

Figure 1: Twitch User Experience



Notes: Screen capture of a typical user experience when viewing a creator's channel. The user can see live-streamed video content, a webcam view of the creator, a chat panel, and information on the current stream (creator's name, stream content tags, time since start of stream, user subscription status, number of viewers).

3.2 Temporary Premium Benefits

When a user purchases more than one gifted subscription at a time, an algorithm determines which users receive those subscriptions. We refer to these purchases as automatically-allocated subscriptions. Figure 2 illustrates an announcement of a user purchasing forty automatically-allocated subscriptions that are distributed across the channel's community.

Twitch describes the algorithm as follows (emphasis added):

> We use an algorithm to help us select gift recipients *starting with eligible viewers* in chat, then followers, mods, and other factors that identify members of a community. Our algorithm also *avoids giving trolls subs.* We are constantly improving our algorithm to detect this behavior.[5]

This mechanism allocates an automatic promotion through which users receive premium benefits without expending any effort to opt in. While it shares features with free trials and gifts, this promotion differs in several ways. Like a free trial, this promotion gives recipients

[5] https://help.Twitch.tv/s/article/gift-subscriptions

limited-time access to the same benefits as standard subscriptions at no cost. However, unlike a free trial, recipients neither sign up to receive benefits nor do they provide payment details or commit to recurring charges after the initial benefit period ends.[6] Additionally, despite the platform label "gift," the algorithmic, impersonal nature of allocation distances this context from conventional gift-giving dynamics.[7] Throughout the paper, we therefore refer to these algorithmically-allocated subscriptions as automatic promotions that grant users temporary premium benefits.

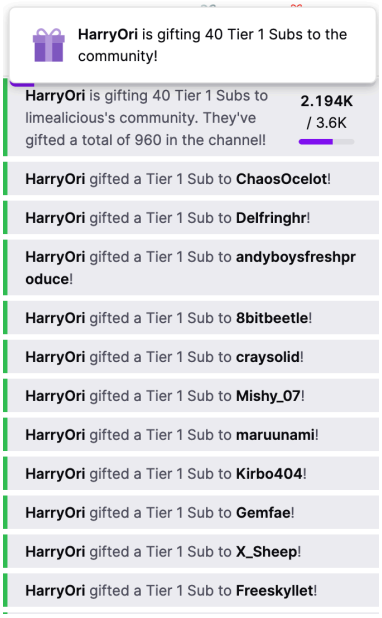
We are interested in estimating the effects of temporary premium benefits on user behavior. While the platform's algorithm generates quasi-random variation in allocation, we do not observe the algorithm's decision rules. A naïve approach might treat this context as a true experiment with uniformly random allocation, in which case the effect of receiving the benefits could be estimated by comparing the post-treatment behaviors of users who received treatment to those that did not receive it. As evidenced by the quote above, allocation correlates with user behaviors. Moreover, the algorithm may balance stated objectives, e.g., avoiding giving trolls the benefits, with unstated objectives, e.g., maximizing future profit.

This allocation algorithm has several implications for causal estimation. First, if we accurately model the allocation process, the residual variation in allocation will be as good as random. Second, if the algorithm prioritizes certain behaviors (e.g., chatting) or avoids users with particular characteristics (e.g., users that already subscribe, or users with bot-like behaviors), it will be important for us to measure behaviors and characteristics that are strongly correlated with those determinants of allocation.

[6] Einav et al. (2025) show how passive enrollment and recurring charges can influence behavior in subscription contexts via inattention and switching costs.

[7] It has been shown that intentionality plays an important role in how people interpret and respond to others' behavior (Clark, 1996). Moreover, recipients attribute less agency and intentionality to algorithmic decisions, particularly for tasks involving human skills, such as gift-giving, leading to reduced emotional reactions (Lee, 2018).

Figure 2: Algorithmically-Allocated Subscription Announcement



Notes: Example of a user purchasing multiple algorithmically-allocated subscriptions that are distributed across the channel's community.

4 Data

In this section, we discuss the data, starting with how it is collected. We then define treatment and how it informs sample construction. We discuss the measurement of behaviors that are important for determining allocation and explaining variation in the effects of receiving temporary premium benefits. Lastly, we motivate and define the outcomes we analyze.

4.1 Data Collection

We combine data from multiple sources to build a comprehensive view of user behavior on the platform. Our analysis focuses on viewership and engagement with the top 100 English-language channels. These channels cover a range of content, including gaming, e-sports, music streaming, and chatting. We collect viewership, chat engagement, and subscription information across the 100 streamers for a nine-month period spanning July 2022 to March 2023. Online Appendix A lists all creators, explains selection criteria, and details the data

collection process.

We collect user-level viewing behavior for each channel, recorded at approximately five minute intervals. These viewing records allow us to measure watch time per channel, the frequency of viewing sessions, and the number of distinct channels watched within different time frames. The analysis focuses on logged-in users, who comprise approximately 63% of channel audiences and demonstrate aggregate viewing patterns that closely mirror overall viewership (Simonov et al., 2023). We focus on logged-in users because their behaviors are observable and they represent the potential user base for subscription purchases and benefits.

We supplement viewership data with the transcripts of channel-specific chat feeds. Chat feeds record every message in a creator's channel, including user, timestamp, and message content. Chat feeds also record channel announcements, including when any user purchases a subscription for themselves or other users.

4.2 Sample

Our unit of analysis is an *eligible* user in a *cohort*. A user is *eligible* if they are observed on a channel within a six-minute window prior to an automatically-allocated subscription and are not currently subscribed to that channel.[8] A *cohort* is the set of all eligible users for a given automatically-allocated subscription purchase. Eligibility is an important criterion that restricts the sample to users who could have received temporary premium benefits. We do not compare eligible users to ineligible users—those who watched the same channel at a different time or who watched other channels—as these users may differ in systematic ways.

Within a cohort, eligible users are *treated* if they receive temporary premium benefits; all other eligible users form a pool from which we sample *control* users. We include all eligible treated users in our sample. We sample control users from the pool of eligible users at a fixed 25:1 ratio of control users to treated users. This fixed sampling ratio ensures that each

[8]The five-minute frequency of data collection ensures at least one observation per channel in the six-minute window leading up to any cohort start time. The algorithm does not award benefits to current subscribers.

treated user in a cohort is equally informative while maintaining a sufficient control pool to estimate the counterfactual behaviors of each treated user. For example, if 40 eligible users received temporary premium benefits, the sample cohort would include the 40 treated users and a random sample of 1,000 control users. Our final estimation sample is comprised of 137,482 treated users and 3,436,355 control users across 13,114 cohorts.[9]

For every eligible user in our final sample, we define a set of covariates and outcomes that are important to both the platform and creators. These covariates relate to watch time, which is one of the most prominent platform metrics, chat, which speaks to social engagement and community, subscription activity, which is a direct revenue-generating action, user retention, as well as measures relating to the composition of activity on the platform, which speak to spillovers.

4.3 Covariates

Our objective is to understand how temporary premium benefits influence multiple measures of user engagement across different time horizons. We begin by outlining pre-treatment behaviors that are likely to predict allocation and that may explain variation in how users respond to temporary premium benefits.

To capture a broad scope of engagement, we distinguish between behavior on the channel where the user received the trial of premium benefits (the *trial* channel) and across all other channels (*other* channels). The composition of engagement may explain differences in behavioral responses. For example, users that engage more with other channels than the trial channel may have greater propensity for cross-channel spillovers (Wang and Goldfarb, 2017). We define four categories of covariates that may be useful in predicting outcomes: watch, chat, future subscription purchases, and overall platform activity. We rely on behavioral covariates, as opposed to user demographics, because observed actions have greater predictive

[9]The final count of control users is 695 short of the count implied by the 25:1 treated to control ratio. In a small number of cohorts (6), such as cohorts that occur in the beginning of a streaming session when the audience is small, there are too few eligible users to exactly maintain this ratio.

power (Matz and Netzer, 2017; Lemmens et al., 2025). We include multiple daily, weekly, and monthly lags of pre-treatment behavior to capture dynamics, such as state dependence and persistence in behavior (Keane, 1997; Seetharaman et al., 1999; Dubé et al., 2010), prior experience on the platform (Ascarza, 2018), and preference for variety (McAlister and Pessemier, 1982; Kim et al., 2002; Datta et al., 2018).

For every eligible user, *watch* behavior is defined as the number of minutes the user spends connected to a particular channel during a specified time period. *Chat* behavior is defined as the number of messages the user writes in a channel's chat feed. *Subscription* behavior is defined as whether or not the user has an active paid subscription during a specified time period.[10] Additionally, we measure several *platform* behaviors: session count, the number of times a user starts watching a channel during a specified time period, and channel variety, the number of unique channels accessed by the user.

Table 1 presents summary statistics for the full list of behavior-channel-time covariates that we use to predict allocation and model treatment response. At the time of treatment, most eligible users are not currently subscribed to any channel—0% on the trial channel (by construction) and 6% across all other observed channels. A small fraction have allowed a prior subscription to lapse in the preceding 30 days—5% on the trial channel and 9% across all other channels. Aggregate watch activity is higher on other channels than on the trial channel, with users watching on average 158 minutes versus 81 minutes in the 24 hours prior to treatment. The distributions of watch and chat behaviors have large variance and are heavily right-skewed.

4.4 Outcomes

The outcomes we define closely parallel the covariates. We continue to distinguish between behavior on the trial and other channels in order to determine which behaviors are changed

[10] One-month subscriptions account for 98% of all paid-subscriptions in our sample. However, there do exist paid subscriptions that last for two or more months. Therefore, we define a user's subscription behavior in a particular period based on the time period in which the subscription was active, regardless of when the payment was made.

by temporary premium benefits, and whether the total level of activity on the platform increases or is redistributed. Our panel also allows us to observe whether behavioral changes persist after the expiration of the benefit period, and whether these effects change in direction or magnitude over time. We define multiple outcome periods, from the day in which the premium benefits were allocated (*Day 1*), to the entire benefit period (*Month 1*), and for each of the three months after the benefit period expires (*Month 2, Month 3, Month 4*).

Outcomes are defined as above and grouped into the same four categories: watch, chat, future subscription purchases, and overall platform activity. Watch behavior is an important metric for both individual content creators and the platform, directly affecting advertising revenue, creator rankings, and growth metrics. Chat behavior is another important measure of engagement for both creators and the platform. Social interactions between viewers and creators enhance the live viewing experience and can indicate engaging or high-quality content (Godes and Mayzlin, 2004). Subscriptions after the expiration of the benefit period help the platform measure the effectiveness of the promotion in converting non-paying users to paid premium users. There are many other behaviors the platform may care about that extend beyond engagement with a particular channel. We measure multiple post-treatment platform-level outcomes: channel variety, session count, and retention (whether we observe the user anywhere on the platform). Table 2 presents the full list of behavior-channel-time outcomes and their means.

Table 1: Covariate Descriptives

		Day Pre	Week Pre				Month Pre	
Behavior	Channel	1	1	2	3	4	1	2
Watch	Trial	157.53	588.27	415.13	382.48	370.01	1858.14	1495.82
		(195.83)	(776.29)	(655.73)	(635.65)	(634.09)	(2427.72)	(2432.46)
	Other	81.01	536.90	522.23	509.41	507.40	2223.61	2256.10
		(197.58)	(910.95)	(880.66)	(867.80)	(878.13)	(3374.32)	(3886.13)
Chat	Trial	1.57	5.98	4.55	4.44	4.35	20.51	17.23
		(12.84)	(48.78)	(44.01)	(48.96)	(44.33)	(171.11)	(153.68)
	Other	0.68	4.42	4.21	4.23	4.28	18.32	17.45
		(9.85)	(43.50)	(41.72)	(40.77)	(41.67)	(156.99)	(156.50)
Subscribe	Trial	0.00†	–	–	–	–	0.05	0.09
							(0.22)	(0.28)
	Other	0.06	–	–	–	–	0.09	0.08
		(0.23)					(0.28)	(0.27)
Session Count	Trial	–	–	–	–	–	23.00	18.37
							(25.12)	(24.80)
	Other	–	–	–	–	–	37.31	37.30
							(52.72)	(57.85)
Channel Variety	Platform	–	–	–	–	–	6.21	5.80
							(5.04)	(5.22)

Notes: Each cell contains the mean of one behavior-channel-time covariate, with standard deviation in parentheses. Time periods denote time intervals defined relative to the cohort start date. D1 is a one day lead, W1 is a one week lead, W2 is a two week lead (excluding the first week) with W3 and W4 defined similarly, M1 is the entire duration (30 days) of the benefit period, and M2, M3, and M4 are the non-overlapping periods two, three, and four months after the cohort start date, respectively. Trial and Other refer to behaviors on the trial channel and all other channels, respectively.

†By construction, the sample only includes users who are not subscribed to the trial channel at baseline.

Table 2: Outcome Descriptives

		Day Post	Week Post				Month Post			
Behavior	Channel	1	1	2	3	4	1	2	3	4
Watch	Trial	178.74 (208.42)	640.18 (842.79)	439.98 (733.39)	389.09 (681.93)	366.80 (666.22)	1939.70 (2669.60)	1376.47 (2403.91)	1220.31 (2306.33)	890.59 (2001.99)
	Other	81.28 (217.74)	556.41 (1198.71)	537.37 (1248.82)	529.94 (1266.62)	510.73 (1229.25)	2278.81 (4865.19)	2040.04 (4790.68)	1915.69 (4455.62)	1450.18 (3885.66)
Chat	Trial	1.58 (13.01)	6.39 (53.18)	4.90 (49.69)	4.20 (43.59)	4.05 (46.04)	20.76 (179.39)	15.99 (161.87)	14.85 (165.75)	13.65 (170.50)
	Other	0.67 (9.65)	4.54 (50.21)	4.33 (49.34)	3.91 (45.18)	3.88 (44.46)	17.82 (176.64)	16.48 (177.65)	15.50 (168.31)	14.39 (159.86)
Subscribe	Trial	-	–	–	–	–	0.11 (0.32)	0.09 (0.28)	0.07 (0.25)	0.06 (0.24)
	Other	–	–	–	–	–	0.09 (0.28)	0.08 (0.27)	0.07 (0.26)	0.07 (0.25)
Session Count	Trial	–	–	–	–	–	22.23 (25.94)	16.11 (23.15)	14.41 (22.52)	10.67 (20.03)
	Other	–	–	–	–	–	36.26 (60.67)	31.28 (57.12)	29.19 (55.80)	22.21 (49.94)
Channel Variety	Platform	–	–	–	–	–	5.96 (4.86)	5.16 (4.70)	4.80 (4.63)	3.81 (4.37)
Retention	Platform	–	–	–	–	–	1.00 (0.05)	0.92 (0.28)	0.88 (0.33)	0.74 (0.44)

Notes: Each cell contains the mean of one behavior-channel-time outcome, with standard deviation in parentheses. Time periods denote time intervals defined relative to the cohort start date. D1 is a one day lead, W1 is a one week lead, W2 is a two week lead (excluding the first week) with W3 and W4 defined similarly, M1 is the entire duration (30 days) of the benefit period, and M2, M3, and M4 are the non-overlapping periods two, three, and four months after the cohort start date, respectively. Trial and Other refer to behaviors on the trial channel and all other channels, respectively.

5 Empirical Strategy

This section presents the empirical strategy for estimating the causal treatment effects of temporary premium benefits. First, we outline the causal framework and identifying assumptions. Next, we describe the application of double-robust machine learning (DML) methods for estimating both average and heterogeneous causal effects. Lastly, we discuss estimation details.

5.1 Causal Framework

Our causal framework follows the canonical potential outcomes framework (Rubin, 1974). Let $Y_{ij}(W_i)$ denote the realization of outcome j should user i receive premium benefits ($W_i = 1$) or not ($W_i = 0$). The causal effect of the benefits on user i's behavior is given

by $Y_{ij}(1) - Y_{ij}(0)$, and the ATE of the benefits on user behavior is $\tau_j \equiv \mathbb{E}[\, Y_{ij}(1) - Y_{ij}(0)\,]$. Incorporating pre-treatment user characteristics and behaviors X_i, the CATE is $\tau_j(x) \equiv \mathbb{E}[\, Y_{ij}(1) - Y_{ij}(0) \mid X_i = x\,]$.

Average treatment effects are identified under many potential sets of restrictions on the relationship between Y_j, W, and X, the simplest case being random assignment of W, together with a stable unit treatment value assumption. A weaker identifying assumption is that of unconfoundedness: $(Y_j(0), Y_j(1)) \perp W \mid X$. That is, after controlling for observed characteristics, the assignment process does not depend on a user's potential outcomes.

The stable unit treatment value assumption requires that the potential outcomes for any one user do not vary with the assignment of premium benefits to other users, i.e., there is no interference across users. CATEs are identified under an additional "sufficient overlap" condition, $0 < P(W = 1 \mid X) < 1$, i.e., the probability of receiving benefits is never deterministic conditional on observables.

5.2 Double-Robust Machine Learning

We use a double/debiased machine learning (DML) approach to estimate causal average treatment effects. In this framework, we separately model benefit allocation and user outcomes. The allocation model, $e(x)$, describes the conditional probability that a user is allocated premium benefits given a set of observed characteristics, x:

$$e(x) = \mathbb{E}[W_i \mid X_i = x].$$

The conditional expected outcome for Y_j is given by

$$m_j(w, x) = \mathbb{E}[Y_{ij} \mid W_i = w, X_i = x].$$

We use gradient-boosted trees to estimate both $e(x)$ and $m_j(w, x)$. Although similar techniques (e.g., random forests, regression forests, neural networks) could also be used to esti-

mate these component models, we selected gradient-boosted trees for their predictive accuracy and flexibility in modeling nonlinear relationships.

To estimate heterogeneous treatment effects, we use the causal forest framework (Athey et al., 2019; Athey and Imbens, 2019). This approach extends DML to deliver consistent estimates of conditional average treatment effects.

The double-robust machine learning approach offers several important advantages. First, like all double-robust estimators, it provides robustness against model misspecification. Second, the component models are highly flexible. The gradient-boosted forests prioritize the most informative variables from a high-dimensional set of candidate covariates, estimating the functional form of the relationship between covariates, allocation, and outcomes without parametric assumptions. Third, this flexibility reduces model misspecification bias in both the allocation and outcome models.

While we do not observe the exact features used in the platform's allocation algorithm, we include a broad set of pre-treatment covariates to increase the likelihood that we account for confounding in both the allocation and outcome models. If the platform's allocation algorithm targets users based on prior watch, chat, subscription, or platform behaviors, then including these covariates helps improve the accuracy of the allocation model. At the same time, these pre-treatment behaviors may also predict post-treatment behaviors, improving the accuracy of the outcome model and richness of the conditional average treatment effects.[11]

5.3 Implementation Details

We estimate a separate outcome model for each outcome. The same allocation model is used to calculate all double-robust average treatment effects, ensuring that differences in outcome estimates reflect differences in behavior rather than allocation model estimation error. While

[11] A similar approach is used by Ellickson et al. (2023), who study the causal effects of targeted email promotions on purchase decisions, though that analysis benefited from observing the exact set of targeting variables used to assign treatment. In our case, we include as many potentially relevant covariates as possible, and rely on the method to identify the relative importance of features. For a deeper discussion of applications of double-robust machine learning methods in the marketing literature, see Lemmens et al. (2025).

cohort timing is not necessarily random, following Athey et al. (2019), all outcomes are demeaned at the cohort level to mitigate context-specific effects. This demeaning subsumes creator, time, and creator-time level effects.

Average treatment effects are calculated using the overlap method outlined in Li et al. (2018), which is preferred in cases of poor overlap (i.e., when the propensities $e(x)$ may be very close to 0 or 1). In our case, one reason for a region of the user covariate space to exhibit poor overlap is if the region corresponds to behavior inconsistent with human activity (e.g., bots that simultaneously view many channels resulting in hundreds of viewing hours per day).

It is important to note that staggered cohorts can lead to changes in the composition of treated and control users over time. In other contexts, such shifts might significantly affect the analysis (Baker et al., 2022; Borusyak et al., 2024). However, the promotion we study is rare—fewer than 1% of users receive temporary premium benefits—making these compositional changes minimal. As a result, we do not explicitly account for compositional changes in cohorts over the sample period.

6 Results

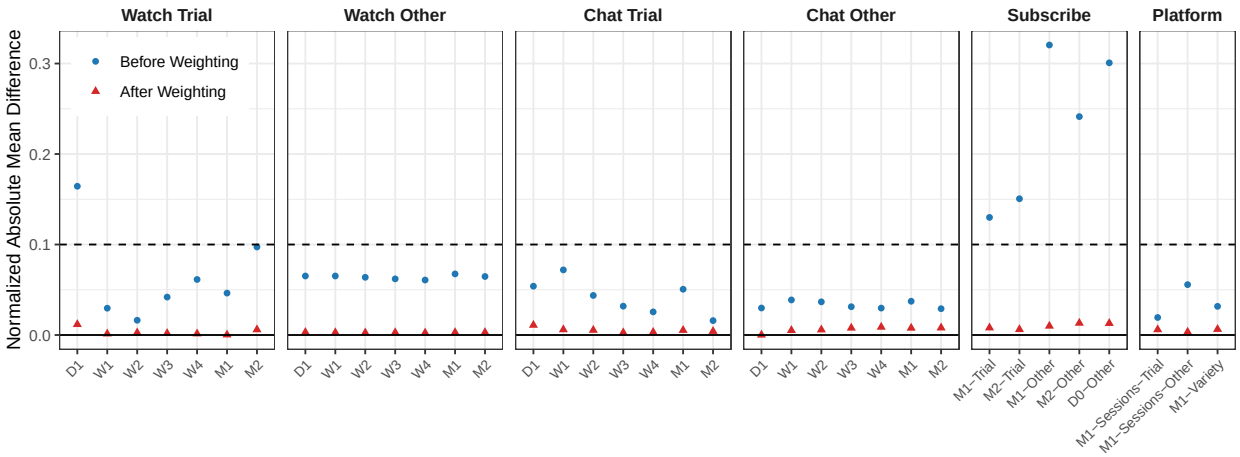
In this section, we assess covariate balance, present average treatment effects, show the robustness of our results to alternative outcome definitions and model specifications, and analyze treatment effect heterogeneity across user segments.[12]

[12]The Online Appendix contains additional estimation details and robustness checks. Online Appendix B presents variable importance statistics for assignment and outcome models, evaluates how accurately the estimated allocation model captures the platform's allocation algorithm, and demonstrates covariate balance between treatment and control groups. Online Appendix C shows the robustness of results to alternative outcome definitions.

6.1 Estimated Assignment

Figure 3 shows normalized absolute mean differences in all pre-treatment covariates between treatment and control groups before and after propensity score weighting. Before weighting, several covariates exceed the conservative 0.1 threshold for covariate balance (Austin, 2011). After applying propensity score weighting, all differences are well below this threshold, indicating that treated and control users are balanced in terms of their observed characteristics. These patterns provide supporting evidence for the the conditional independence assumption required for causal identification.

Figure 3: Covariate Balance



Notes: Normalized absolute mean differences in pre-treatment covariates between treatment and control groups with (blue circles) and without (red triangles) propensity score weighting. Time periods denote intervals relative to cohort start: D0 (baseline), D1 (one day prior), W1–W4 (weeks 1–4 prior), M1–M2 (months 1–2 prior). Trial and Other refer to behaviors on the trial channel versus all other channels, respectively. The dashed line at 0.1 indicates the conventional balance threshold.

6.2 Average Treatment Effects

Table 3 presents average treatment effects for watch, chat, and subscription outcomes across trial and other channels, and over multiple time horizons. Each watch and chat effect should be interpreted as an average change in the level of the outcome, while each subscribe outcome effect should be interpreted as a change in the probability of purchasing a subscription.

Table 3: Average Treatment Effects

		Day Post	Week Post				Month Post			
Outcome	Channel	1	1	2	3	4	1	2	3	4
Watch	Trial	1.330	22.850	23.221	23.527	21.994	97.777	46.055	27.884	15.828
		(0.496)	(2.082)	(1.912)	(1.777)	(1.752)	(6.818)	(6.384)	(5.963)	(5.298)
	Other	4.865	28.728	26.545	23.906	25.903	113.390	113.970	105.219	82.484
		(1.033)	(6.512)	(6.198)	(6.180)	(5.993)	(25.993)	(23.674)	(22.326)	(19.886)
Chat	Trial	1.106	3.048	1.608	1.500	1.145	7.530	1.472	1.172	1.406
		(0.037)	(0.116)	(0.108)	(0.108)	(0.094)	(0.342)	(0.319)	(0.295)	(0.301)
	Other	0.084	0.499	0.388	0.556	0.373	1.864	1.421	1.767	1.283
		(0.021)	(0.087)	(0.121)	(0.146)	(0.097)	(0.363)	(0.373)	(0.508)	(0.425)
Subscribe	Trial							-0.073	-0.018	0.059
								(0.001)	(0.001)	(0.001)
	Other						0.042	0.047	0.030	0.025
							(0.001)	(0.001)	(0.001)	(0.001)

Notes: Each cell shows the overlap-weighted ATE from a double-robust model with gradient boosted allocation and outcome models. The allocation model, $\hat{e}(x)$ is common across every cell; the outcome models, $\hat{m}(w, x)$, are specific to each cell. Time periods are mutually exclusive within each grouping (Day Post, Week Post, Month Post) but overlap across groupings. Standard errors clustered at the cohort level are in parentheses.

Watch

Watch time on the trial channel increases by 98 minutes during the trial month, with a 23 minute increase in the week premium benefits are received. These elevated watch levels persist throughout the month-long benefit period and continue until the third month after the benefit period concludes. Promotion recipients also increase their viewership of other channels by 113 minutes during the benefit period, indicating positive viewership spillovers across the platform. These positive spillover effects persist in each of the three months after the benefit period concludes. While the increase in watch time on the trial channel is larger during the benefit period, the increase in watch time on other channels is larger during each of the three months after the benefit period ends.

These results show that promotion recipients exhibit significant and sustained increases in viewership across the platform. Taken together, the promotion has an expansionary effect on platform watch time rather than merely shifting time from other channels to the trial channel. The persistent increases beyond the 30-day benefit period cannot be explained only by concurrent access to the functional benefits of a subscription. This suggests that

temporary premium benefits influence some users' content preferences and affinity towards the platform.

Chat

Users who receive the promotion also show significant increases in the propensity to use the platform's chat feature during and after the benefit period. Average chat volume during the benefit period increases by 7.5 messages on the trial channel and 1.9 messages across all other channels. These increases are large as only 50% of users chat in the month prior to treatment on any of the observed channels. While the large initial increase on the trial channel is, in part, explained by promotion recipients writing "thanks" or "thank you" in chat to the user who gifted the subscription, chat levels remain elevated across the trial and other channels for the three months after the benefit period concludes. These effects suggest that the promotion's impact extends well beyond these initial reactions, indicating lasting shifts in user behavior and deeper platform engagement.

Paid Subscriptions

Temporary premium benefits have large effects on subsequent subscription behavior that vary by channel and time horizon. Benefit recipients are initially 7.3 percentage points less likely to be subscribed to the trial channel in the month immediately following the benefit period, but this negative effect reverses to a positive 5.9 percentage point increase by the third month after the benefit period ends. Conversely, subscription rates for other channels increase across all time horizons. Promotion recipients are 4.2 percentage points more likely to be subscribed to other channels during the benefit period. Treated users are also 4.7 percentage points more likely to subscribe to other channels in the month after the benefit period ends, with that pattern continuing for months three and four. These effects are substantial, corresponding to a 40-62% lift during the four months after the promotion was received. Despite the temporary reduction in propensity to subscribe to the trial channel,

the net platform-wide effects on paid subscriptions overall, and during each of the three post-benefit months, are positive.

The initial reduction in trial channel subscriptions may reflect several behavioral mechanisms. Users may engage in intertemporal substitution, delaying purchases after receiving free access to the same benefits. Alternatively, content satiation may lead users to explore other creator content after intensive exposure to the trial channel. Another possible explanation is that receiving premium benefits delays the next opportunity to subscribe for treated users—they cannot purchase a new subscription until the start of month two at the earliest. In contrast, control users can purchase subscriptions throughout the benefit period. This creates a reduction in observed treatment group subscriptions in month two. The impact of this mechanism diminishes in months three and four as both groups have more equal opportunity to subscribe.

Platform Outcomes

Table 4: Additional Platform Outcomes

Outcome	Month 1	Month 2	Month 3	Month 4
Channel Variety	0.089	0.075	0.059	0.060
	(0.010)	(0.010)	(0.010)	(0.010)
Trial Session Count	1.416	0.670	0.461	0.288
	(0.064)	(0.062)	(0.060)	(0.058)
Other Session Count	1.650	1.775	1.481	1.056
	(0.270)	(0.247)	(0.254)	(0.233)
	(0.001)	(0.001)	(0.001)	(0.001)
Retention	0.000	0.009	0.009	0.012
	(0.000)	(0.001)	(0.001)	(0.001)

Notes: Each cell shows the overlap-weighted ATE from a double-robust model with gradient boosted allocation and outcome models. The allocation model, $\hat{e}(x)$ is common across every cell; the outcome models, $\hat{m}(w, x)$, are specific to each cell. Standard errors clustered at the cohort level are in parentheses.

Temporary premium benefits generate small positive and lasting changes in platform-level engagement. Benefit recipients watch, on average, 0.1 more channels (1.5% increase), tune in

for 1.4 additional sessions on the trial channel (6.4% increase), and tune in for 1.7 additional sessions on the other channels (4.6% increase) The promotion increases the probability of retention—any subsequent activity on the platform—by 0.1 percentage points in Months 2, 3 and 4. This corresponds to a decrease in churn between 5-11% in each of the three months following the benefit period. While these effects are generally small, we expect them to be lower bounds as we do not observe activity on every channel on the platform.

Overall, temporary premium benefits generate significant platform-wide expansionary effects. Users who receive benefits watch more, chat more, and are more likely to be subscribed in the long run to both the trial channel and other channels on the platform. These results highlight the importance for platforms to measure and understand the comprehensive effects of their marketing interventions, as relevant effects may not be immediately evident and may extend beyond the immediate context of the treatment.

6.3 Robustness

We next show the robustness of our average treatment effect estimates to alternative outcome definitions and model specifications.

In our main specification, we present the watch and chat effects in levels (minutes and counts). However, log-transformed outcomes may be desirable in this context for several reasons. First, heavy users can disproportionately influence the average treatment effects in levels, whereas the log transformation reduces the impact of outliers. This point is particularly important in this context given the heavily right-skewed distribution of consumption on the platform. Second, the platform may not equally value the same absolute increase for light and heavy users. For example, Gentzkow et al. (2024) shows that the CPM (cost per thousand impressions) for heavy TV viewers is lower than that of lighter viewers, reflecting their reduced value to advertisers.

In Table 5, we show the results with log-transformed outcomes. With logged outcomes, the effects have the same sign, are slightly larger in magnitude, and are more precisely

Table 5: Average Treatment Effects with Logged Outcomes

		Day Post	Week Post				Month Post			
Outcome	Channel	1	1	2	3	4	1	2	3	4
log(1+Watch)	Trial	0.050	0.098	0.180	0.175	0.166	0.117	0.125	0.093	0.087
		(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.004)	(0.008)	(0.008)	(0.008)
	Other	0.013	0.040	0.046	0.040	0.040	0.071	0.065	0.055	0.058
		(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
log(1+Chat)	Trial	0.199	0.249	0.114	0.091	0.082	0.289	0.063	0.046	0.048
		(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
	Other	0.011	0.028	0.023	0.022	0.022	0.056	0.038	0.032	0.030
		(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Subscribe	Trial							-0.073	-0.018	0.059
								(0.001)	(0.001)	(0.001)
	Other						0.042	0.047	0.030	0.025
							(0.001)	(0.001)	(0.001)	(0.001)

Notes: Each cell shows the overlap-weighted ATE from a double-robust model with gradient boosted allocation and outcome models. The allocation model, $\hat{e}(x)$ is common across every cell; the outcome models, $\hat{m}(w, x)$, are specific to each cell. Time periods are mutually exclusive within each grouping (Day Post, Week Post, Month Post) but overlap across groupings. Standard errors clustered at the cohort level are in parentheses.

estimated. We find that watch time of the trial channel increases by 12.4% ($\exp(\hat{\beta}) - 1$) during the benefit period. As before, this increase is persistent: we find a statistically significant 9.1% increase three months after the benefit period ends.

These estimated effects should be interpreted with caution as they reflect a combination of extensive and intensive margin changes (Chen and Roth, 2024). In our context, this is distinction is important because the typical user does not engage in chat behavior. Therefore, the 33.4% increase in chat behavior on the trial channel during the benefit period is at least partially driven by changes in the extensive margin. In contrast, every eligible user in our sample watches content on the platform. As such, the watch estimates primarily reflect changes in the intensive margin of behavior.

Next, we show how our estimates vary by re-estimating average treatment effects with six alternative model specifications. Instead of re-estimating all outcomes, we focus on changes in behavior during the benefit period (Month 1 in Table 3). Specification 1 estimates each treatment effect as a simple unweighted difference in means between treated and control users. Specification 2 is a linear outcome model comprised of a treatment indicator and the full set of user-level pre-treatment covariates. Specification 3 is a logit assignment model

with the full set of user-level pre-treatment covariates. Specification 4 is a double robust regression model, using the logit propensities of Specification 3 as assignment weights together with the linear outcome model of Specification 2. Specifications 5 and 6 estimate average treatment effects using gradient-boosted outcome-only and assignment-only models, respectively. Specification 7 presents our preferred doubly-robust estimates with gradient-boosted component models.

Table 6: Treatment Effect Estimates Across Models

			Regression Models			Gradient Boosted Models		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Behavior	Channel	Mean	$\hat{e}$	$\hat{m}$	$\hat{e}, \hat{m}$	$\hat{e}$	$\hat{m}$	$\hat{e}, \hat{m}$
Watch	Trial	33.126	140.718	100.386	96.655	54.186	101.343	97.777
		(9.914)	(9.735)	(6.945)	(8.772)	(10.450)	(12.082)	(7.930)
	Other	36.174	-53.107	6.509	-4.390	-71.816	29.545	113.390
		(15.807)	(21.187)	(16.035)	(20.624)	(34.675)	(37.003)	(27.770)
Chat	Trial	14.671	11.190	7.202	5.614	4.123	5.776	7.530
		(0.667)	(0.824)	(0.608)	(1.071)	(0.518)	(0.537)	(0.385)
	Other	14.357	5.132	1.522	1.522	-0.288	2.023	1.864
		(0.839)	(0.728)	(0.586)	(0.630)	(0.531)	(0.497)	(0.439)
Subscribe	Other	0.114	0.046	0.047	0.047	0.043	0.048	0.042
		(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: Standard errors in parentheses are calculated from a block bootstrap of cohorts with 200 iterations, maintaining the ratio of treated and control users within each cohort. Mean: simple difference in means. $\hat{e}$: inverse propensity score weighting, with propensity scores estimated via logistic regression or gradient boosting. $\hat{m}$: outcome regression estimated via OLS or a T-learner with separate gradient boosted models for treated and control groups. $\hat{e}, \hat{m}$: doubly robust estimators combining both propensity score weighting and outcome regression. In each regression model, we include both the level and quadratic term for increased flexibility.

The estimates are shown in Table 6.[13] Although the effects are directionally consistent with our preferred specification, the difference in means specification unsurprisingly yields markedly different estimates. We do not expect this specification to return unbiased or consistent estimates since treatment is not randomly allocated. Moreover, treatment is strongly positively correlated with particular outcomes, such as chat activity and watch levels, which

[13]The gradient-boosted outcome-only model is implemented as a T-Learner, and does not have a closed form solution for standard errors. For the sake of comparability, we bootstrap standard errors for all models in Table 6, including our preferred specification. This results in small differences between the standard errors here and those reported in Table 3.

explains why we see attenuated effects in the robust specifications.[14] The two double-robust specifications, columns 4 and 7, return similar estimates with the exception of the Watch Other outcome, which is not statistically distinguishable from zero in the double-robust linear regression model. Watch time on other channels, which can theoretically range from zero to simultaneous consumption of all channels during the entire month, has the greatest potential for extreme outliers.[15] Consistent with large outliers, the logged-outcome results in Table 5 exhibit much smaller standard errors than our main specification, which uses levels. While the double-robust models both provide protection against model misspecification and confounding in treatment assignment, we believe our preferred specification provides the best estimates given its flexibility. Additionally, the gradient-boosted specification lends itself to heterogeneity analysis.

6.4 Heterogeneous Effects

Identifying how different users respond to treatment is a necessary step in improving the efficacy of targeted marketing efforts. Our heterogeneity analysis focuses on behavior over the three months following the end of the benefit period. We focus on three binary outcome variables that are relevant to platform decision-making: trial channel subscriptions, other channel subscriptions, and user retention. Paid subscriptions to the trial channel measure the effectiveness of promotional benefits in converting free to premium users. Paid subscriptions to other channels capture revenue-generating spillovers. User retention speaks to long-term preferences for engaging with content on this platform instead of alternatives. Each binary outcome has a simple interpretation as a change in probability.

To define user segments, we focus on pre-treatment characteristics that are predictive of and have a clear theoretical connection to the outcomes of interest. We find that watch behavior on the trial channel and other channels in the month prior to treatment have high variable importance for predicting subscription and retention outcomes. This is consistent

[14] See Online Appendix B for more details on variable importance in the assignment and outcome models.
[15] See Online Appendix C for further discussion of the role of outliers and additional outcome specifications.

with prior research showing that past usage patterns are strong predictors of future customer behavior and lifetime value (Fader et al., 2005a,b; Fader and Hardie, 2009; Lambrecht and Misra, 2017; Ascarza, 2018).

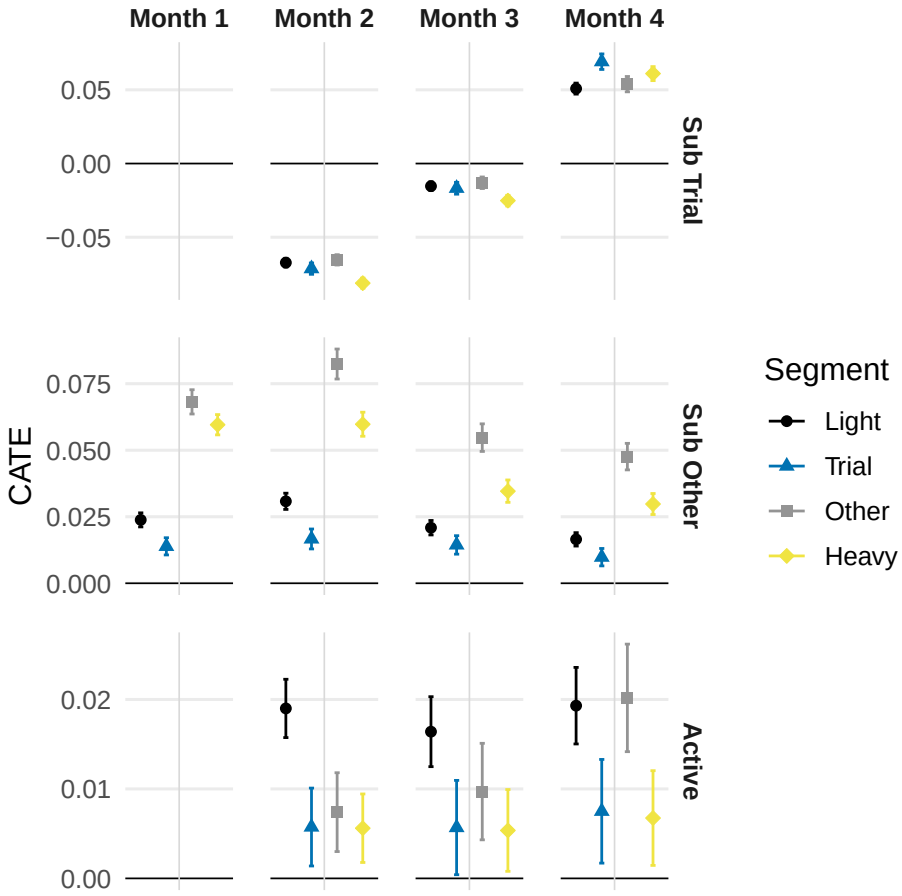
To estimate user-level CATEs, we use causal forests (Athey and Imbens, 2019; Athey et al., 2019) that build on our previous assignment and outcome models. After estimating the heterogeneous effects, we use a doubly-robust linear projection of the user-level effects onto median splits of our two pre-treatment covariates. This allows us to succinctly characterize how treatment response varies across user segments. For ease of interpretability, we focus this analysis along these two dimensions. In Section 7, we take a more agnostic, granular, and data-driven approach to targeting based on the full set of observable user characteristics.

Figure 4 presents heterogeneous treatment effects for each outcome and time period, across four user segments: *light watchers* (users that are below the median watch level on the trial and other channels), *trial watchers* (users that are above the median on the trial but below the median on other channels), *other watchers* (users that are below the median on the trial but above the median on other channels, and *heavy watchers* (users that are above the median on the trial and other channels).

Time period is the dominant factor in determining effect size and direction for subscriptions to the trial channel. In the short run (Months 2 and 3), users with below-median trial channel watch levels (*light watchers* and *other watchers*) exhibit larger (i.e., less negative) treatment effects. In the longer run (Month 4) this pattern reverses, with *heavy watchers* and *trial watchers* being most likely to purchase paid subscriptions to the trial channel.

The promotion has positive cross-channel effects for each of the four user segments, increasing the probability of a paid subscription to a non-trial channel. These positive spillovers are larger for users with watch levels above the median on other channels, and largest for *other watchers*. These results suggest that a user who experiences subscription benefits outside of their preferred context may be persuaded to pay for a subscription in a context that more closely aligns with their preferences.

Figure 4: Heterogeneous Treatment Effects



Notes: This figure shows heterogeneous treatment effects and 95% confidence intervals for subscription and retention outcomes. Users are segmented based on pre-treatment watch time on the trial channel and across all other channels. Standard errors clustered at the cohort level.

In the month following the benefit period, the net effect on subscriptions is positive for the *other watchers* segment. While we cannot reject zero net effect for the *heavy watchers* segment, the remaining two segments have a negative net effect until Month 3. By Month 4, all segments have positive net subscription effects. We note that both the other channel subscription and net subscription effects are likely lower bounds. First, we do not observe all other channels on the platform. Second, the binary outcome definition does not account for contemporaneous subscriptions to multiple channels.

Premium benefits have a positive effect on retention for each segment in each time period.

Retention effects are largest among *light watchers*, who also make up the segment with the highest baseline probability of exiting the platform. For each of the four segments, retention effects are largest in the longer term (Month 4), when the baseline probability of churn is higher.

These heterogeneity results reveal tensions between targeting to achieve different outcomes. That is, the strongest long-run treatment effects for the three outcomes are for distinct segments—trial channel subscription effects are strongest among the *trial watchers*; other channel subscription effects are the strongest for the *other watchers*; and retention effects are strongest among the *light watchers*. These tensions highlight the need for platforms to formalize the trade-offs inherent in targeting for different objectives.

7 Multi-Objective Targeting

As shown in the previous section, the effects of premium benefits vary across user segments. Moreover, the users that are most responsive on one dimension of interest may exhibit less favorable responses on other dimensions. In this section, we formalize these trade-offs using a multi-objective optimization framework (Rafieian et al., 2024). We continue to focus on three long-run outcomes—paid subscriptions to the trial channel, paid subscriptions to other channels, and user retention—which together reflect the platform's ability to monetize and sustain user engagement. Using causal forest estimates, we identify Pareto efficient targeting policies—policies where improving one outcome necessarily worsens another—and evaluate the performance of different policies along the Pareto frontier. This framework allows us to answer three questions: (1) How do optimal targeting policies compare to the platform's current targeting algorithm? (2) What are the opportunity costs of targeting to optimize a single outcome? (3) How sensitive are Pareto-optimal policies to shifts in the platform's relative value of each outcome?

7.1 Problem Definition

Let Y_1, Y_2, and Y_3 denote the three outcomes of interest: trial channel subscriptions, other channel subscriptions, and user retention. A targeting *policy*, denoted as $\pi : X \to [0, 1]$, is a mapping from covariates X to the probability of assignment. We focus on deterministic policies, where a user with covariates x is either *targeted*, $\pi(x) = 1$, or *not targeted*, $\pi(x) = 0$.[16]

For each outcome j, we evaluate the performance of a policy π using the average treatment effect on the targeted (ATT): the expected conditional average treatment effect, $\tau_j(x)$, among users for whom $\pi(x) = 1$:

$$\rho_j(\pi) = \mathbb{E}_X \left[\tau_j(X) \mid \pi(X) = 1\right]$$

A policy π is *Pareto optimal* if no alternative policy π' can improve at least one outcome's ATT without reducing another. Our goal is to identify the set of Pareto optimal policies, which forms the *Pareto frontier*.

We trace out the Pareto frontier using a linear scalarization algorithm. The algorithm optimizes a weighted sum of performance measures, $\sum_{j=1}^{3} \beta_j \rho_j(\pi)$, where the weights $\{\beta_1, \beta_2, \beta_3\}$ are non-negative and sum to 1. The policy that optimizes the combined objective $\sum_{j=1}^{3} \beta_j \rho_j(\pi)$ targets the set of users with the largest weighted sum of treatment effects. By re-estimating the targeting policy for different weight vectors in the simplex, we obtain a set of Pareto optimal policies. The union of these policies approximates the Pareto frontier, which illustrates trade-offs between objectives and provides a view of potential outcomes in the policy space.[17]

If targeting is costless and not scarce (i.e., the platform can give out as many promotions as it wants at zero cost), the optimal policy for a given value of β, $\pi_\beta(x)$, targets an eligible

[16] We use *targeted* and *not targeted* to differentiate between users who receive premium benefits under a hypothetical policy, and continue to use *treated* and *control* to refer to users who receive premium benefits under the platform's algorithm.

[17] See Rafieian et al. (2024) for details on the properties of the Scalarization with Causal Effects Algorithm.

user with characteristics x if the weighted sum of the user's CATEs, $\sum_j \tau_j(x)\beta_j$, is greater than or equal to zero. If we instead constrain the number of targeted users to Z, the optimal policy targets the Z eligible users with the highest weighted sum of CATEs. Formally, the corresponding optimal policy is defined as:

$$\pi_\beta(x) = \begin{cases} 1 & \text{if } \sum_{j=1}^{3} \beta_j \tau_j(x) \geq \eta^{(Z)}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\eta^{(Z)}$ is the Z-th largest value of $\sum_{j=1}^{3} \beta_j \tau_j(x)$ across eligible users. The scalarization algorithm can be adapted to constrain the number of targeted users at different levels. For example, in our application the number of targeted users can be constrained to match (or be a multiple of) the total number of treated users in the data or the number of treated users within each cohort.

7.2 Estimation

Our search for Pareto optimal policies is based on user-level CATEs. While our causal forests return user-level CATE estimates, and are trained using honest sampling to mitigate overfitting, using these CATE estimates for both policy definition and evaluation introduces bias. Specifically, targeting the Z eligible users with the highest-weighted sum of CATEs will result in upward-biased policy performance estimates (Xu et al., 2025). This bias arises because estimation errors in the CATEs are correlated with the selection rule: users selected for targeting are more likely to have positive estimation errors, which results in overoptimistic policy evaluation among the targeted users.

To address this issue, we first split the full sample by cohort into three subsamples. On the first subsample, we estimate a causal forest model that is used in defining the optimal policy (i.e., to determine which users are targeted). On the second, independent, subsample, we estimate a second causal forest model that is used only for performance evaluation. We then

apply both models to the third subsample: we use the first model's CATE estimates to define the targeting policy, and the second model's CATE estimates to evaluate its performance. This procedure decouples the selection of targeted users from evaluation, yielding more conservative and less biased performance estimates. This three-way split is more conservative than a standard two-way split because policy performance is evaluated using model-based CATE estimates rather than directly observed outcomes. Thus, it ensures both the policy selection model and the evaluation model are applied out-of-sample (Athey and Wager, 2021).

To ensure balanced targeting across cohorts and reduce the sensitivity of performance measure estimates to outliers, we adopt a cohort-level targeting constraint: within each cohort, we constrain the total number of targeted users to be five times the number of observed treated users in that cohort.

We estimate the Pareto frontier by performing a grid search over β.

7.3 User-Level Heterogeneity and Example Policies

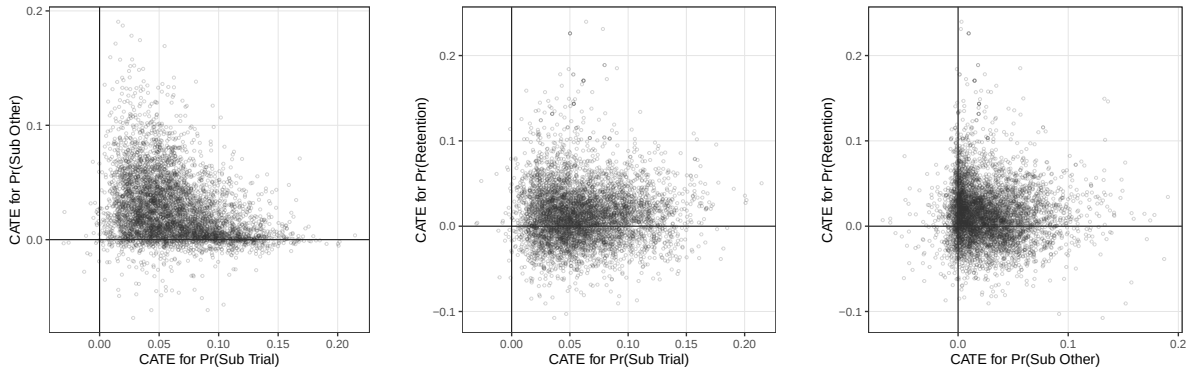
Before discussing optimal policies, we first examine the distribution of estimated CATEs and illustrate a few example policies.

Figure 5 presents pairwise comparisons of estimated user-level CATEs, $\hat{\tau}_j(x)$, for our three outcomes. Each distribution exhibits substantial heterogeneity. There is a negative relationship between large trial-channel and other-channel subscription effects. We also see that users who exhibit large retention effects tend to have modest subscription effects.

Figure 6 illustrates three example Pareto optimal policies plotted with respect one pair of effects—trial and other subscriptions. Panel (a) illustrates a single-objective policy that places all weight on trial channel subscriptions ($\beta_1 = 1, \beta_2 = 0, \beta_3 = 0$). This policy identifies users most responsive to trial channel targeting, but is agnostic to users with high other-channel subscription effects. Note that the boundary between targeted and non-targeted users is not perfectly delineated because of the cohort-level targeting constraint; if targeting

Figure 5: Pairwise Comparisons of Individual-Level CATEs

(a) Sub Other vs Sub Trial (b) Retention vs Sub Trial (c) Retention vs Sub Other



Notes: Pairwise comparisons of individual-level CATEs for subscriptions to trial channels, subscriptions to other channels, and retention. All outcomes are measured four months post-trial. Points are a random sample of 5,000 users (for graphical clarity).

were unconstrained, the boundary would be defined by a vertical line at a threshold level of trial subscription effects.

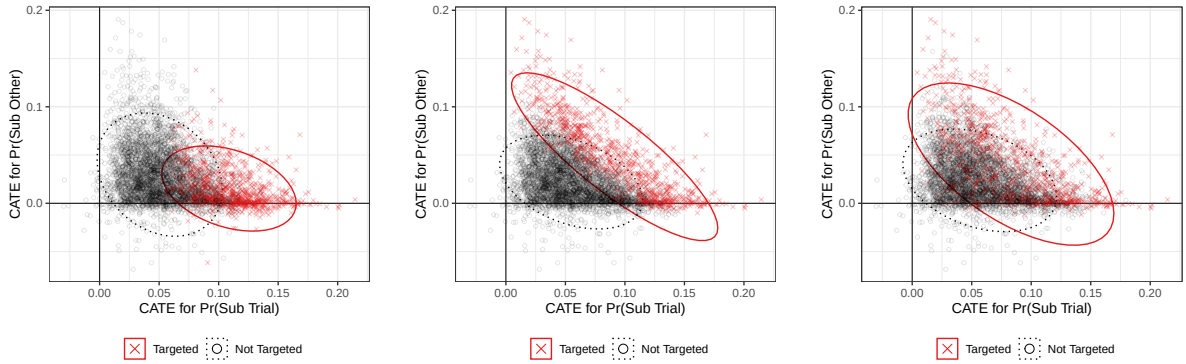
Panel (b) shows a multi-objective placing equal weights on trial and other channel subscriptions ($\beta_1 = 1/2, \beta_2 = 1/2, \beta_3 = 0$). Targeted users are now concentrated in the upper-right region of the joint distribution. The targeted set includes users with strong effects on either subscription dimension, as well as users that have moderately strong effects on both dimensions. Given the negative correlation between user-level CATEs within the targeted region, 55% of users targeted in policy (a) are also targeted in policy (b).

Panel (c) attaches equal weights to all three outcomes ($\beta_1 = \beta_2 = \beta_3 = 1/3$). In this example, the boundary between targeted and non-targeted users appears less distinct than in the previous two examples. This is because retention, the third outcome, is weighted in the targeting decision but is not visualized in the two-dimensional plot. The policies depicted in (a) and (c) target distinct user groups, with 43% of users who are targeted by policy (a) also being targeted by policy (c).

In this empirical context, the substantial heterogeneity in user-level response to treatment within each individual dimension creates scope for targeted policies to outperform uniform

Figure 6: Example Policies

(a) $\beta = (1, 0, 0)$ (b) $\beta = (1/2, 1/2, 0)$ (c) $\beta = (1/3, 1/3, 1/3)$



Notes: Panels display individual-level CATEs for trial channel subscriptions (x-axis) and other channel subscriptions (y-axis), with each targeted user marked with a red x. Panel (a) shows a single-objective policy maximizing trial channel subscriptions. Panel (b) shows a multi-objective policy weighting trial and other channel subscriptions equally. Panel (c) shows a policy giving equal weight to all three outcomes. Each policy targets the same total number of users (five times the observed treatment count within each batch). Points shown are a random sample of 5,000 users (for graphical clarity). Ellipses are drawn to contain 90% of users in each group to help distinguish targeted from not targeted users.

allocation in terms of the ATT. Moreover, the covariation in treatment responses between outcomes—in particular, the fact that responses are not strongly positively correlated—highlights the importance of explicitly accounting for all important outcomes in a multi-objective targeting framework.

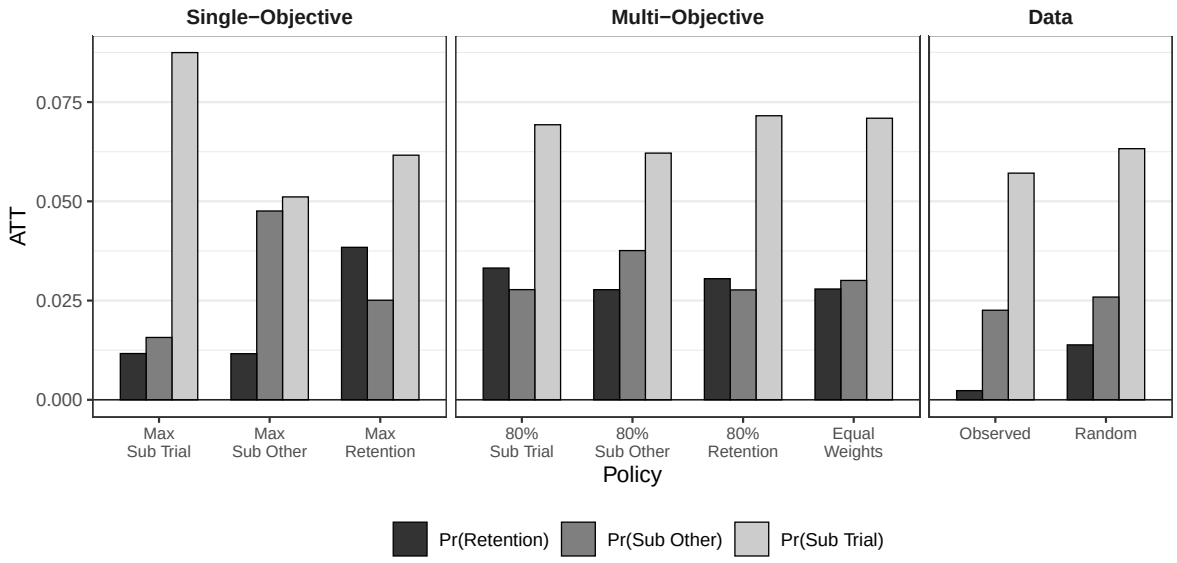
7.4 Policy Performance

We now quantify the performance of Pareto frontier policies and compare them to a set of non-Pareto optimal benchmark policies. Figure 7 summarizes ATTs for policies in three groups: single-objective policies, sample multi-objective policies, and benchmark policies (observed and random allocation).[18] Specifically, we look at the three multi objective policies on the Pareto that achieve 80% of the optimal single objective performance for one measure, while maximizing the combined performance of the other two measures. The results in this section align closely with those presented earlier, but they differ in two important ways.

[18] In Online Appendix B, we visualize the Pareto frontier and provide the numerical estimates visualized in Figure 7.

First, the main results section emphasized population-level effects, reporting ATEs estimated with overlap weighting. In this section, we instead evaluate targeting policies derived from individual-level CATEs. Because our goal is policy evaluation rather than population-level inference, the ATT estimates are computed directly on the holdout sample without applying overlap weighting.

Figure 7: Policy Performance



Notes: Bars display average treatment effects on the targeted (ATTs) for selected policies. Single-objective policies maximize one outcome. Multi-objective policies achieve 80% of the maximum ATT for one outcome while optimizing the sum of the other two, or give equal weight to all three. Observed reflects the platform's actual allocation. Random simulates uniform random assignment.

The ATTs of the observed allocation are similar to random assignment and markedly different from any of the estimated Pareto optimal policies. This finding is consistent with the platform's stated objective of "identify[ing] members of a community", and provides further evidence that the allocation algorithm is not designed to optimize the subscription or retention outcomes we study here.

Policies that optimize single outcomes achieve higher ATTs on their target dimensions, but forego potential improvements on other dimensions. For example, moving from the Equal Weights policy to a policy that maximizes subscriptions on the trial channel (Max Sub Trial) increases trial subscriptions by 1.7 percentage points (23% increase over Equal

Weights policy), but decreases other channel subscriptions by 1.4 points (48% decrease) and retention by 1.6 points (58% decrease). Similarly, shifting from Equal Weights to Max Retention increases retention by 1.1 points (38% increase), but reduces trial subscriptions by 0.9 points (13% decrease) and other subscriptions by 0.5 points (17% decrease). These trade-offs highlight the high opportunity cost of optimizing one objective in isolation.

In contrast, policies that balance objectives, such as Equal Weights, perform well across all outcomes simultaneously. The Equal Weights policy achieves an ATT of 0.071 on trial subscriptions, 0.030 on other subscriptions, and 0.028 on retention—representing 63–81% of each single-objective's maximum, and avoiding the foregone benefits that come from optimizing a single outcome. The Pareto frontier is also relatively flat near the Equal Weights policy: comparing the three 80% policies, improvements in outcomes are within ±1.0 percentage points of those from Equal Weights. Putting these targeting effects in context, a platform-wide shift from the observed allocation to the equal weights policy would result in 213 more subscriptions and 256 additional retained users per 10,000 targeted users in the fourth month following treatment. The total effect of the policy, which would include the first three months after targeting and the months after month four, is likely much larger.

8 Conclusion

In this paper, we study the effects of temporary premium benefits on user behavior. Our analysis centers on a multi-channel social live-streaming platform where users can purchase premium subscriptions to specific creator channels. We leverage quasi-exogenous variation in the promotional allocation of premium benefits together with a flexible double-robust estimation strategy to measure causal effects across multiple outcomes and time horizons. The promotion we study not only affects future subscription purchases, but also a wide variety of platform-relevant behaviors, including watch activity and social engagement. These effects vary meaningfully across time and for different user segments, and spill over to other chan-

nels on the platform. Users who receive benefits on a channel they watch less are more likely to respond by subscribing to other channels that they watch more regularly. We quantify the trade-offs in designing targeting policies to optimize different objectives and show that single-outcome targeting carries high opportunity costs due to the potential for promotions to influence a broad set of platform-relevant outcomes.

Our results have several implications for platforms and marketing managers. First, firms benefit from taking a holistic view when evaluating marketing interventions. A narrow focus on individual metrics or short-term outcome windows may misrepresent time-varying treatment effects and obscure trade-offs between objectives. Second, platforms can exploit cross-channel spillovers by targeting users outside of their preferred contexts. This result applies specifically to firms that sell multiple products or services. Although free access to premium benefits may cannibalize purchases in the short-term, multi-service firms can offset these negative effects by internalizing positive spillovers across their portfolio. For example, a single-product firm (e.g., Netflix) cannot capture cross-service spillovers from promotions like free trials. In contrast, a multi-product firm (e.g., Disney, which offers multiple streaming services and other entertainment products) stands to benefit when a promotion on one service boosts purchases or engagement on another. Importantly, a multi-service firm must analyze user data across its entire portfolio, as evaluating each service in isolation would miss effects that determine overall profitability.

Our analysis leaves several questions for future research. We document strong cross-channel spillovers but do not identify the mechanisms behind them. Spillovers could be driven by many different mechanisms, including from substitution, complementarity, or quality differences across channels. Understanding which of these mechanisms play a role would inform promotion strategies in other contexts. If spillovers are driven by substitution, platforms should target adjacent content that builds interest without giving away what users would otherwise pay for—for example, offering trial access to comedy content to users that prefer drama content to build interest in a full subscription. If complementarity drives

spillovers, a social media platform could use its recommendation algorithms to position ads next to complementary content—for example, sequencing an all-wheel drive automobile ad after outdoor-lifestyle content might enhance the appeal or attention to the ad. Distinguishing between these mechanisms would require experimental manipulation of the degree of complementarity between focal and adjacent contexts—variation our observational data do not provide. Relatedly, while our reduced-form approach measures changes in behavior that are suggestive of preferences, we do not identify the underlying preference structure that maps time allocation to willingness-to-pay.

Finally, implementing optimal targeting in practice requires firms to balance experimentation with exploitation. Continued experimentation provides the variation needed to learn about preferences and treatment effects, while exploitation increases short-term performance using current knowledge. How platforms can dynamically manage this trade-off—particularly as user preferences and content offerings evolve—remains an important operational challenge.

References

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

Athey, Susan and Guido W. Imbens (2019). "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics*, 11(1): 685–725.

Athey, Susan, Julie Tibshirani and Stefan Wager (2019). "Generalized Random Forests." *The Annals of Statistics*, 47(2): 1148 – 1178.

Athey, Susan and Stefan Wager (2021). "Policy Learning with Observational Data." *Econometrica*, 89(1): 133–161.

Austin, Peter C (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research*, 46(3): 399–424.

Baker, Andrew C, David F Larcker and Charles CY Wang (2022). "How Much Should We Trust Staggered Difference-in-Differences Estimates?" *Journal of Financial Economics*, 144(2): 370–395.

Bapna, Ramnath and Akhmed Umyarov (2015). "Do your online friends make you pay? A Randomized Field Experiment on Peer Influence in Online Social Networks." *Management Science*, 61(8): 1902–1920.

Bawa, Kapil and Robert Shoemaker (2004). "The Effects of Free Sample Promotions on Incremental Brand Sales." *Marketing Science*, 23(3): 345–363.

Borusyak, Kirill, Xavier Jaravel and Jann Spiess (2024). "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies*: rdae007.

Chae, Boyoun, Andrew T Stephen, Yakov Bart and David A Yao (2017). "Spillover Effects in Seeded Word-of-Mouth Marketing Campaigns." *Journal of Marketing Research*, 54(1): 143–162.

Chen, Jiafeng and Jonathan Roth (2024). "Logs with Zeros? Some Problems and Solutions." *The Quarterly Journal of Economics*, 139(2): 891–936.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey and James Robins (2018). "Double/Debiased Machine Learning for Treatment and Structural Parameters."

Clark, Herbert H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Datta, Hannes, Bram Foubert and Harald J Van Heerde (2015). "The Challenge of Retaining Customers Acquired with Free Trials." *Journal of Marketing Research*, 51(2): 217–234.

Datta, Hannes, George Knox and Bart J Bronnenberg (2018). "Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery." *Marketing Science*, 37(1): 5–21.

Dubé, Jean-Pierre, Günter J. Hitsch and Peter E. Rossi (2010). "State Dependence and Alternative Explanations for Consumer Inertia." *RAND Journal of Economics*, 41(3): 417–445.

Einav, Liran, Ben Klopack and Neale Mahoney (2025). "Selling Subscriptions." *American Economic Review*, 115(5): 1650–71.

Ellickson, Paul B., Wreetabrata Kar and James C. Reeder (2023). "Estimating Marketing Component Effects: Double Machine Learning from Targeted Digital Promotions." *Marketing Science*, 42(4): 704–728.

Fader, Peter S and Bruce GS Hardie (2009). "Probability Models for Customer-Base Analysis." *Journal of Interactive Marketing*, 23(1): 61–69.

Fader, Peter S, Bruce GS Hardie and Ka Lok Lee (2005a). ""Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model." *Marketing Science*, 24(2): 275–284.

Fader, Peter S, Bruce GS Hardie and Ka Lok Lee (2005b). "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis." *Journal of Marketing Research*, 42(4): 415–430.

Förderer, Jann, Dominik Gutt and Brad N Greenwood (2023). "Star Power and Content Creator Ecosystems: Evidence from Twitch." *Information Systems Research*, 34(2): 675–693.

Foubert, Bram and Els Gijsbrechts (2016). "Try It, You'll Like It—or Will You? The Perils of Early Free-Trial Promotions for High-Tech Service Adoption." *Marketing Science*, 35(5): 810–826.

Gentzkow, Matthew, Jesse M Shapiro, Frank Yang and Ali Yurukoglu (2024). "Pricing Power in Advertising Markets: Theory and Evidence." *American Economic Review*, 114(2): 500–533.

Godes, David and Dina Mayzlin (2004). "Using Online Conversations to Study Word-of-Mouth Communication." *Marketing Science*, 23(4): 545–560.

Goldfarb, Avi, Catherine Tucker and Yanwen Wang (2022). "Conducting Research in Marketing with Quasi-Experiments." *Journal of Marketing*, 86(3): 1–20.

Huang, Yufeng and Ilya Morozov (2025). "The Promotional Effects of Live Streams by Twitch Influencers." *Marketing Science*.

Iyengar, Raghuram, Young-Hoon Park and Qi Yu (2022). "The Impact of Subscription Programs on Customer Purchases." *Journal of Marketing Research*, 59(6): 1101–1119.

Kan, Michael (2020). "Netflix Ends Free Trials in the US." URL https://www.pcmag.com/news/netflix-ends-free-trials-in-the-us, accessed June 18, 2025.

Keane, Michael P. (1997). "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior." *Journal of Business & Economic Statistics*, 15(3): 310–327.

Kim, Jaehwan, Greg M Allenby and Peter E Rossi (2002). "Modeling Consumer Demand for Variety." *Marketing Science*, 21(3): 229–250.

Kim, Jungsik, Sangwon Lee, Changhyun Park and Wei Zhang (2025). "Gift Diffusion and Social Transmission in Live Streaming Communities." *Journal of Interactive Marketing*, 65: 123–140.

Lambrecht, Anja and Kanishka Misra (2017). "Fee or Free: When Should Firms Charge for Online Content?" *Management Science*, 63(4): 1150–1165.

Lee, Min Kyung (2018). "Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management." *Big Data & Society*, 5(1): 2053951718756684.

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

Lemmens, Aurélie, Jason M. T. Roos, Sebastian Gabel, Eva Ascarza, Hernán A. Bruno, Brett R. Gordon, Ayelet Israeli, Elea McDonnell Feit, Carl F. Mela and Oded Netzer (2025). "Personalization and Targeting: How to Experiment, Learn & Optimize." *International Journal of Research in Marketing*.

Li, Fan, Kari Lock Morgan and Alan M. Zaslavsky (2018). "Balancing Covariates via Propensity Score Weighting." *Journal of the American Statistical Association*, 113(521): 390–400.

Lin, Yan, Dai Yao and Xingyu Chen (2021). "Happiness Begets Money: Emotion and Engagement in Live Streaming." *Journal of Marketing Research*, 58(3): 417–438.

Lu, Shijie, Dai Yao, Xingyu Chen and Rajdeep Grewal (2021). "Do Larger Audiences Generate Greater Revenues Under Pay What You Want? Evidence from a Live Streaming Platform." *Marketing Science*, 40(5): 964–984.

Matz, Sandra C and Oded Netzer (2017). "Using Big Data as a Window into Consumers' Psychology." *Current Opinion in Behavioral Sciences*, 18: 7–12.

McAlister, Leigh and Edgar Pessemier (1982). "Variety Seeking Behavior: An Interdisciplinary Review." *Journal of Consumer Research*, 9(3): 311–322.

Oestreicher-Singer, Gal and Lior Zalmanson (2013). "Content or Community? A Digital Business Strategy for Content Providers in the Social Age." *MIS Quarterly*, 37(2): 591–616.

Pattabhiramaiah, Adithya, S Sriram and Puneet Manchanda (2019). "Paywalls: Monetizing Online Content." *Journal of Marketing*, 83(2): 19–36.

Rafieian, Omid, Anuj Kapoor and Amitt Sharma (2024). "Multiobjective Personalization of Marketing Interventions." *Marketing Science*.

Reza, Syed Mashrur, Teck-Hua Ho, Qiong Ling and Yun Shi (2021). "Free Trial Promotions and Consumer Subscription Decisions in Online Streaming Services." *Journal of Marketing*, 85(6): 171–190.

Rubin, D.B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5): 688–701.

Sahni, Navdeep S, Dan Zou and Pradeep K Chintagunta (2017). "Do Targeted Discount Offers Serve as Advertising? Evidence from 70 Field Experiments." *Management Science*, 63(8): 2688–2705.

Seetharaman, P. B., Andrew Ainslie and Pradeep K. Chintagunta (1999). "Investigating Household State Dependence Effects across Categories." *Journal of Marketing Research*, 36(4): 488–500.

Shah, Denish, V. Kumar and Kihyun Hannah Kim (2014). "Managing Customer Profits: The Power of Habits." *Journal of Marketing Research*, 51(6): 726–741.

Simonov, Andrey, Raluca M. Ursu and Carolina Zheng (2023). "Suspense and Surprise in Media Product Design: Evidence from Twitch." *Journal of Marketing Research*, 60(1): 1–24.

von Wangenheim, Florian and Tomás Bayón (2007). "Behavioral Consequences of Overbooking Service Capacity." *Journal of Marketing*, 71(4): 36–47.

Wang, Kitty and Avi Goldfarb (2017). "Can Offline Stores Drive Online Sales?" *Journal of Marketing Research*, 54(5): 706–719.

Xu, Sikun, Raphael Thomadsen and Dennis J. Zhang (2025). "Winner's Curse in Data-Driven Decision-Making: Evidence and Solutions.", URL https://sics.haas.berkeley.edu/pdf_2025/paper_xtz.pdf.

Yang, Jeremy, Dean Eckles, Paramveer Dhillon and Sinan Aral (2024). "Targeting for Long-Term Outcomes." *Management Science*, 70(6): 3841–3855.

Yoganarasimhan, Hema, Elnaz Barzegary and Abhijit Pani (2023). "Design and Evaluation of Optimal Free Trials." *Management Science*, 69(10): 6215–6238.

Online Appendix

Contents

A Data Collection

A.1 Creator Sample

Twitch has over 1 million content creators (streamers), each with their own channel; however, many of those streamers have little to no viewers. For tractability, we focus the analysis on the top 100 English language streamers. We used a Twitch data aggregator[1] to select the 100 highest-ranked English language streamers by total viewership in the 90 days leading up to the start of the sample (July 1, 2022), and who are also ranked in the top 150 by watch time in the 30 days leading up to the start of the sample. The 90 day selection criterion helps define streamers with a sustained presence on the platform, while the 30 day criterion helps identify streamers who were currently active and popular on the platform. Table A.1 contains the full list of 100 streamers included in our sample. While the majority of these channels are operated by individuals, a select few (e.g. esl_csgo) are managed by organizations. These organization-managed channels may not have many active subscribers, but they are useful for measuring spillovers in watch and engagement.

Table A.1: Creator Channels in Sample

39daph	dreamleague	loltyler1	saintone
aceu	elajjaz	lpl	shahzam
adinross	esfandtv	lvndmark	shivfps
amouranth	esl_csgo	mande	shroud
asmongold	esl_dota2	maximilian_dood	sinatraa
aydan	fextralife	mizkif	smitegame
barbarousking	foolish_gamers	moistcr1tikal	sodapoppin
beyondthesummit	forsen	moonmoon	summit1g
blastpremier	fuslie	nickeh30	swagg
boxbox	gamesdonequick	nickmercs	sweetdreams
brawlhalla	gorgc	ninja	symfuhny
brucedropemoff	gtawiseguy	nmplol	sypherpk
buddha	hasanabi	northernlion	tarik
castro_1021	hiko	otzdarva	tenz
chess	iitztimmy	penta	tfue
chilledchaos	ironmouse	pestily	thebausffs
classybeef	jinnytty	pokelawls	topsonous
clix	k3soju	pokimane	trainwreckstv
coconutb	kyedae	prod	tsm_imperialhal
cohhcarnage	kyle	quickybaby	uberhaxornova
daltoosh	lck	quin69	valorant
deuceace	lcs	rainbow6	xqc
disguisedtoast	lec	ranboolive	yourragegaming
distortion2	leekbeats	rocketleague	zealsambitions
doublelift	lirik	roshtein	zerkaa

[1] https:\sullygnome.com

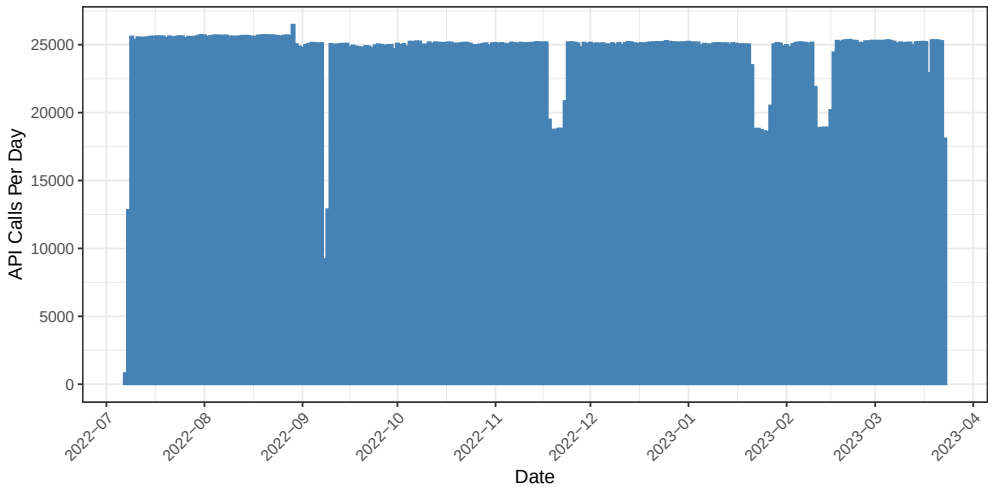
A.2 Data Collection

Viewing Behavior

We collected watch data by querying the https://tmi.twitch.tv/ API endpoint. This endpoint returned a JSON file containing the broadcaster's name, live status, and a list of logged-in chatters (users) watching the channel. We rotated through 100 streamers and queried the endpoint for each streamer approximately once per five minutes from July 2022 until the endpoint was officially shut down in early April 2023.[2] The approximately five-minute frequency of collection ensures at least one observation per channel in the six-minute window leading up to any cohort start time, which we use in our definition of eligible users.

Short data gaps occurred due to power outages and reaching device storage limits. Figure A.1 shows the distribution of successful API calls over the sample period. The maximum height of the bars is determined by the cadence of our data collection procedure. We had four separate computers running, each collecting 1/4 of the data. The first dip (09/2022) affected all computers; the subsequent dips each only affected one computer. The dips resulted from human/technology/utilities-related failures in our data collection infrastructure rather than Twitch service outages that could be correlated with treatment. We therefore have no reason to believe these short outages will affect the results in any substantive way.

Figure A.1: Viewing Behavior Collection Outages



Notes: This figure documents the number of successful API calls per day.

[2]See https://discuss.dev.twitch.com/t/legacy-chatters-endpoint-shutdown-details-and-timeline-april-2023/43161 for the official announcement.

Chat and Subscription Behavior

Channel-specific chat logs were collected using the Chatty application,[3] which allowed for the simultaneous monitoring of all channels in our sample. For each channel, we recorded the complete chat transcript, including all messages sent and subscription status change notifications. These data were collected on a different computer than the viewing behavior.

[3] https://chatty.github.io/

A.3 Data Cleaning and Processing

Our estimation sample is comprised of cohorts with treatment dates ranging from September 2022 to December 2022 to ensure each cohort includes a two-month pre-treatment and four-month post-treatment observation period for all eligible users.

We limit our analysis to cohorts with at least five treated users. The primary motivation for this filter is computational ease, as each cohort used in the estimation requires computing cohort-specific covariates and outcomes for a separate set of control users. Cohorts with five or more treated users account for 80% of treated users across the entire set of algorithmically-allocated premium benefits (i.e., cohorts of size two or more) in our data. The average number of treated users per cohort in the final estimation sample is 10.5.

We further refine our sample to include only eligible users with at least 30 minutes and no more than 21,600 minutes (equivalent to twelve hours per day) of watch time across all observed channels during the month prior to treatment. This filter, which removes fewer than 2% of users, excludes users that exhibit extreme watch levels that is likely more consistent with bot activity than human behavior.

B Estimation

B.1 Double/De-biased Machine Learning

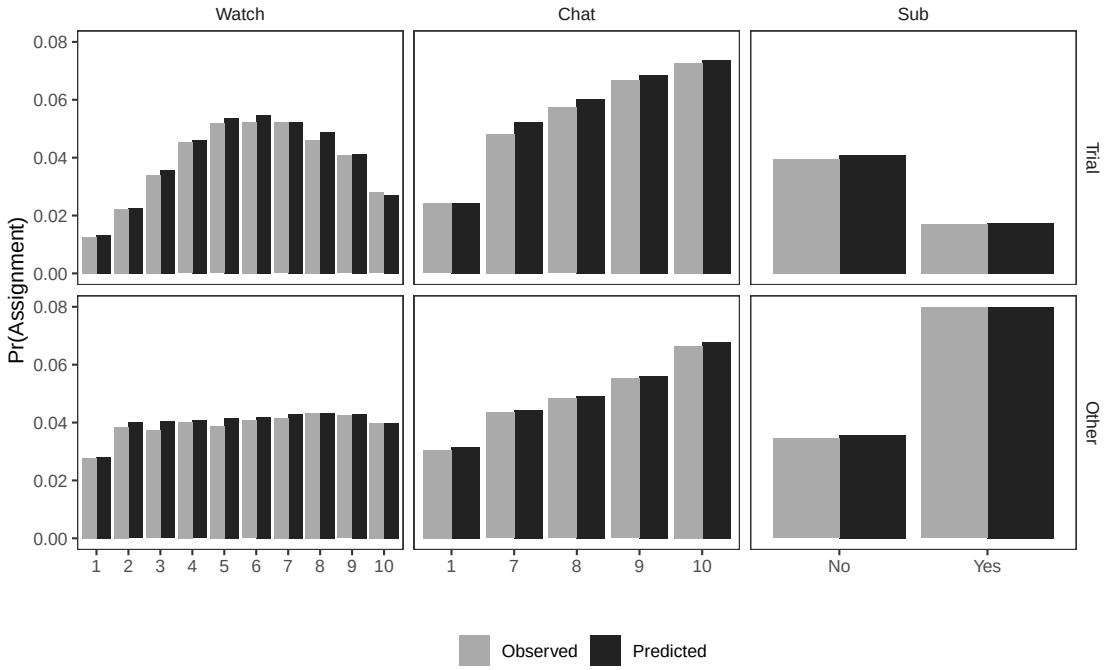
Tuning Parameters

Hyperparameter tuning across every assignment and outcome model would be computationally intensive. We instead tune hyperparameters for three models: the assignment model, one representative continuous outcome model (trial channel watch during the trial month), and one representative binary outcome model (paid subscriptions to other channels during the trial month). We use 5-fold cross-validation with cohort-level splits to sequentially tune seven hyperparameters over a coarse grid of parameters values: number of rounds (nrounds), maximum tree depth (max_depth), minimum child weight (min_child_weight), minimum loss reduction (gamma), row subsampling fraction (subsample), column subsampling fraction (colsample_bytree), learning rate (eta), and loss weight on treated units (scale_pos_weight). The three outcomes converge on similar optimal hyperparameters, from which we choose one set to apply across all models: nrounds = 1000, max_depth = 6, min_child_weight = 1, gamma = 1, subsample = 0.6, colsample_bytree = 0.9, eta = 0.01, and scale_pos_weight = 1.

Estimated Assignment

We do an out-of-sample prediction exercise to illustrate how treatment assignment probability varies with pre-treatment covariates. While our double-robust estimation approach requires only one of the two models to be correctly specified, an accurate assignment model increases our confidence in our identifying assumptions. We split the sample into training cohorts (50%) and holdout cohorts (50%), estimate the assignment model on the training sample using our tuned hyperparameters, and compare predicted probabilities to observed treatment rates in the holdout sample.

Figure B.2: Prediction Exercise



Notes: Linear projection of user-level estimated assignment probability onto deciles of user-level characteristics. Watch Decile: trial channel watch time in the last month. Chat Decile: count of chat messages on the trial channel in the last month. Fewer than 40% of users chat, so the first decile contains all users that do not engage in that behavior. The average assignment probability of 1/26 in each panel reflects the 25:1 control to treatment ratio.

Figure B.2 displays how predicted and observed assignment probabilities vary with pre-treatment watch, chat, and subscription behaviors across both trial and other channels. The projections show several features of the assignment algorithm. First, allocation is not uniformly random. More active users on the trial channel more frequently receive the promotion, but this relationship reverses beyond a threshold; very high activity reduces assignment probability. This inverted U-shape suggests the algorithm rewards engagement while discouraging bot-like behavior, which aligns with the platform's statement on how allocation works. Interestingly, assignment seems to close to uniformly random with respect to watch time on other channels. Overall, predicted and observed assignment very closely align.

Table B.2 presents pre-treatment covariate descriptives separately for the treated and control users. There are small differences in baselines between the two groups, as shown in balance tests reported in Figure 3. Treated users have lower pre-treatment engagement with the trial channel across multiple metrics: their average watch time on day 1 pre-treatment is 129.51 minutes compared to 158.65 minutes for control users, and their chat activity is similarly lower. Treated users also show higher subscription rates to other channels at baseline (0.12 versus 0.05) and greater engagement with the broader platform. These

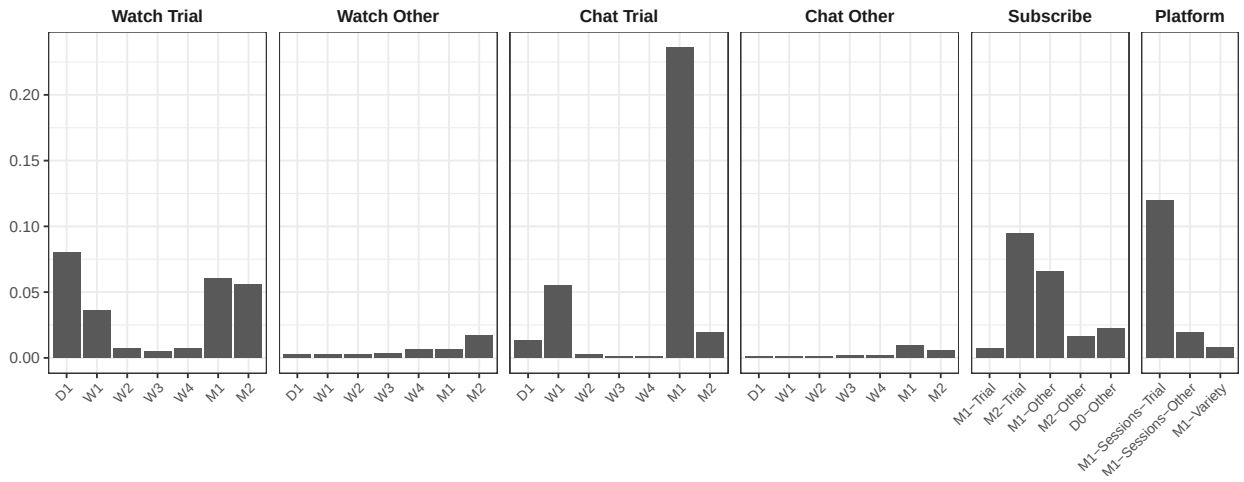
differences highlight the importance of accounting for pre-treatment covariates.

Variable Importance

Figure B.3 shows variable importance measures from the assignment model. Chat activity on the trial channel during the first pre-treatment month (M1) is the strongest predictor of assignment. Session counts and watch activity on the trial channel also predict assignment, as do subscribe-related covariates. In contrast, watch time and chat activity on other channels are not strong predictors of treatment assignment. This pattern demonstrates that user behavior on the trial channel is a key determinant of assignment, while activity elsewhere on the platform has less influence.

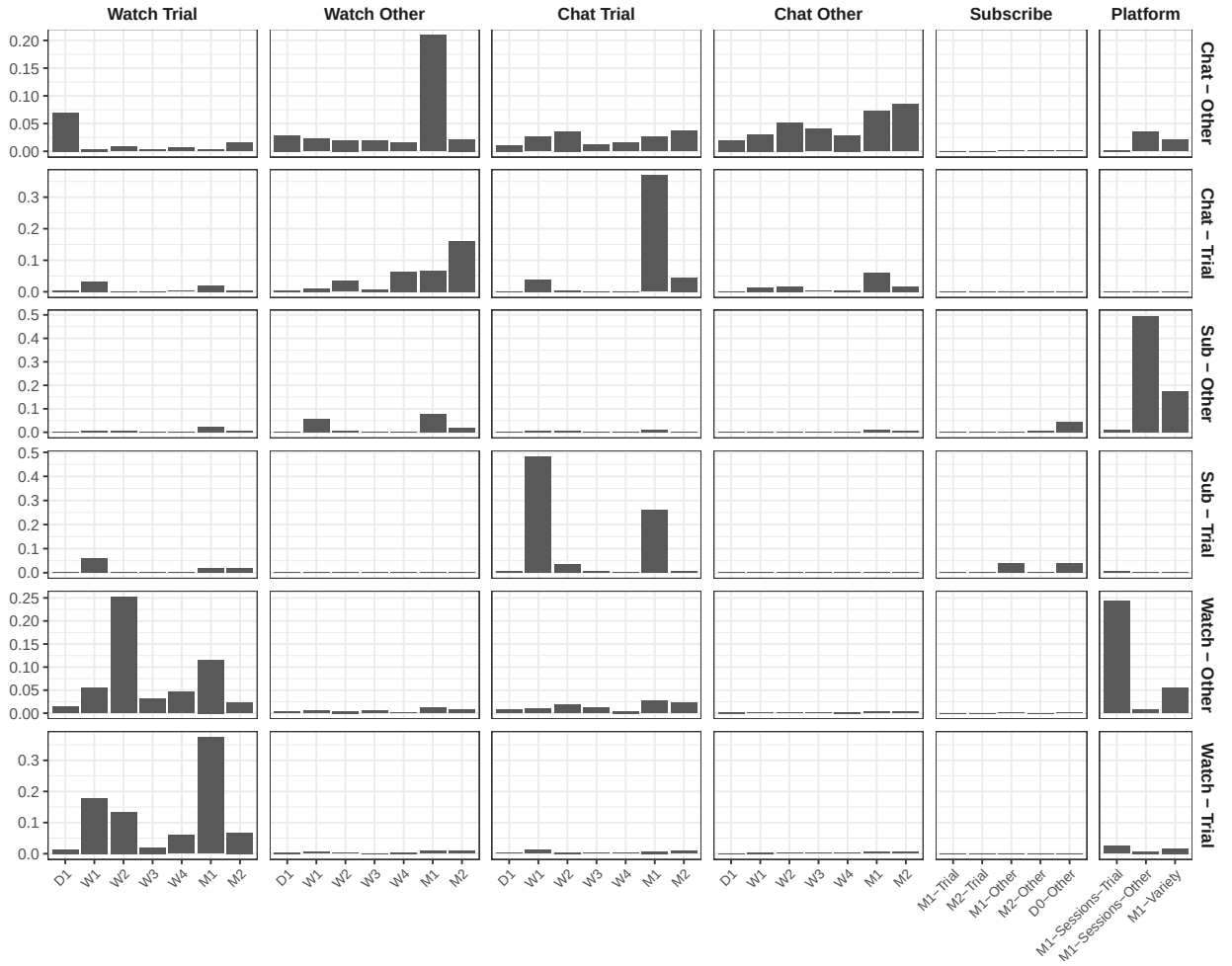
Figure B.4 shows variable importance measures for all covariates (columns) for the primary one-month post-treatment outcomes (rows). Unlike assignment, subscription activity does not strongly predict any outcomes. Chat outcomes are predicted by multiple sets of covariates, with chat activity on the respective channel or channel group generally being the strongest predictor. For subscription outcomes, platform-wide session counts (a function of watch behavior) are the strongest predictors. For watch-related outcomes, pre-treatment watch behavior on the trial channel is the strongest predictor. Overall, each group of covariates is a meaningful predictor of at least one group of outcomes or assignment.

Figure B.3: Assignment Model Variable Importance



Notes: Variable importance measures from the assignment model across different covariate sets (columns). Importance represents the average improvement in prediction accuracy attributable to splits on each variable. Higher values indicate greater importance for predicting treatment effect heterogeneity. Time periods: D1 = day 1, W1-W4 = weeks 1-4, M1-M4 = months 1-4 (all pre-treatment).

Figure B.4: Causal Forest Variable Importance



Notes: Variable importance measures from gradient boosting models for each outcome (rows) across different covariate sets (columns). Importance represents the average improvement in prediction accuracy attributable to splits on each variable. Higher values indicate greater importance for predicting treatment effect heterogeneity. Time periods: D1 = day 1, W1-W4 = weeks 1-4, M1-M4 = months 1-4 (all pre-treatment).

Table B.2: Covariate Descriptives by Treatment

Treated Users								
		Day Pre	Week Pre				Month Pre	
Behavior	Channel	1	1	2	3	4	1	2
Watch	Trial	129.51 (159.88)	585.52 (649.34)	417.09 (573.48)	365.45 (545.67)	341.14 (542.04)	1802.34 (1980.52)	1307.29 (2007.39)
	Other	81.14 (169.91)	560.62 (879.92)	549.57 (874.17)	531.42 (861.43)	526.26 (872.56)	2318.64 (3326.75)	2294.00 (3725.41)
Chat	Trial	2.28 (11.67)	9.36 (40.98)	6.20 (35.75)	5.54 (40.08)	5.10 (36.36)	27.51 (133.93)	18.79 (120.21)
	Other	1.25 (11.60)	8.00 (53.73)	7.58 (52.72)	7.48 (51.19)	7.41 (52.26)	32.53 (194.16)	29.92 (198.54)
Subscribe	Trial	0.00†	–	–	–	–	0.02 (0.15)	0.05 (0.21)
	Other	0.12 (0.33)	–	–	–	–	0.18 (0.38)	0.15 (0.35)
Session Count	Trial	–	–	–	–	–	24.43 (20.65)	17.59 (21.47)
	Other	–	–	–	–	–	40.33 (52.70)	39.35 (57.05)
Channel Variety	Platform	–	–	–	–	–	6.69 (5.14)	6.17 (5.35)

Control Users								
		Day Pre	Week Pre				Month Pre	
Behavior	Channel	1	1	2	3	4	1	2
Watch	Trial	158.65 (197.05)	588.38 (780.94)	415.05 (658.80)	383.16 (638.98)	371.17 (637.47)	1860.37 (2443.89)	1503.36 (2447.63)
	Other	81.01 (198.61)	535.95 (912.16)	521.13 (880.90)	508.53 (868.04)	506.65 (878.34)	2219.81 (3376.16)	2254.58 (3892.42)
Chat	Trial	1.55 (12.88)	5.85 (49.06)	4.48 (44.30)	4.40 (49.28)	4.32 (44.62)	20.23 (172.42)	17.17 (154.86)
	Other	0.66 (9.78)	4.27 (43.03)	4.08 (41.22)	4.10 (40.30)	4.15 (41.19)	17.75 (155.29)	16.95 (154.56)
Subscribe	Trial	0.00†	–	–	–	–	0.05 (0.23)	0.09 (0.29)
	Other	0.05 (0.22)	–	–	–	–	0.08 (0.28)	0.08 (0.27)
Session Count	Trial	–	–	–	–	–	22.94 (25.28)	18.40 (24.92)
	Other	–	–	–	–	–	37.19 (52.72)	37.22 (57.88)
Channel Variety	Platform	–	–	–	–	–	6.19 (5.03)	5.78 (5.21)

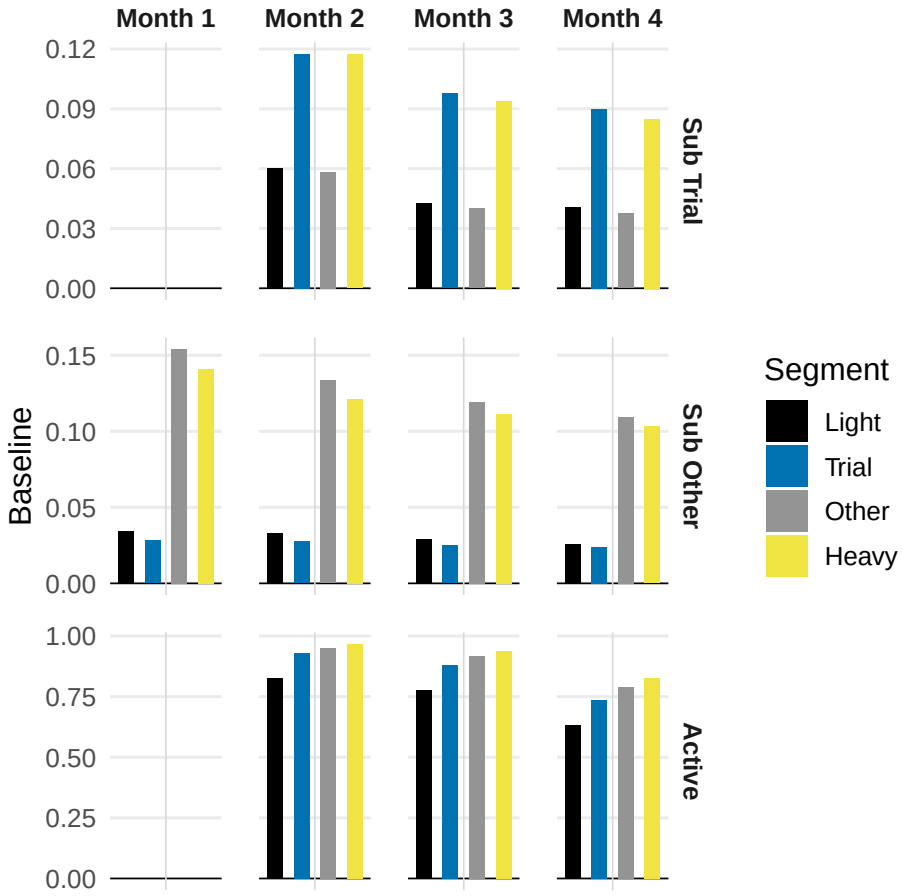
Notes: Each cell contains the mean of one behavior-channel-time covariate, with standard deviation in parentheses. Time periods denote time intervals defined relative to the cohort start date. D1 is a one day lead, W1 is a one week lead, W2 is a two week lead (excluding the first week) with W3 and W4 defined similarly, M1 is the entire duration (30 days) of the benefit period, and M2, M3, and M4 are the non-overlapping periods two, three, and four months after the cohort start date, respectively. Trial and Other refer to behaviors on the trial channel and all other channels, respectively.

†By construction, the sample only includes users who are not subscribed to the trial channel at baseline.

B.2 CATE Baselines

Figure B.5 shows the baselines for the CATEs reported in Figure 4. As expected, users that watch more of the trial channel (trial and heavy segments) are more likely to subscribe to the trial channel. Likewise users that watch more of the other channels (other and heavy segments) are more likely to subscribe to an other channel. Month one trial baselines are omitted because treated users are, by definition, subscribed. Retention rates on the platform are high in months one and two, meaning that even small treatment effects may be indicative of a large percentage change in the propensity to churn. By construction, all eligible users are observed, and thus retained, on the platform in the month 1.

Figure B.5: Heterogeneous Treatment Effects by Watch Level - Baselines



Notes: This figure shows the baselines for the CATEs reported in Figure 4.

B.3 Additional Policy Frontier Figures and Tables

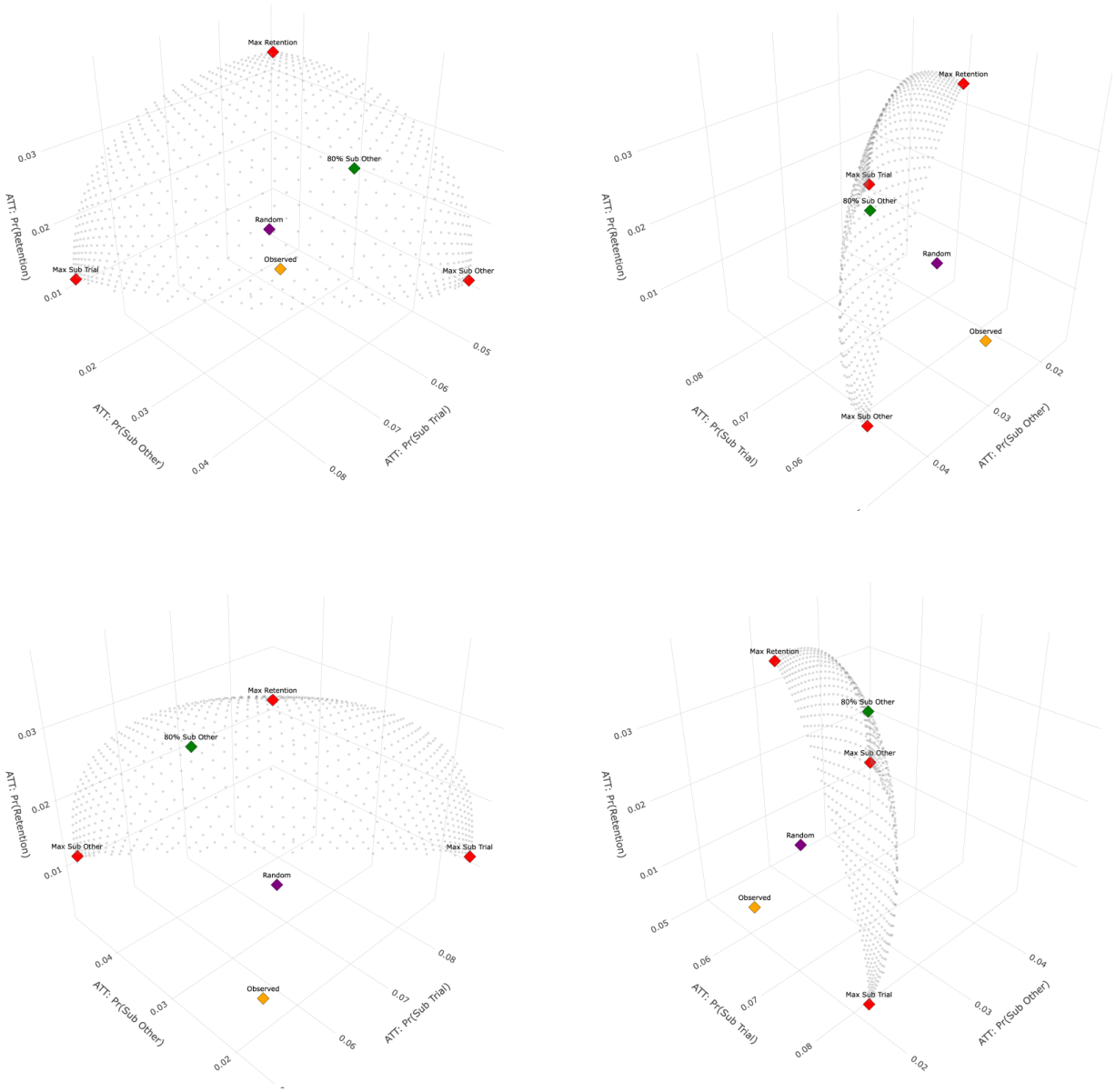
Figure B.6 shows the estimated Pareto frontier across the three multi-objective optimization outcomes. The four figures (different rotations of the same frontier) show that the Pareto-optimal policies outperform both the observed platform allocation and random assignment across all three outcomes. Table B.3 presents the specific estimates for Figure 7.

Table B.3: Multi-Objective Policy Frontier

Policy	$(\beta_1, \beta_2, \beta_3)$	ATT Pr(Sub Trial)	ATT Pr(Sub Other)	ATT Pr(Retention)
Max DV1 (SO)	(1.00, 0.00, 0.00)	0.0875	0.0157	0.0117
Max DV2 (SO)	(0.00, 1.00, 0.00)	0.0511	0.0476	0.0116
Max DV3 (SO)	(0.00, 0.00, 1.00)	0.0616	0.0251	0.0384
Equal Weights	(0.33, 0.33, 0.33)	0.0709	0.0301	0.0279
80% Max DV1	(0.28, 0.25, 0.47)	0.0693	0.0278	0.0332
80% Max DV2	(0.23, 0.45, 0.33)	0.0621	0.0376	0.0277
80% Max DV3	(0.33, 0.28, 0.40)	0.0716	0.0277	0.0305
Observed	—	0.0571	0.0226	0.0023
Random	—	0.0633	0.0259	0.0138

Notes: Average Treatment Effects on the Targeted shown in Figure 7.

Figure B.6: 3D Policy Frontier



Notes: Four rotations of the same 3D Pareto frontier across three outcomes: trial channel subscriptions, other-channel subscriptions, and retention at four months. Red diamonds mark policies maximizing individual outcomes; the green diamonds an example balanced policy (80% Sub Other); purple diamonds the observed platform allocation and a random policy.

C Robustness Checks

User engagement on the Twitch platform is highly right-skewed and characterized by outliers. To reduce the influence of anomalous activity on our results, we filter out users that exhibit activity inconsistent with human behavior and consider transformations of our outcomes. We next show the robustness of our results to these sample and outcome definitions.

C.1 Watch Time Filter

In Table C.4, we show the robustness of the main results to alternative filtering thresholds for removing users with extreme watch levels from the sample. Our main estimation sample uses a threshold of 21,600 minutes (equivalent to twelve hours per day) of watch time across all observed channels during the month prior to treatment. We consider two alternative thresholds at 14,400 minutes (eight hours per day on average) and 28,800 minutes (sixteen hours per day on average). In general, we find that removing additional users from the far right tail of the watch distribution leads to an increase in the size of estimated watch and chat treatment effects, suggesting that the 12 hour per day threshold we use is conservative.

Table C.4: Average Treatment Effects for Alternative Watch Filters

		Maximum Average Daily Watch		
		8 hours	12 hours	16 hours
Watch	Trial	109.387	97.777	88.428
		(6.738)	(6.818)	(6.875)
	Other	144.980	113.390	81.797
		(25.446)	(25.993)	(26.028)
Chat	Trial	7.618	7.530	7.453
		(0.335)	(0.342)	(0.338)
	Other	2.281	1.864	1.696
		(0.340)	(0.363)	(0.374)
Subscribe	Other	0.041	0.042	0.042
		(0.001)	(0.001)	(0.001)

Notes: Estimated ATEs for Month 1 outcomes. Each column corresponds to a different sample which is constructed by removing eligible users with average daily watch time during the month prior to treatment exceeding the indicated threshold (8, 12, or 16 hours per day). The allocation model, $\hat{e}(x)$ is common across every cell in a given column; the outcome models, $\hat{m}(w, x)$, are specific to each cell. Standard errors clustered at the cohort level are in parentheses.

C.2 Threshold Watch and Chat Outcomes

As an alternative to the level and logged watch outcomes presented in the main text, in Table C.5 we present a sequence of binary outcomes that indicate whether a user's watch time exceeds a given threshold. Each treatment effect can be interpreted as a change in the probability of watch time exceeding that threshold. On the trial channel during the benefit period, the greatest changes occur in the 4-16 hour range, corresponding to moderate levels of monthly viewership. There is no meaningful change in the probability of becoming a very heavy user with 128 hours of total watch time (equivalent to greater than 4 hours of watch time per day). On other channels, the largest changes are at smaller thresholds, between 30 minutes and 4 hours during the benefit period. Beyond the 16 hour threshold, treatment effects are small and tend towards statistical insignificance. Together, these results suggest that expansionary effects are driven by relatively lighter users on the platform rather than heavy users.

Table C.5: Watch Threshold Effects

	Threshold: Pr($Watch \geq x$)								
Channel	30 min	1 hr	2 hrs	4 hrs	8 hrs	16 hrs	32 hrs	64 hrs	128 hrs
Trial	0.008	0.013	0.020	0.026	0.032	0.028	0.019	0.007	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Other	0.011	0.010	0.009	0.008	0.006	0.003	0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: Estimated ATEs for binary outcome variables defined as 1 if a user's watch level exceeds the given threshold during the Month 1 period. Standard errors clustered at the cohort level are in parentheses.

Table C.6 shows similar treatment effects for chat threshold outcomes. Once again, the largest effects are for thresholds indicative of light to moderate levels of engagement. We also see that effects are driven by a mixture of intensive and extensive margin changes in behavior. The probability of sending at least one chat increased by 0.11 on the trial channel and 0.03 on other channels, showing an increase in the propensity to chat. All other levels of engagement show smaller but significant increases in chat activity by users that already engaged in this behavior.

Table C.6: Chat Threshold Effects

	Threshold: $\Pr(Chat \geq x)$								
Channel	1	2	4	8	16	32	64	128	256
Trial	0.114	0.091	0.073	0.058	0.044	0.032	0.020	0.012	0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Other	0.025	0.019	0.014	0.011	0.008	0.004	0.003	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)

Notes: Estimated ATEs for binary outcome variables defined as 1 if a user's count of chat messages exceeds the given threshold during the Month 1 period. Standard errors clustered at the cohort level are in parentheses.

C.3 Types of Subscriptions

As a subsidiary of Amazon, Twitch offers special benefits to Amazon Prime customers. During the time of our data collection, users were able to subscribe for free to one channel of their choice each month by linking their Amazon Prime account to Twitch. In the average treatment effects presented in Section 6, we show the effect of temporary premium benefits on paid subscriptions. In Table C.7, we show the effect of temporary premium benefits on user propensity to subscribe using Prime benefits.

Table C.7: Paid and Prime Subscriptions

		Month Post			
Outcome	Channel	1	2	3	4
Subscribe (Paid)	Trial		-0.073	-0.018	0.059
			(0.001)	(0.001)	(0.001)
	Other	0.042	0.047	0.030	0.025
		(0.001)	(0.001)	(0.001)	(0.001)
Subscribe (Prime)	Trial		-0.016	-0.002	-0.003
			(0.000)	(0.001)	(0.001)
	Other	0.042	0.035	0.029	0.027
		(0.001)	(0.001)	(0.001)	(0.001)

Notes: Each cell shows the overlap-weighted ATEs from a double-robust model with gradient boosted allocation and outcome models. Standard errors clustered at the cohort level are in parentheses.

Similar to the negative short-run effect of premium benefits on paid subscriptions to the trial channel, treated users are also less likely to activate a Prime subscription on the trial channel. The Prime subscription effects are smaller in magnitude than the paid subscription effects on the trial channel. In contrast, treated users show an approximately equal increase in propensity to activate a Prime subscription on a non-trial channel as they do to purchase a subscription to a non-trial channel. These results suggest the automatic promotion we study interacts with the opt-in Prime promotion to further encourage exploration and deeper engagement on the platform.