Optimizing User Engagement through Adaptive Ad Sequencing

Omid Rafieian[*]

Cornell Tech and Cornell University

June 30, 2019

Abstract

Mobile in-app advertising has grown exponentially in the last few years. In-app ads are often shown in a sequence of short-lived exposures for the duration of a user's stay in an app. The current state of both research and practice ignores the dynamics of ad sequencing and instead adopts a myopic framework to serve ads. In this paper, we propose a unified dynamic framework for adaptive ad sequencing that optimizes user engagement in the session, e.g., the number of clicks or length of stay. Our framework comprises of two components – (1) a Markov Decision Process that captures the domain structure and incorporates inter-temporal trade-offs in ad interventions, and (2) an empirical framework that combines machine learning methods such as Extreme Gradient Boosting (XGBoost) with ideas from the causal inference literature to obtain counterfactual estimates of user behavior. We apply our framework to large-scale data from the leading in-app ad-network of an Asian country. We document significant gains in user engagement from adopting a dynamic framework. We show that our forward-looking ad sequencing policy outperforms all the existing methods by comparing it to a series of benchmark policies often used in research and practice. Further, we demonstrate that these gains are heterogeneous across sessions: adaptive forward-looking ad sequencing is most effective when users are new to the platform. Finally, we use a descriptive approach to explain the gains from adopting the dynamic framework.

Keywords: advertising, personalization, adaptive interventions, dynamic policy, Markov Decision Process, machine learning, reinforcement learning

1 Introduction

1.1 Adaptive Interventions in Mobile Advertising

Consumers now spend a significant portion of their time on mobile devices. The average time spent on mobile devices by US adults has grown steadily over the last few years (eMarketer, 2019a). This demand expansion, in turn, has amplified marketing activities towards mobile users. In 2018, mobile advertising generated over \$71 billion in the US, accounting for roughly double the share of its digital counterpart, desktop advertising (eMarketer, 2019b). Most of this growth in mobile advertising is attributed to in-app ads, i.e., ads shown inside mobile apps. Indeed, in-app advertising is now the dominant channel for mobile advertising, generating over 80% of ad spend in the mobile advertising category (eMarketer, 2018).

Two key features of mobile in-app ads have contributed to their growth. First, the mobile app ecosystem has excellent user tracking ability, thereby allowing "personalization" of ad interventions, i.e., targeting of users based on their prior behavioral history (Han et al., 2012). Second, in-app ads are usually short-lived and dynamic in nature: each ad intervention is shown for a fixed amount of time (e.g., 30 seconds or one minute) inside the app, and is then followed by another ad intervention. As such, a user can see multiple ad exposures within a session.[1] This is in contrast with the common practice in desktop advertising, where ads remain fixed throughout a session. Short-lived ads together with personalization potential make in-app advertising amenable to "adaptive interventions", i.e., targeting of ads based on time-varying behavioral information about users to maximize user engagement.

We illustrate the general schema for a mobile publisher's ad sequencing problem in Figure 1.[2] In this problem, the publisher has to decide which ad to serve in each period that the user stays in the app. We characterize three separate pieces of information that the publisher can incorporate when deciding which ad to show in any impression: (1) the pre-session information, which consists of user characteristics as well as the behavioral history of the user up until the current session, (2) the session-level information, which is the sequence of ads the user has seen so far within the session, as well as his response to each, and (3) the future information, which captures the publisher's expectations on how the sequence is likely to evolve in the next periods. The current research and practice mostly focus on the first two pieces and overlook the future information (Li et al., 2010; Lu et al., 2010; Rafieian and Yoganarasimhan, 2020b). One reason is that capturing future information often adds significantly to the complexity of the problem. In addition, the returns

---

[1]A session is an uninterrupted time that a user spends inside an app.

[2]In this paper, we use the publisher, ad-network, and platform interchangeably, when we refer to the agent who makes the ad placement decision.
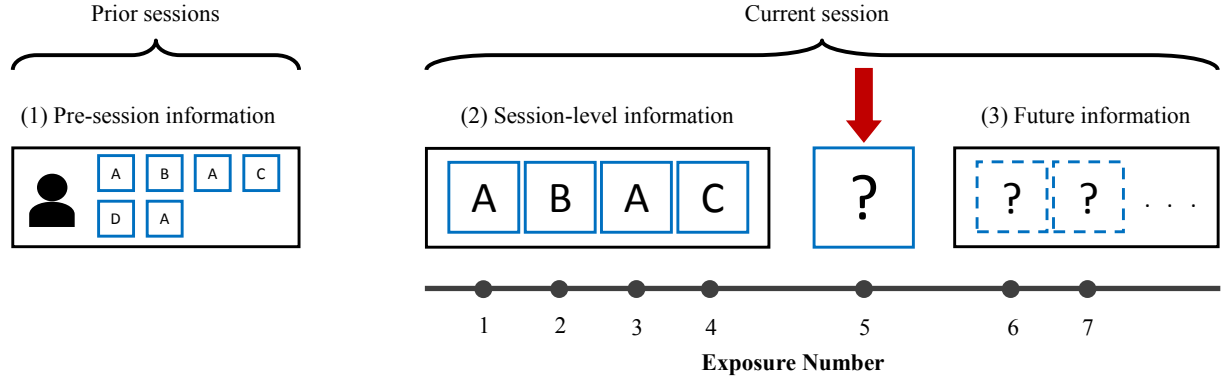
3

Figure 1: A visual schema the publisher's ad sequencing decision and different types of information available.

from adopting a forward-looking model are not clear. Thus, the publisher's decision on whether to use a dynamic framework boils down to whether incorporating future information helps her achieve a better outcome.

In principle, if there is no interaction between sequential ad interventions, incorporating future information will not improve the ad sequencing policy. However, at the advertiser level, the extant literature has documented various empirical effects that highlight such inter-temporal trade-offs. Prior findings on spillover and carryover effects of advertising rule out the independence of ads shown in a sequence (Rutz and Bucklin, 2011; Sahni, 2015a). In particular, in the context of mobile in-app advertising, Rafieian and Yoganarasimhan (2020a) find that a higher variety of previous ads shown within a sequence leads to a higher likelihood of click on the next ad. At the same time, they show that an ad will have a higher chance of being clicked if it has been shown more within the sequence of prior ads. These findings suggest a natural inter-temporal trade-off: higher variety by definition implies fewer repeat exposures of each ad. While these effects are well-established at the advertiser level, neither research nor practice has looked into how to collectively incorporate these findings to dynamically sequence ads to optimize publishers' outcomes. Thus, to the extent that these effects have significant impact on user's engagement with ads, it is important to develop a dynamic framework for ad sequencing and examine the gains from adopting such a framework.

1.2 Research Agenda and Challenges

In this paper, we propose a dynamic framework that incorporates the inter-temporal trade-offs in ad sequencing and develops an optimal policy that optimizes user engagement with ads in a session. User engagement with ads or the match between users and ads is particularly an important outcome for publishers, as it is the main channel through which the publisher can create value by

ad sequencing. We can use different metrics to measure user engagement with ads depending on the context of advertising. In the context of mobile in-app advertising, click is a reasonably good metric for the user engagement with ads because of two reasons. First, all ads are mobile apps whose objective is to get more clicks and installs (e.g., performance ads). Second, clicks are directly linked to revenues as most in-app ad publishers employ cost-per-click or cost-per-install as their monetization strategy.

Therefore, we use a click-maximizing objective and seek to answer the following three questions in this paper:

1. How can we develop a dynamic theoretical framework that incorporates the inter-temporal trade-offs in ad sequencing and designs a policy that maximizes the expected number of clicks per session?

2. How can we empirically evaluate the performance of this dynamic ad sequencing policy relative to other benchmark policies?

3. What are the gains from using a dynamic framework to allocate ads? What explains the differences in outcomes and interventions under a fully dynamic sequencing policy and other benchmark policies?

We need to overcome three major challenges to satisfactorily answer this set of questions. First, to capture the dynamics of this problem, we need a theoretical framework that incorporates inter-temporal trade-offs in the publisher's ad allocation decision. Second, to develop a click-maximizing dynamic sequencing policy, we need to obtain accurate personalized counterfactual estimates of user behavior that allows us to evaluate all feasible policies (and not only those implemented in the data). As such, we need a setting with enough randomization in the ad allocation process that enables us to exploit this randomization and obtain accurate counterfactual estimates of user behavior. Finally, to measure the gains from the dynamic framework and explore the differences in sequencing strategies across sessions, we need to have a solution concept that determines the optimal dynamic policy and an evaluation approach that accurately estimates the outcomes under the new optimal policy for each session.

1.3 Our Approach

In this paper, we present a unified three-pronged framework that addresses these challenges and develops a forward-looking adaptive sequencing policy to maximize user engagement with ads. We present an overview of our approach in Figure 2. As shown in the top row of this figure, we start with a theoretical framework that models the domain structure of our problem and allows us to
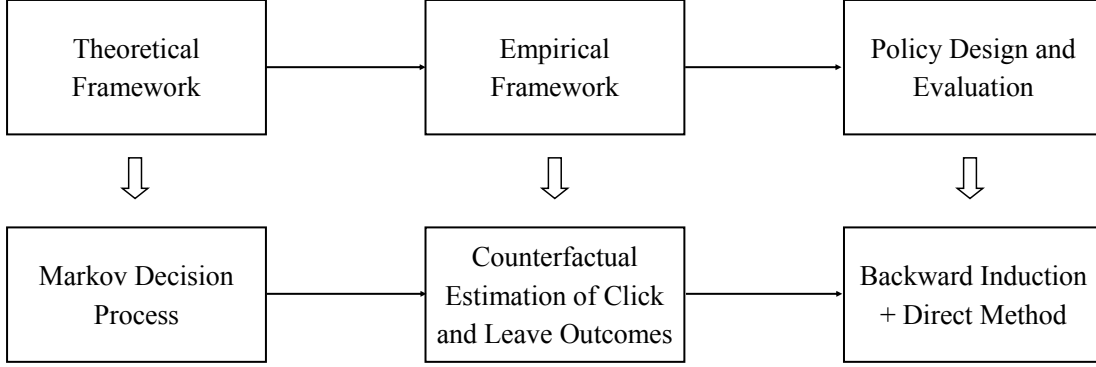
5

Figure 2: An overview of our approach. The top row presents our general framework and the bottom row shows the specific approach we take in this paper.

identify key empirical tasks required for the policy design and evaluation. The bottom row follows the same flow and illustrates the specifics of our approach to the ad sequencing problem in mobile in-app advertising:

Theoretical Framework: We start with a Markov Decision Process (MDP henceforth) that characterizes the structure of adaptive ad interventions. Starting from a theoretical framework allows us to address all the challenges discussed above. First, to capture the dynamics of the problem, we specify a domain-specific MDP with a rich set of state variables that incorporates the inter-temporal trade-offs identified in the literature. Our MDP characterizes the reward at any time period as well as how the state evolves in future periods, given any action taken by the publisher. Since our goal is to optimize the number of clicks per session, we define the reward as the expected probability of click on any ad selected. This probability is also informative of the available history in the next period, and thereby captures a probabilistic component of state transitions in our MDP. Another probabilistic factor that affects the future state is the expected probability of the user leaving the session after an intervention. It determines with what probability the user will be available to receive the next intervention. Together, these two probabilistic outcomes help us define the empirical tasks required in our problem: personalized counterfactual estimation of click and leave outcomes.

Empirical Framework: To obtain accurate counterfactual estimates for the click and leave outcomes, we use the filtering strategy proposed by Rafieian and Yoganarasimhan (2020b) that filters out the ads that *could have never been shown* in a session. This is because we cannot estimate the outcome reliably for these ads as they are not within the joint distribution of the training set that is used for model fitting. However, if an ad *could have been shown* within a session (i.e., has a non-zero propensity score), our outcome estimates will be accurate for this ad in this session as long as we control for all other covariates affecting the propensity of seeing this particular ad.

Further, to personalize these counterfactual estimates, we employ machine learning methods that can capture more complex relationships between the covariates and the outcome. In particular, we use the Extreme Gradient Boosting (XGBoost henceforth) method developed by Chen and Guestrin (2016), which is a fast and scalable version of Boosted Regression Trees (Friedman, 2001).

Policy Design and Evaluation: In order to develop the optimal dynamic policy, we can use our counterfactual estimates for click and leave probabilities and numerically solve for the value functions, given the state variables. In our case, we employ a non-stationary finite horizon MDP and determine the optimal dynamic policy using backward induction. Finally, we use a direct method evaluation method to evaluate the performance of any policy at the session level. This method uses the outcome estimates for both click and leave to simulate the stochasticity within the session and evaluate the outcome/outcomes of interest.

1.4 Findings and Contribution

We apply our framework on the data from a leading mobile in-app ad-network of a large Asian country. Our setting has some notable features that make it amenable to our research goals. First, the ad-network uses a short-lived ad format where ad interventions last for a short period of time and change within the session. Second, the extent of randomization in ad allocation is quite high because of two reasons: (1) the ad-network runs a quasi-proportional auction that employs a probabilistic allocation rule, and (2) the ad-network only allows limited targeting on broad categories. These two features help us estimate click and leave outcomes for a wide range of counterfactual ads. This task would not have been possible if the ad-network had run a deterministic mechanism such as the second-price auction or allowed micro-level targeting.

We first discuss the results from our first-stage machine learning models for both click and leave outcomes. We evaluate these predictive models on a hold-out test set, using various goodness-of-fit measures. We show that both the click and leave models achieve high out-of-sample predictive accuracy. We then show that dropping variables that are defined either at the ad- or session-level from the predictive model leads to a significant drop in the performance of both click and leave models. These results provide preliminary evidence for the gains from the dynamic framework since the publisher can actively influence these variables by adopting an adaptive forward-looking policy.

Next, we focus on the main goal in this paper – evaluating the gains from *adaptive forward-looking sequencing policy*. This policy incorporates all three pieces of information in Figure 1. To establish the performance of our adaptive forward-looking sequencing policy and identify where these gains come from, we define a set of benchmark policies that use a combination of different pieces of information:

- *Random sequencing policy*: This policy allocates exposures to ads randomly. It serves as a baseline for capturing how well we can do without any model.

- *Non-adaptive single-ad policy*: This policy uses the pre-session information to allocate ads within the session. Since this information does not change within the session, the optimal allocation based on only this information is to select a single ad that maximizes publisher's rewards. As such, this policy simulates the case where the ad slot is fixed throughout the session and helps us identify the opportunity costs of using a fixed ad slot.

- *Adaptive myopic sequencing policy*: This policy uses both pre-session and session-level information and allocates each impression to the ad with the highest probability of click in that impression, regardless of how it affects the expected future rewards.

We find that all model-based policies lead to substantial gains compared to the random sequencing policy, in terms of the expected number of clicks per session. In particular, the adaptive forward-looking sequencing policy increases the expected number of clicks by 80.36% relative to the random sequencing policy. We then show that the expected number of clicks is 7.87% higher under the adaptive forward-looking sequencing policy as compared to the non-adaptive single-ad policy. This finding demonstrates the opportunity cost of using a fixed ad slot throughout the session, which supports the current industry trend of using short-lived ad slots. Finally, we show that the adaptive forward-looking sequencing policy results in 1.50% increase in the expected number of clicks per session, compared to the adaptive myopic sequencing policy. This suggests that choosing the best match at any point will not necessarily create the best match outcome at the end of the session. Rather, the right action sometimes is to show the ad that is not necessarily the best match at the moment but transitions the session to a better state in the future. Together, these findings establish the benefits of adopting an adaptive and forward-looking approach to allocate ads. This has important implications for publishers and ad-networks, especially since the current practice in the industry overlooks the dynamics of ad sequencing.

Next, we explore the heterogeneity in the gains from the adaptive forward-looking sequencing policy compared to both non-adaptive single-ad and adaptive myopic sequencing policies, across the sessions. We examine how the relative gains from the adaptive forward-looking sequencing changes, as the number of prior sessions a user has participated in increases. On the one hand, we expect more data on the user to benefit the adaptive forward-looking sequencing policy more than other policies, as the publisher can take prior usage patterns of the user into account to improve the adaptive forward-looking sequencing policy. On the other hand, sequencing effects are shown to become smaller as the user becomes more experienced (Rafieian and Yoganarasimhan, 2020a). We find evidence for the latter: the relative gains from adopting adaptive forward-looking sequencing

policy diminishes as the user participates in more sessions. We further provide some descriptive evidence that the source for these diminishing returns seems to be the number of distinct ads the user has seen: if a user has seen many distinct ads over time, it is harder to affect their decision through adaptive forward-looking sequencing of ads.

Finally, we present some descriptive analysis to better understand how and why the adaptive forward-looking and adaptive myopic sequencing policies differ. We find that the adaptive forward-looking sequencing policy tends to repeat the same ad in consecutive exposures more than the adaptive myopic sequencing policy. This is likely because repeating an ad consecutively is not the optimal decision if the publisher only takes the reward at that moment into account. However, it is the right decision if the publisher is forward-looking, as it increases the number of prior exposures for an ad and strengthens the session-level carryover effects.

In sum, our paper makes three contributions to the literature. First, from a methodological standpoint, we propose a unified dynamic framework that theoretically characterizes the domain structure of the mobile in-app advertising environment, and an empirical approach that allows us to break the problem into composite machine learning tasks. To our knowledge, this is the first paper to collectively incorporate temporal effects of advertising documented in the literature at the advertiser level, and propose a dynamic framework for the sequential allocation that characterizes optimal policy design for publishers. The generality of our framework makes it applicable to the contexts where advertisers have to make dynamic decisions such as the attribution problem. Second, from a substantive point-of-view, we establish the gains from an adaptive forward-looking sequencing policy as compared to other benchmarks that are often used in research and practice. This finding is of importance, as the current practice in this industry ignores the dynamics of ad allocation problem. Third, from a managerial perspective, we quantify the opportunity costs of using fixed ad slots. Our framework can help marketing practitioners and ad-networks to design the ad slot that is optimal in their context.

2 Related Literature

Our paper relates and contributes to several streams of literature.

First, our paper relates to the marketing literature on personalization and targeting in digital platforms. Early papers in this stream build Bayesian frameworks that exploit behavioral data and personalize marketing mix variables (Ansari and Mela, 2003; Manchanda et al., 2006; Arora and Henderson, 2007). Other recent papers in this area use a combination of feature generation and supervised machine learning frameworks to provide more scalable solutions for personalization and targeting policies for large-scale data (Yoganarasimhan, 2018; Rafieian and Yoganarasimhan, 2020b). While all these papers focus on prescriptive or substantive frameworks to study personalization,

they all study this phenomenon from a static point-of-view. Our paper extends this literature by offering a scalable and dynamic framework to develop personalized targeting policies.

Second, our paper relates to the literature on the temporal effects of advertising, such as spillover effects in search advertising (Rutz and Bucklin, 2011; Sahni, 2016), carryover effects in display advertising (Johnson et al., 2016a), temporal interactions between multiple advertising channels (Li and Kannan, 2014), effects of temporal spacing in search advertising Sahni (2015b), and the effects of variety of previous ads in mobile in-app advertising context (Rafieian and Yoganarasimhan, 2020a). While these papers establish the presence of these temporal effects from advertisers' perspective, they do not address how a publisher can use this information to optimally show ads in sequences. In this paper, we view this problem from a publisher's perspective who wants to maximize users' engagement with ads and develop a dynamic framework that incorporates all possible temporal effects that have been documented in the literature.

Third, our paper relates to the literature on dynamic policy design in digital advertising. Given the complexity of solving a dynamic policy, prior works often simplify the problem to avoid the curse of dimensionality. Urban et al. (2013) focus on 16 cognitive-style segments and combine dynamic programming with a Bayesian framework to infer segment membership. In the context of linear video ads, Kar et al. (2015) incorporate ad-specific leave probability as the only source of inter-temporal trade-off and use a cascade model to obtain a dynamic policy. Using the context of mobile in-app advertising, Sun et al. (2017) theoretically examine the problem of short-lived ads in mobile in-app advertising and theoretically derive the dynamic policy by only focusing on certain aspects of the dynamics in this problem. Closely related to our problem, Theocharous et al. (2015) present a framework to develop and evaluate dynamic policies and study the gains at the advertiser level. Our paper differs from these papers since we are the first to present a dynamic framework that incorporates all established temporal effects of advertising and examines the gains from adopting an adaptive forward-looking sequencing policy, from the publisher's perspective.

Finally, our paper relates to the growing literature on machine learning applications in marketing. The vast majority of papers in this stream focus on prediction tasks and use various supervised or unsupervised learning algorithms to achieve a better predictive accuracy (Toubia et al., 2007; Hauser et al., 2010; Dzyabura and Yoganarasimhan, 2018). A narrower body of works in this area brings machine learning methods to policy design questions. Using a static approach, some recent papers develop machine learning methods to design optimal policies in various contexts such as pricing (Dubé and Misra, 2017), ad placement (Rafieian and Yoganarasimhan, 2020b), and CRM campaigns (Hitsch and Misra, 2018). Incorporating the dynamics of exploration-exploitation trade-off, Schwartz et al. (2017) offer a multi-armed bandit approach in a display advertising context.

Our paper adds to this literature by fully incorporating the dynamics of the ad sequencing problem through an MDP and linking it to smaller machine learning tasks.

3   Setting and Data

3.1   Setting

Our data come from a leading mobile in-app advertising network of a large Asian country that had over 85% of the market share around the time of this study. Figure 3 summarizes most key aspects of the setting. We number the arrows in Figure 3 and explain what each step of the ad allocation process in details below:

1. The ad-network designs an auction to sell ad slots. In our setting, the ad-network runs a quasi-proportional auction with a cost-per-click payment scheme. As such, for a given ad slot and a set of participating ads $\mathcal{A}$ with a bidding profile $(b_1, b_2, \ldots, b_{|\mathcal{A}|})$, the ad slot is allocated to ad a with the following probability:

$$q_a^p(b; z) = \frac{b_a z_a}{\sum_{j \in \mathcal{A}} b_j z_j},\tag{1}$$

   where z_a is ad a's quality score, which is a measure reflecting the profitability of ad a. The ad-network does not customize quality scores across auctions.[3] The payment scheme is cost-per-click and is similar to Google's sponsored search auctions. That is, ads are first ranked based on their product of bid and quality score, and the winning ad pays the minimum amount that guarantees their rank if a click happens on their ad.

2. Advertisers participating in the auction choose: (a) design their banner, (b) specify the areas in which they want to show their ad, and (c) submit their bid. Figure 3 shows an example of auction with four different ads.

3. Whenever a user starts a new session in an app (in Figure 3, we use a messaging app as an example), a new impression is being recognized, and a request is sent to the publisher to run an auction.

4. The auction takes all the participating ads into account and selects the ad probabilistically based on the weights shown in Equation (1). It is worth noting that all the participating ads have the chance to win the ad slot. This is in contrast with more widely used deterministic

---

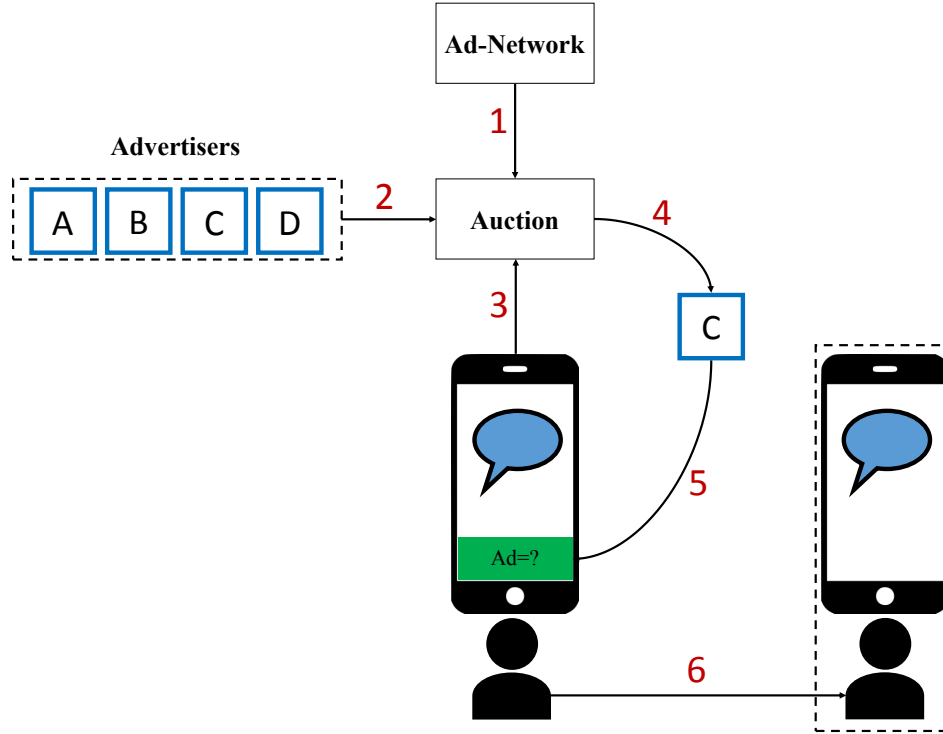[3]In our data collection period, each ad just had one quality score.

Figure 3: A visual schema of our setting

mechanisms like second-price auctions, where the ad with the highest product of bid and quality score always wins the ad slot.

5. The selected ad is placed at the bottom of the app, as shown in Figure 3.

6. Each ad exposure lasts one minute. During this time, the user makes two key decisions: (a) whether to click on the ad, and (b) whether to stay in the app or leave the app and end the session. If the user clicks on the ad, the corresponding advertiser has to pay the amount determined by auction. After one minute, if the user continues using the app, the ad-network treats the continued exposure as a new impression and repeat steps 3 to 6 are repeated until the user leaves the app. We assume that a user has left the app when the time gap until the next exposure exceeds 5 minutes. Consistent with this definition, we define a session as the time interval between the time a user comes to an app and the time she leaves the app.[4]

---

[4]There are obviously various ways to define a session based on the time gap between two consecutive exposures. We show that our results are robust to different definitions.

3.2 Data

We have data on all impressions and clicks for the one month period from 30 September 2015, to 30 October 2015. Overall, we observe 1,594,831,699 impressions along with 14,373,293 clicks in the data, implying a 0.90% CTR. We now describe our raw variables and sampling procedure.

3.2.1 Raw Variables

Each impression contains the following raw information:

- *Time and date:* The exact time-stamp of the impression.
- *App information:* The identifier for a mobile app that shows ads through the ad-network.
- *User information:* An identifier that is unique to each mobile device and serves as our user ID.
- *GPS information:* The exact latitude and longitude of the user at the time of the impression.
- *Targeting variables:* The set of variables that advertisers can target on. There are five main categories that advertisers can target: province, hour of the day, smartphone brand, connectivity type, and Mobile Service Provider (MSP). If an advertiser decides to exclude a certain sub-category within these variables (e.g., Samsung smartphones), his ad will not be shown in impressions in that sub-category.
- *Ad information:* The set of variables related to the ad shown in the impression. It consists of an ad identifier and the potential cost-per-click.[5][6]
- *Click outcome:* A binary variable indicating whether the user clicked on the ad. This is our primary outcome of interest to measure the match between users and ads. While click is generally an imperfect outcome for the effects of ads, it is reasonable to use it as our primary outcome for two reasons. First, all ads are mobile apps whose objective is to get more clicks or installs (e.g., performance ads). Second, the ad-network uses a cost-per-click auction, which directly links the click outcome to the ad-network's revenues.

3.2.2 Sampling Procedure

Our sampling procedure consists of two essential steps – (1) user sampling, and (2) app sampling. We describe each step in greater details below:

- *User sampling:* Since we want to optimally sequence ads within the session, our optimal intervention depends on users' past history. As such, we only focus on users for whom we can exploit their entire history. The challenge is that there is no variable in our data identifying new users. As illustrated in Figure 4, our approach is to split our data into two parts based on a date (October 22), and keep users who are active in the second part of the data (October 22 to October

---

[5]The potential cost-per-clik is the amount that the ad would have paid if the user had clicked on their ad.
[6]We do not have the data on the banner creatives and its format, i.e., whether it is a jpeg file or an animated gif.
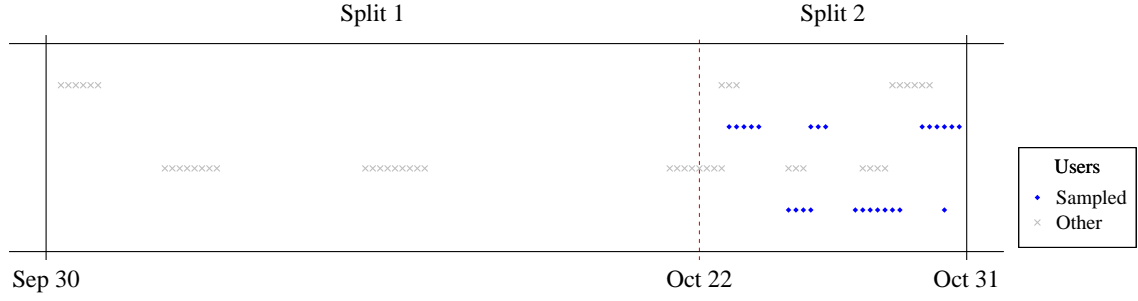
Figure 4: Schema for identification of new users.

30), but not in the first part (September 30 to October 22). This sampling scheme guarantees that the users who are identified as new users have not had any activity in the platform at least for the last three weeks. We drop all the other users from our data.

- *App sampling:* We only focus on the most popular mobile app in the platform, which is a messaging app that has over 30% share of total impressions. As such, we drop new users who do not use this app. There are a few reasons why we focus on this app. First, this is the only app whose identity is known to us. Second, we expect the sequencing effects to be context-dependent, and focusing on one app helps us perform a cleaner analysis. Finally, it takes users a relatively long time to learn how to use certain apps (e.g., games), and learning effects can interact with sequencing effects. However, this messaging app is widely popular in the country and easy to use, so we expect users to pay more attention to ads from the beginning.

Overall, our sampling procedure gives us a total of 8,323,778 impressions shown to a set of 94,884 unique new users. Over 40% of these users use other apps in addition to the messaging app. In our data, there are 6,955,995 impressions shown inside the focal messaging app that corresponds to 1,271,068 unique sessions. For our analysis, we only focus on the impressions shown in the messaging app. However, we use impressions shown in other apps for feature generation.[7]

3.3 Summary Statistics

We now present some summary statistics on the data. As mentioned earlier, these statistics correspond to the impressions shown in the messaging app.

---

[7]Our sampling procedure is almost identical to that of Rafieian and Yoganarasimhan (2020a). However, the number of impressions and sessions is slightly different, because we need to drop users with missing information on the latitude and longitude. Rafieian and Yoganarasimhan (2020a) use those impressions because latitude and longitude do not play a role in their analysis.

3.3.1 Shares of Categorical Variables

Since most raw variables presented in §3.2.1 are categorical, we cannot show the mean and standard deviation for them. Instead, we present the number of categories as well as the shares of the top three sub-categories within each variable in Table 1.

Variable	Number of categories	Share of top categories		
		1st	2nd	3rd
Province	31	24.58%	9.54%	7.50%
Hour of the Days	24	8.48%	8.03%	7.25%
Smartphone Brand	7	44.72%	38.07%	10.11%
Connectivity Type	2	50.54%	49.46%	
MSP	3	50.15%	44.07%	5.77%
Ad	327	18.43%	8.03%	7.50%

Table 1: Summary statistics of the categorical variables. This includes the number of categories and the percentage shares for the top sub-categories within each variable.

As shown in Table 1, one province accounts for a quarter of all impressions. The second row shows that there are certain hours of the day with more user activity. We find that these hours are late at the night when users are not at work. Table 1 shows that two major smartphone brands constitute over 80% of all impressions. Finally, we find that ads have different shares: the top three ads account for roughly 35% of all impressions. On the other hand, we observe that most ads have a very small share. This is mostly because these ads ran short campaigns. Later in §4.2.2, we show the full distribution of ad shares and provide a more detailed discussion.

3.3.2 Distribution of Session-Level Outcomes

Our goal in this paper is to examine how much we can improve session-level user engagement with ads through optimal sequencing of ads. As such, the key outcomes are defined at the session level. Figure 5 shows the empirical CDF of two main outcomes of interest in this study – session length, and the total number of clicks made in a session, which is our primary outcome of interest. We measure session length by the number of exposures shown within any session. Figure 5a shows how this outcome varies across sessions. As shown in this figure, around 50% of all sessions end in only two exposures. Further, the empirical CDF in Figure 5a shows that the vast majority of sessions do not last for more than 10 exposures and only a small fraction of them last for 30 or more exposures.

In Figure 5b, we show the empirical CDF for our primary outcome of interest – the total number of clicks per session. As expected, most sessions end with no clicks being made on ads shown within the session, and the percentage of sessions with at least one click amounts to 7.77%. This is a reasonably high percentage in this industry. Interestingly, there are sessions with more than one

(a) Session Length

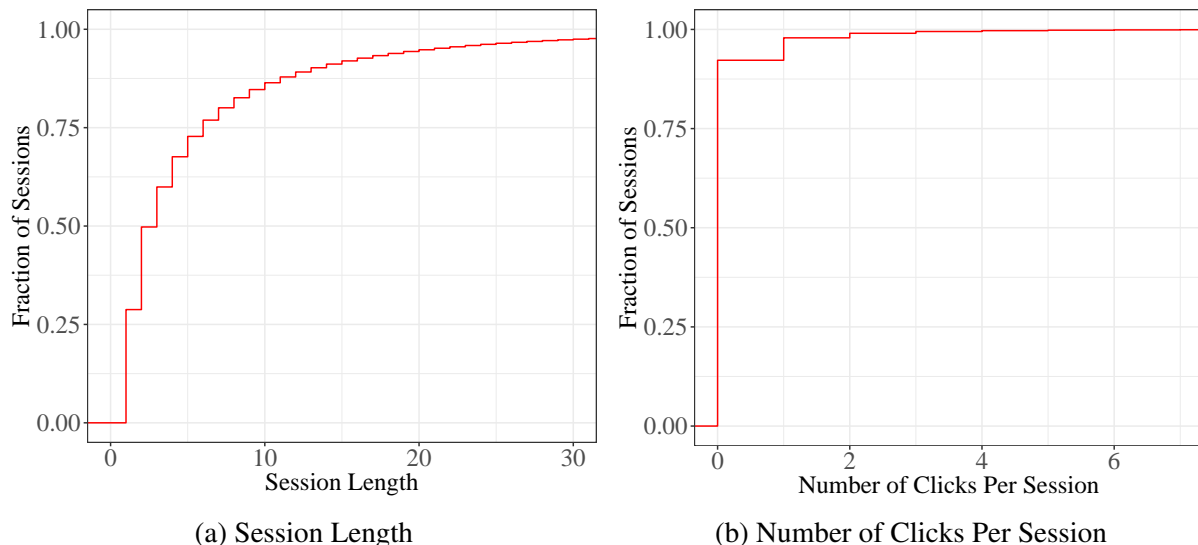(b) Number of Clicks Per Session

Figure 5: Empirical CDF of the session length and total number of clicks per session.

click. Further exploration suggests that these sessions are typically much longer than other sessions, with an average length of over 15 exposures.

3.3.3 Distribution of Micro-Interventions

A central piece of our study is the sequence of ads shown within the session. These sequences are determined by the publisher's ad placement decision at a given exposure. As such, each ad exposure can be treated as a micro-intervention that forms the whole sequence. We focus on three binary micro-interventions at any time period given the history of prior ads shown within the session – (1) *repeat*, (2) *breadth-increasing change*, and (3) *breadth-constant change*. Figure 6 illustrates these micro-interventions. This figure presents a case where the publisher wants to select the fourth ad. We define a micro-intervention as *repeat* if the publisher selects the last ad shown within the session. On the other hand, if the publisher shows any other ad, it is called a *change*, meaning that the current exposure shows a different ad from the last ad. In line with Rafieian and Yoganarasimhan (2020a), we then decompose *change* into two parts – (1) *breadth-constant change*, where the publisher changes the last ad but shows an ad that has been shown before in the session, and (2) *breadth-increasing change*, where the publisher changes the last ad and shows an ad that has not been shown before, thereby increasing the breadth of variety in the session. These three micro-interventions are mutually exclusive, i.e., only one of them takes the value one in any time period.

An important point to notice is that these binary micro-interventions together determine many characteristics of a sequence. For example, if prior interventions in a sequence are mostly breadth-
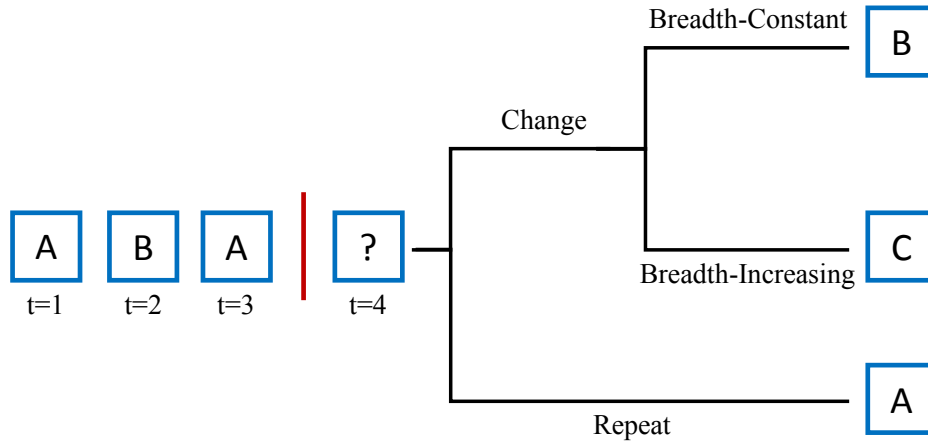
16

Figure 6: An example of three micro-interventions at exposure number 4

increasing changes, we will have a high breadth of variety in that sequence. We present the distribution of these micro-interventions in our data in Figure 7. Each line in this figure represents the percentage of a specific micro-intervention at different points within the session. A few patterns emerge from Figure 7. First, the lines for both *change* and *repeat* are flat over time, illustrating the independence and stability of the auction across different exposures. Further, repeat accounts for only 20% of micro-interventions, which is mostly due to the probabilistic nature of the auction: each ad has a chance of being shown proportional to the product of its bid and quality score. This creates a great degree of randomization in the ad allocation process. Finally, we observe a decreasing pattern in breath-increasing changes. By definition: a random change is less likely to show a completely new ad because we have a finite set of ads competing in a given auction. The fraction of breadth-increasing changes is over 30% across all exposures, indicating the extent of variation in ad allocation within the session.
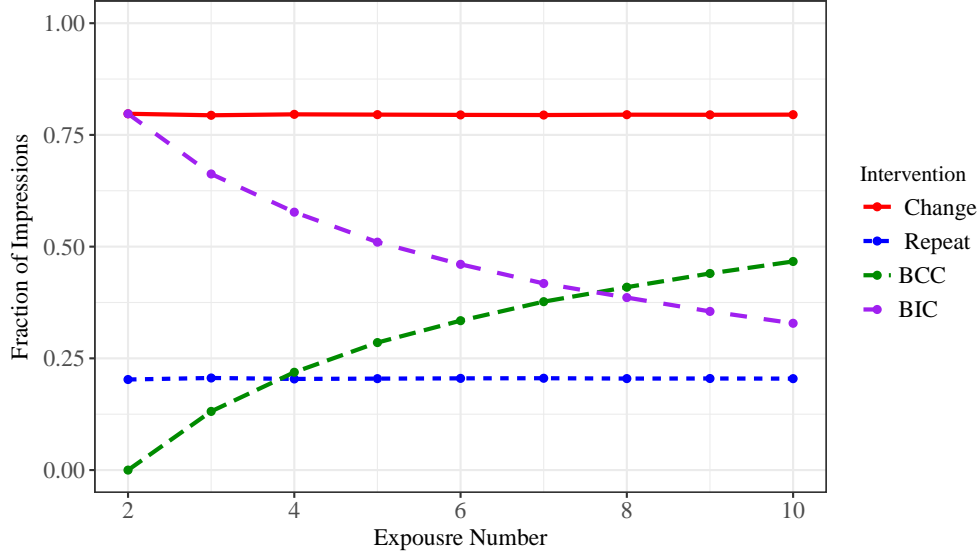
Figure 7: Distribution of micro-interventions at different exposure numbers within the session

4   Dynamic Framework for Sequencing of Ads

We now present our dynamic framework for sequencing of ads. This section proceeds as follows. We start with the motivation for the use of a dynamic framework in our context in §4.1. Next, in §4.2 we define the primitives of our framework and then specify an MDP that incorporates publishers' current and expected future rewards.

4.1   Motivation

The micro-interventions presented in §3.3.3, by and large, depend on the history of the sequence. Further, they determine how a sequence evolves, thereby shaping the history of the sequence for future exposures. Thus, in principle, if the within-sequence history affects which micro-intervention is most effective at any point, it is not clear whether that intervention is the optimal action from a dynamic point-of-view. In fact, the optimal action will be the one with the right balance between how effective it is now and where it transitions the entire sequence. For example, suppose that the publisher wants to fill two impressions with two ads A and B. Now, consider a case where ad A is generally a better ad, so the publisher's optimal decision in both impressions is to allocate them to ad A. However, showing ad A after ad B generates the best overall outcome, as it differentiates ad A and generates a significantly higher click probability. In that sense, the best possible myopic action at a point may not be the optimal decision from a forward-looking perspective.

The main question is whether the within-sequence history affects the outcomes that the publisher cares about. Prior literature has offered some evidence on such effects, including the effects of

multiple exposures and temporal spacing of ads (Sahni, 2015a; Johnson et al., 2016b). In mobile in-app advertising setting, Rafieian and Yoganarasimhan (2020a) show that users' clicking behavior on the next ad depends on the variety of the prior sequence, as measure by the total number of changes and the breadth of variety (sum of all breadth-increasing changes). While they document the positive effects of variety, they also find that the number of prior exposures of an ad within the session positively affects users' clicking behavior. Given these two findings, it is not clear how the publisher should manage these micro-interventions at any given point. On the one hand, more change creates a higher variety in the sequence, thereby increasing the probability of click on the next ad. On the other hand, repeating an ad will increase the number of prior exposures of that ad which is shown to positively affect the probability of click. Thus, the publisher faces various inter-temporal trade-offs when allocating ads within the session.

We address this challenge by developing a dynamic framework that: (1) captures the inter-temporal trade-offs in publishers' ad placement decision in the session, and (2) uses both pre-session and adaptive session-level information to personalize the sequence of ads for the user in any given session. Our framework incorporates both how an action affects the outcome in the current period, as well as the externalities that influence future exposures. In the next sections, we present the details of our framework.

4.2 Model Setup

We specify an MDP that captures the inter-temporal trade-offs in publishers' decision problem by taking into account both current and expected future rewards. An MDP is characterized by a set of primitives that serve as inputs into the objective function that the decision-maker seeks to maximize. We present a generic definition of these primitives below and discuss the specifics of each in the next sections.

1. Time Period (t): The first component we need to specify is the time unit. Since exposures are shown sequentially in our case, we treat each ad exposure as a time period wherein the publisher needs to decide on her actions. $t = 1$ indicates the first exposure in a session.

2. State Space ($\mathcal{S}$): The state space consists of all the information the publisher has about an exposure, which affects her decision-making process.

3. Action Space ($\mathcal{A}$): The action space contains the set of actions the publisher can take. In our case, this action is to show one ad from the ad inventory every time an impression is recognized. As such, $\mathcal{A}$ is the full ad inventory in our problem.

4. Transition Function (P): This function determines how the current state transitions to the

future state given the action made at that point. As such, we can define $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ as a stochastic function that calculates the probability $P(s' \mid s, a)$ where $s, s' \in \mathcal{S}$ and $a \in \mathcal{A}$. Note that this is a crucial component of an MDP since publishers cannot control the dynamics of the problem if the next state is not affected by the current decision. In §4.2.3, we discuss the components of the transition function in our problem in detail.

5. Reward Function (R): This function determines the reward for any action a at any state s. As such, we can define this function as $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$. This function can take different forms depending on the publisher's objective. In our case, since the publisher is interested in optimizing user engagement, she can use different metrics that reflect user engagement such as the probability that the user clicks on the ad. In §4.2.4, we discuss our choice of reward function in greater details.

6. Discount Factor (β): The rate at which the publisher discounts the expected future rewards. In other words, it is the weight that the publisher assigns to the future relative to the current period.

With all these primitives defined, we can now write the publisher's maximization problem as follows:

$$\operatorname*{argmax}_{a} \left[R(s, a) + \beta \mathbb{E}_{s'|s,a} V(s') \right], \tag{2}$$

where $V(s')$ is the value function incorporating expected future rewards at state s' if the publisher selects ads optimally. Following Bellman (1966), we can write this value function for any state $s \in \mathcal{S}$ as follows:

$$V(s) = \max_{a} R(s, a) + \beta \mathbb{E}_{s'|s,a} V(s') \tag{3}$$

As shown in Equation (2), the optimization problem consists of two key elements – the current period reward and the expected future rewards. The publisher chooses the ad that maximizes the sum of these two elements.

4.2.1 State Variables

The state variables contain all the information that the publisher can use for any given exposure in a session. As discussed earlier, the publisher can take two pieces of information into account: (1) pre-session information, and (2) session-level information. Pre-session information contains any data on the user up until the current session, including his demographic variables and behavioral history. For any session i, we denote the pre-session state variables by X_i. It is important to notice that the pre-session variables are not adaptive, i.e., it does not change within the session and hence not have t subscript.

20

On the other hand, session-level variables are adaptive and change within the session. At any given point in a session, this information captures the information about the prior sequence of ads shown to the user as well as user's actions after seeing each ad. Let $G_{i,t}$ denote the session-level state variables. We can write:

$$G_{i,t} = \langle A_{i,1}, Y_{i,1}, A_{i,2}, Y_{i,2}, \ldots, A_{i,t-1}, Y_{i,t-1} \rangle, \tag{4}$$

where $A_{i,s}$ denotes the ad shown in exposure number s and $Y_{i,s}$ denotes whether the user clicked on this ad. As a result, $G_{i,t}$ is the sequence of all ads and actions within the session up to the current time period. Overall, we define the state variables as $S_{i,t} = \langle X_i, G_{i,t} \rangle$, i.e., a combination of both pre-session and session-level variables.

4.2.2 Action Space: Ad Inventory

As mentioned in §4.2, the publisher's action space in our case is the ad inventory. At each point of time, the publisher chooses one ad from the inventory to show to the user. We only focus on the top 15 ads in the ad inventory due to four key reasons. First, as shown in Figure 8, the top 15 ads generate over 70% of all impressions in the focal messenger app. So they collectively account for a significant portion of our observed data. Second, these top ads are more stable in terms of their budget. Since we want to run counterfactual policies, it is important to make sure that advertisers' budgets do not cause any problem for the reliability of our results. Third, we have more data for these ads as these are shown in a wide variety of state variables. This makes our estimates for their outcomes more accurate and less noisy. Finally, limiting the action space makes the problem computationally more tractable when we want to numerically solve for the optimal policy.

4.2.3 Transition Function

We now characterize the law-of-motion, i.e., how state variables transition given the publisher's action at any point. As mentioned earlier, we are interested in the probability of the next state being s', given that action a is taken in state s in the current period, i.e., $P(s' \mid a, s)$. Suppose that the user is in state $S_{i,t} = \langle X_i, G_{i,t} \rangle$ at exposure t in session i. The only time-varying factor in $S_{i,t}$ that can transition is $G_{i,t}$, which is the history of the sequence. Given the definition of $G_{i,t}$ in Equation (4), we can determine the next state if we know user's decision to click on the current ad and/or continue staying in the session. There are three mutually exclusive possibilities for state transitions:

1. *Click and stay:* If the user clicks on ad $A_{i,t}$ and stays in the session, we can define the next state as follows:

$$S_{i,t+1} = \langle X_i, G_{i,t}, A_{i,t}, Y_{i,t} = 1 \rangle, \tag{5}$$
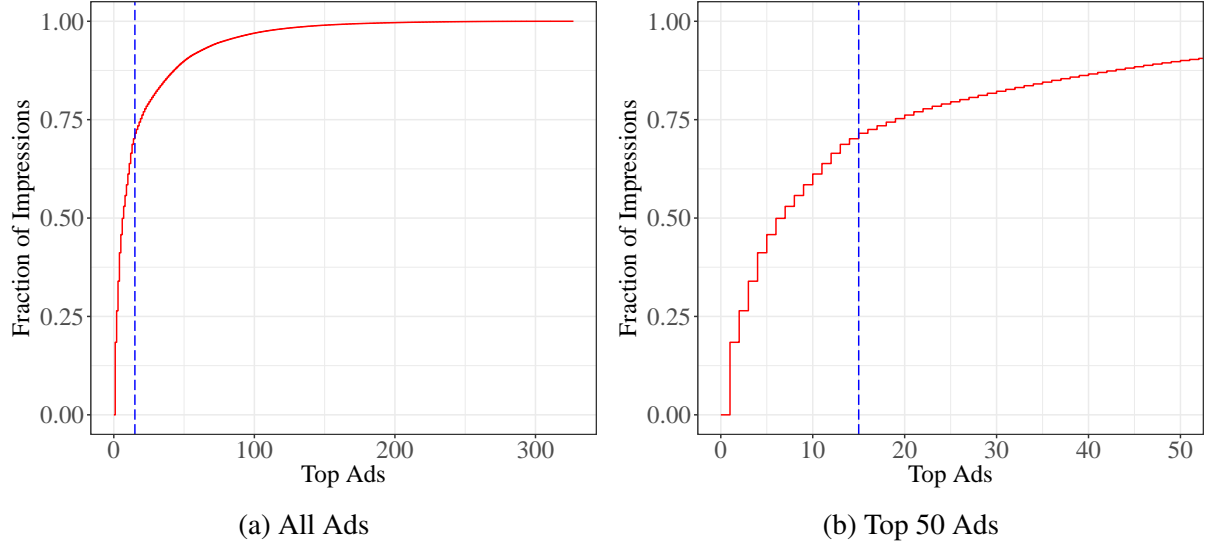
(a) All Ads (b) Top 50 Ads

Figure 8: Cumulative fraction of impressions associated with the top ads. The figure in the left shows the distribution for all 327 ads. The figure in the right zooms in the top 50 ads.

where $Y_{i,t} = 1$ indicates that the user has clicked on the ad shown in exposure number t.

2. *No click and stay:* If the user does not click on ad $A_{i,t}$ and stays in the session, we can similarly define the next state as follows:

$$S_{i,t+1} = \langle X_i, G_{i,t}, A_{i,t}, Y_{i,t} = 0 \rangle, \tag{6}$$

where $Y_{i,t} = 0$ indicates that the user has not clicked on the ad shown in exposure number t.

3. *Leave:* Regardless of user's clicking outcome, if the user decides to leave, the entire session is terminated and there is no more decision to be made. Thus, we can write:

$$S_{i,t+1} = \varnothing \tag{7}$$

Figure 9 visually presents the three possibilities presented above. This figure illustrates an example where the publisher shows an ad in the fourth exposure in a session. It shows three possibilities and how each forms the next state. Based on this characterization, we can now define
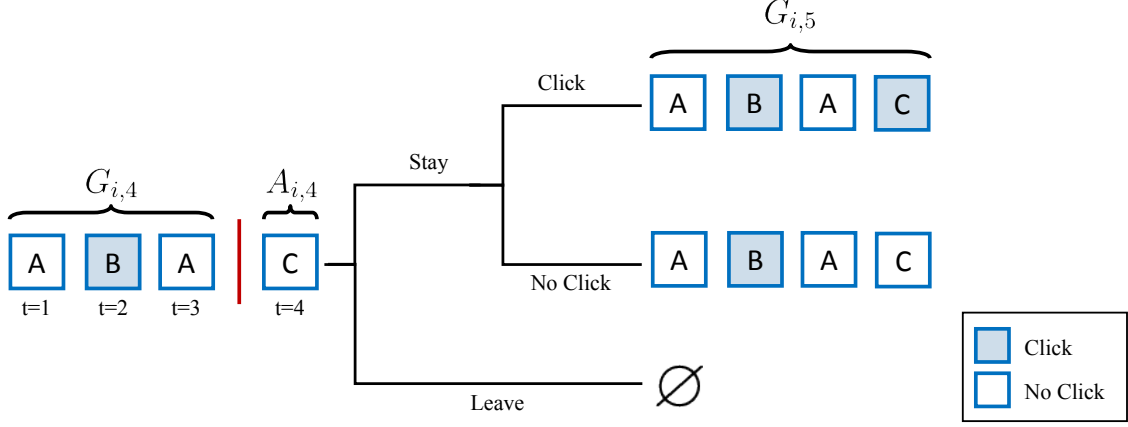
Figure 9: An example illustrating the state transitions.

the transition function for any pair of action and state as follows:

$$P(S_{i,t+1} \mid a, S_{i,t}) = \begin{cases} \big(1 - P(L_{i,t} \mid a, S_{i,t})\big) P(Y_{i,t} \mid a, S_{i,t}) & \text{Equation (5)} \\ \big(1 - P(L_{i,t} \mid a, S_{i,t})\big)\big(1 - P(Y_{i,t} \mid a, S_{i,t})\big) & \text{Equation (6)} \\ P(L_{i,t} \mid a, S_{i,t}) & \text{Equation (7)} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Equation (8) illustrates that the publisher needs to accurately estimate two user-level outcomes given any ad shown – click and leave probabilities. In §5, we discuss our approach to obtain these estimates.

4.2.4   Reward Function

Another piece of an MDP that needs to be defined is the reward function. The reward function can take different forms that vary with the publisher's objective and main outcome of interest. We primarily focus on the number of clicks per session as our main objective. In our case, clicks are particularly good measures of the user engagement with ads because of two reasons. First, all ads in our study are mobile apps that want more clicks and installs. In the literature, this type of ads is called performance ads and their match value is generally assumed to be the probability of click (Arnosti et al., 2016). Second, click is the main source of revenue for the publisher, since the advertiser only pays when a click happens.

There are other advantages in using the click as the main outcome of interest. First, it is a well-recorded outcome that is realized immediately in the data. Second, click is a function of users'

behavior, whereas other outcomes usually involve other players such as advertisers (e.g., publisher's revenues). Thus, the publisher's optimization problem only depends on inferring users' behavior, which is a feasible task in a data-rich environment.

Given that publishers want to maximize the number of clicks made per session, we can define the reward function as the probability of click for a pair of state and action. For exposure number t in session i, we can write:

$$R_t(a; S_{i,t}) = P(Y_{i,t} \mid a, S_{i,t}) \tag{9}$$

This is the probability of click on ad a if shown in the current state.

4.2.5 Discount Factor

The discount factor in Equation (2) reflects the relative importance of the expected future rewards compared to the current period reward from the publisher's perspective. As such, it generally takes a positive value close to one, if the publisher is forward-looking, i.e., they incorporate the expected future rewards in their decision problem. If the discount factor is zero, it means that the publisher is myopic and only cares about the reward in the current period.

In our dynamic framework, the publisher's decision is to select an ad for each impression in a session. The entire session happens in just a few minutes, depending on the number of exposures the user chooses to stay for. Given the short time horizon of the optimization problem, a risk-neutral publisher must value the current and expected future rewards equally, indicating that β is very close to 1.

4.2.6 Policy Definition

We characterize a dynamic policy as a mapping $\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$, that assigns a probability to any action $a \in \mathcal{A}$ taken in any given state $s \in \mathcal{S}$. For a deterministic policy, $\pi(a \mid s)$ will take value one only for one ad for any given state. However, we define the policy as a probability function to allow for non-deterministic policies as well.

In sum, the main goal of our framework is to develop a policy π^* that maximizes the expected reward given the initial set of state variables and the transition function. We later discuss how we develop and evaluate this policy in §6.2 and §6.3, respectively.

5 Empirical Strategy

In this section, we present our empirical strategy to estimate the primitives of the dynamic framework defined in §4. This involves the estimation of probabilistic components of the transition function as characterized in §4.2.3, as well as estimation of unknown components in the reward function as presented in §4.2.4. Together, this gives us two estimands – (1) leave outcome, and (2) click

outcome. The task in both cases is to accurately predict the outcome for a pair of state and action. However, given the broader task of taking the best action at any state that we are interested in, we need to obtain these outcome estimates not only for the ad that is shown in the data but also for the counterfactual ads that are not shown in the data. Below is a formal definition of these two tasks:

Task 1: For any set of state variables observed in the data, we want to accurately estimate the click probability for all ads if shown in that impression. That is:

$$\hat{y}_{i,t}(a; S_{i,t}) = P(Y_{i,t} \mid a, S_{i,t}), \forall a \in \mathcal{A}_i \tag{10}$$

Task 2: For any set of state variables observed in the data, we want to accurately estimate the leave probability for all ads if shown in that impression. That is:

$$\hat{l}_{i,t}(a; S_{i,t}) = P(L_{i,t} \mid a, S_{i,t}), \forall a \in \mathcal{A}_i \tag{11}$$

We discuss our empirical strategy for Tasks 1 and 2 and helps us set the scope of our predictive model in §5.1. We then describe the learning algorithm that we use and its advantages in §5.3. Finally, in §5.4, we present the estimation results on our click and leave probability model.

5.1 Filtering Strategy for Counterfactual Estimation

For the task of outcome prediction like ours, we can use machine learning methods that capture more complex relationships between the covariates and outcomes (Mullainathan and Spiess, 2017). However, these methods are only able to predict the outcome for observations within the joint distribution of the observed data. This implies that if an ad could have never been shown in an observation in the actual data, our outcome estimate for that ad may not be reliable. Thus, if there is no randomization in the ad allocation process, we cannot reliably estimate the outcome for counterfactual ads.

This informs our empirical strategy for counterfactual estimation of click and leave outcomes. While ads are selected through a deterministic allocation rule in most commonly used auctions such as second-price, the quasi-proportional auction in our setting has the advantage of using a probabilistic rule that induces randomization in the ad allocation process. The key issue with a deterministic allocation rule is that the ad that is actually shown in the data is the only ad that *could have been shown*, and other counterfactual ads *could have never been shown*. We argue that if an ad could have been shown in a session (i.e., has non-zero propensity score), this observation could have been generated within the joint distribution of the training data. Thus, we are able to accurately estimate the outcome for these ads in both factual or counterfactual situations.
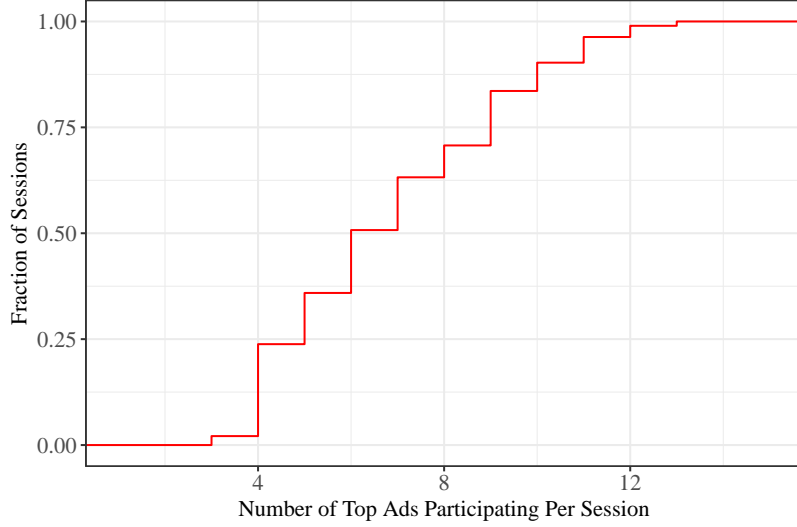
Figure 10: Empirical CDF of the number of ads competing per session.

To reflect this idea, we employ a filtering strategy similar to that in Rafieian and Yoganarasimhan (2020b). Here, for each session i, we identify the set of ads that could have been shown in that session and call it $\mathcal{A}_i$. We filter out the set of ads that are not in $\mathcal{A}_i$. This step sets the scope of our ability to generate counterfactual estimates. Figure 10 shows the empirical CDF of the number of ads participating in the auction for a session. As shown in this figure, the number of ads competing for each impression is quite variable across sessions.

Further, the extent of randomization in our problem allows us to make the unconfoundedness assumption: for any exposure number t in session i, the set of potential outcomes for all ads in $\mathcal{A}_i$ is independent of the actual ad that is shown in that exposure, conditional on the state variables. We can write:

$$\{Y_{i,t}(a)\}_{a \in \mathcal{A}_i} \perp\!\!\!\perp A_{i,t} \mid S_{i,t},$$

where $Y_{i,t}(a)$ is the potential outcome at the exposure t in session i when ad a is being shown. We can even weaken this assumption as only conditioning on demographic variables D_i yields the conditional independence presented above. This is because the ad allocation is random, controlling for the propensities determined by auction. Since only targeting variables can affect propensities in an auction, we only need to control for them. Thus, if we properly control for D_i in our predictive model, our outcome estimates preserve consistent treatment effects.

5.2 Feature Generation

As discussed earlier, our goal is to estimate click and leave outcomes for any combination of ad and state variables, as shown in Equations (10) and (11). A major challenge in estimating these
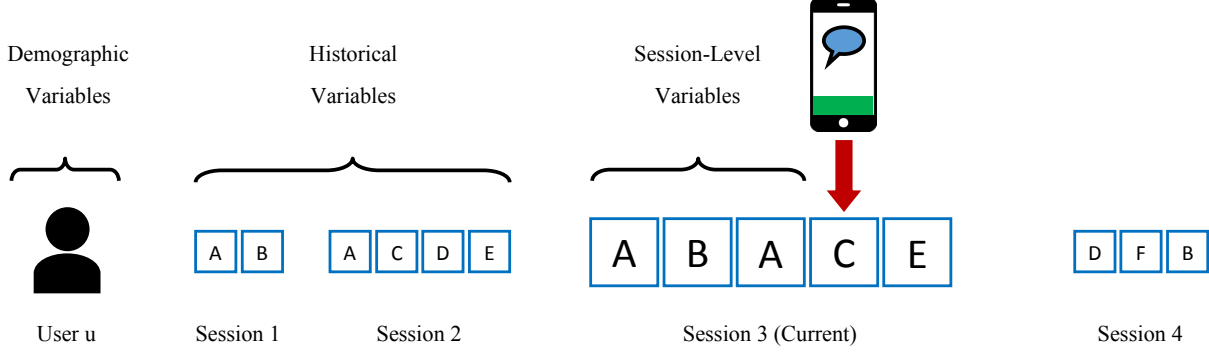
26

Figure 11: A visual schema for our feature generation and categorization.

equations is that the set of inputs is quite large, containing the entire sequence of prior ads shown to the user. In this section, we present a feature generation framework that maps a combination of state variables and ads ($\langle S_{i,t}, a \rangle$) to a set of meaningful features that we can give as inputs to our learning algorithm. Ideally, we need our final set of features to fully represent $\langle S_{i,t}, a \rangle$ in a lower dimension without any information loss. Thus, we generate a set of features that help us predict users' clicking behavior and app usage based on the prior literature on advertising.

We categorize these features into three groups: (1) demographic features, (2) historical features, and (3) session-level features. Demographic and historical features relate to the pre-session state variables (X_i), whereas session-level features relate to the session-level variables ($G_{i,t}$). Figure 11 provides an overview of our feature generation and categorization. In this example, the user is at her fourth exposure in her third session. The features for this particular exposure include the observable demographic features, historical features generated from the prior sessions, and session-level features that are generated from the first three exposures shown in the current session. Clearly, we do not use any information from the future to generate a feature: at any point, we only use the prior history up to that point. In the following sections, we describe all these features in detail.

5.2.1 Demographic Features

This includes the variables that we already observe in our data (see §3.2.1), such as the province, latitude, longitude, smartphone brand, mobile service provider (MSP), and connectivity type. For any session i, we use D_i to denote the set of demographic features. These features do not transition based on the ad that the publisher shows at any time period. As such, we do not use subscript t for

them.[8] We include these features because of two reasons. First, these features help predict both users' clicking behavior and app usage. Second, the targeting variables are the main confounding source, and controlling them guarantees that we control for propensity score of ads when estimating the outcomes.

5.2.2 Historical Features

Historical features reflect the user's past activity prior to the current session. While demographic features are available in the data, we need to generate historical features based on the pre-session information. These features are not adaptive, i.e., they remain constant within the session. We follow the approach in Rafieian and Yoganarasimhan (2020b) to generate these features. We refer the reader to that paper for details on why these features help predict user-level outcomes.

Let i, u, t, p, and a denote the session, user, exposure number, app, and ad respectively. Since we only focus on the top app, using the subscript p indicates that we only use impressions in that app to generate a feature. Otherwise, we calculate the feature using all impression. Below, we present the detailed set of our historical features along with their definition:

- $Imp_{i,u}$: The total number of impressions user u has seen prior to session i.
- $Click_{i,u}$: The total number of clicks user u has made prior to session i.
- $Imp_{i,u,p}$: The total number of impressions user u has seen in the top app prior to session i.
- $Click_{i,u,p}$: The total number of clicks user u has made in the top app prior to session i.
- $Imp_{i,u,t}$: The total number of impressions user u has seen at exposure number t prior to session i.
- $Click_{i,u,t}$: The total number of clicks user u has made at exposure number t prior to session i.
- $Imp_{i,u,a}$: The total number of impressions of ad a that user u has seen prior to session i.
- $Click_{i,u,a}$: The total number of click that user u has made on ad a prior to session i.
- $Space_{i,u,a}$: The time space (in minutes) between session i and the last time ad a is shown to user u prior to session i. It takes value zero if there was no prior exposure of ad a.
- $LastSessionLength_{i,u}$: The length of last session (in number of exposures) that user u was exposed to prior to session i.
- $AvgSessionLength_{i,u}$: The average length of the sessions (in number of exposures) that user u was exposed to prior to session i.
- $LastFreeTime_{i,u}$: The free time (in minutes) user u has had between her last session and session i.

---

[8]One could argue that features such as latitude and longitude may change within the session. While this is possible, it is unlikely to happen as a result of the publisher's ad interventions. Further, the sessions are usually short, and we rarely observe such a change in our data.

- *AvgFreeTime$_{i,u}$*: The average free time (in minutes) user u has had between her sessions prior to session i.
- *Breadth$_{i,u}$*: The total number of distinct ads that user u has seen prior to session i.
- *GiniSimpson$_{i,u}$*: The Gini-Simpson index for ads that user u has seen prior to session i (Simpson, 1949). This metric captures the diversity of prior ad exposures by calculating the probability that two random exposures from the past were of different ads. A higher Gini-Simpson index means that the user has seen a more diverse set of ads. We can write the Gini-Simpson index as follows:

$$GiniSimpson_{i,u} = 1 - \sum_{a \in \mathcal{A}} \frac{Imp_{i,u,a}(Imp_{i,u,a} - 1)}{Imp_{i,u}(Imp_{i,u} - 1)} \tag{12}$$

For a session i, we denote all these features as H_i. Like demographic features, publisher's actions will not change historical features within the session. However, for the next session that the user participates in, the historical features are updated. Further, it is worth noting that three historical features, *Imp$_{i,u,a}$*, *Click$_{i,u,a}$*, and *Space$_{i,u,a}$*, are ad-specific. It means that these features will change when different ads are selected.

5.2.3 Session-Level Features

The session-level features are key to our analysis as we are interested in optimal sequencing of ads within the session. Further, we allow these features to change by the exposure number. That is, depending on the prior exposures within the session, these features will evolve. Below is a list of session-level temporal features:

- *Imp$_{i,t}$*: The total number of impressions the user has seen in session i prior to exposure number t. For any exposure number t, this feature is $t - 1$.
- *Click$_{i,t}$*: The total number of clicks the user has made in session i prior to exposure number t.
- *Imp$_{i,a,t}$*: The total number of impressions of ad a that user has seen in session i prior to exposure number t.
- *Click$_{i,a,t}$*: The total number of clicks that the user has made on ad a in session i prior to exposure number t.
- *Space$_{i,a,t}$*: The number of exposures between exposure number t and the last time ad a was shown in session i. It takes value 0 when there is no prior exposure of ad a in the session.
- *Breadth$_{i,t}$*: The total number of distinct ads that the user has seen within session i prior to exposure number t. We can define this feature as follows:

$$Breadth_{i,t} = \sum_{a \in \mathcal{A}} \mathbb{1}(Imp_{i,a,t} > 0) \tag{13}$$

29

- *Changes$_{it}$*: The total number of consecutive changes of ads prior to the exposure number t within the session i. We can write:

$$Change_{i,t} = \sum_{j=2}^{t-1} \mathbb{1}(A_{i,j} \neq A_{i,j-1}), \tag{14}$$

where $A_{i,j}$ is the ad shown at exposure number j in session i.

- *GiniSimpson$_{i,t}$*: The Gini-Simpson index for the ads shown within session i prior to exposure number t. Following the same logic in Equation (12), we can write:

$$GiniSimpson_{i,t} = 1 - \sum_{a} \frac{Imp_{i,a,t}(Imp_{i,a,t} - 1)}{(t-1)(t-2)} \tag{15}$$

For any session i and exposure number t, we denote all session-level features by $O_{i,t}$. As such, this is the only set of features that has subscript t, indicating that it changes within the session. Therefore, the publisher's actions affect the transition of these features in the session. One could argue that historical features also change within the session as user's history accumulates after each exposure. It is worth noting that we do not update the history within the session because session-level temporal features capture that information. As a result, not updating historical features will not result in any information loss.

5.3 Learning Algorithm

Here we describe the learning algorithm that we use to estimate the click and leave models. Since our goal is to estimate these outcomes as accurately as possible, we need to build a model that is able to capture complex relationships between covariates and outcomes. For both outcomes, we use Extreme Gradient Boosting (XGBoost henceforth) method developed by Chen and Guestrin (2016), which is a fast and scalable version of Boosted Regression Trees (Friedman, 2001). There are some key reasons why we use XGBoost as our main optimization method. First, it has been shown to outperform most existing methods in most prediction contests, especially those related to human decision-making like ours (Chen and Guestrin, 2016). Second, Rafieian and Yoganarasimhan (2020b) show that in the same context, XGBoost achieves the highest predictive accuracy compared to other methods.

There are important implementation details in making an XGBoost model that we summarize as follows. First, following the arguments in Rafieian and Yoganarasimhan (2020b), we use logarithmic loss as our loss function. Second, we split our data into two parts – training and test sets. To tune parameters of XGBoost, we split the training set into two parts and use one as a hold-out validation

Evaluation Metric	Training	Test
RIG	0.2847	0.2768
R^2	0.1899	0.1872
AUC	0.8560	0.8531

Table 2: Performance of the click estimation model on training and test sets

set to prevent the model from over-fitting. Finally, to select the hyper-parameters accurately, we conduct a grid search over a large set of hyper-parameters and select those that give us the best performance on a hold-out validation set. Finally, we evaluate the performance of our model on both training and test set. Note that the test set is not used at any stage to select the model. Thus, the evaluation on the test set demonstrates an out-of-sample predictive accuracy.

5.4 Results

5.4.1 Results from the Click Estimation Model

We use three different evaluation metrics to evaluate the predictive performance of our model:

- *Relative Information Gain (RIG):* This is defined based on the log loss, which is our loss function in the XGBoost model. This metric reflects the percentage improvement in log loss compared to a baseline model that simply predicts the average CTR for all impressions.
- *R-Squared (R^2):* This is the most commonly used metric in marketing and economics, and intuitively calculates the percentage of variance in the outcome that our model can explain.
- *Area Under the Curve (AUC):* It determines how well we can identify *true positives* without identifying *false positives*. This score ranges from 0 to 1 and a higher score indicating better performance.

The results of these three metrics are shown in Table 2. We find that our model achieves over 28.47% and 27.68% RIG on training and test sets, respectively. This predictive accuracy is quite substantial compared to the literature (Yi et al., 2013; Rafieian and Yoganarasimhan, 2020b). Further, the results on R^2 also indicate that our model reaches an excellent predictive performance. Given the inherent noise and variability in clicks, explaining over 18% of the variance in this outcome requires a very powerful model. Finally, the results on the last metric document a very good classification power.

5.4.2 Results from the Leave Estimation Model

We now discuss the results from the leave model. The leave outcome is an important piece of the transition function, since the leave outcome by the user terminates the entire session. We use the same evaluation metrics to evaluate the predictive performance of our leave estimation model. Table 3 summarizes the results in terms of these metrics on both the training and test sets. We find that

31

Evaluation Metric	Training	Test
RIG	0.0975	0.0949
R^2	0.0942	0.0918
AUC	0.7164	0.7135

Table 3: Performance of the leave estimation model on training and test sets

Model	Click Model			Leave Model		
	RIG	R^2	AUC	RIG	R^2	AUC
Full model	0.2768	0.1872	0.8531	0.0949	0.0918	0.7135
Without session-level	0.2245	0.1280	0.8334	0.0881	0.0889	0.7055

Table 4: Performance of models with/without session-level features

our leave estimation model explains over 9% of the variance in the leave outcome on both training and test sets.

Overall, it is worth noting that we should not expect the leave model to to have high predictive ability for a variety of reasons. First, clicking decision is mainly related to the ad shown, whereas leave decision is somewhat independent of ad exposures, especially in the messenger app where people's usage stems from users' messaging behavior that is unobserved to the researcher. Second, we focus on new users for whom we do not have a long panel. Hence, our features reflecting their past usage may not be very informative.

5.4.3 Preliminary Evidence on the Gains from the Adaptive Framework

An adaptive framework allows the publisher to use real-time session-level information to make decisions. Intuitively, the publisher can benefit from an adaptive framework for decision-making if session-level features play an important role in driving the main outcome of interest. In other words, if the user's decision to click on ads is only a function of his demographic or historical features and does not change based on session-level features, using an adaptive framework becomes ineffective. Thus, an intuitive test that can provide some preliminary evidence on the effectiveness of an adaptive framework is to drop session-level features from the main click and leave models and see whether it results in a drop in the predictive performance of these models.

The model without session-level features exclude the following features: $Imp_{i,t}$, $Click_{i,t}$, $Breadth_{i,t}$, $Changes_{i,t}$, $GiniSimpson_{i,t}$, $Imp_{i,a,t}$, $Click_{i,a,t}$, $Space_{i,a,t}$. We compare the performance of this model with the full model for both click and leave estimation models. The results on the performance of these models on the same test set are presented in Table 4. As shown in this table, the performance of the model drops when we exclude session-level features. This finding suggests and publishers can benefit from using adaptive frameworks for ad sequencing.

6 Ad Sequencing Policies

As discussed earlier, our main goal is to use our dynamic framework and develop a sequencing policy that maximizes user engagement in the session. We primarily measure user engagement by the number of clicks per session. To develop sequencing policies, we can use our estimates for the transition and reward functions to solve for the optimal policy in the MDP specified in §4.

In this section, we first present a series of other baseline policies that are commonly used in practice in §6.1. It allows us to compare the performance of the adaptive forward-looking policy developed by our framework with other benchmarks and establish the gains from adopting a dynamic framework. We then discuss the empirical evaluation of all these policies in §6.2. Finally, in §6.4, we present our results on different sequencing policies, explore heterogeneity in gains from the dynamic sequencing policy across sessions, and offer some explanation for the performance of different methods.

6.1 Definition of Sequencing Policies

We present a series of benchmark policies to examine the publisher's gains from adopting a dynamic framework for ad sequencing. We consider benchmark policies that drop modeling components in our dynamic framework and/or reflect the current norm in research and practice. As such, our comparison allows us to pin down how valuable each modeling component is and how much we can improve over the current practice. Starting with the adaptive forward-looking policy, which is based on our dynamic framework, we present the list of policies that we consider as follows:

- **Adaptive forward-looking sequencing policy:** This sequencing policy uses the dynamic framework in §4 with the expected probability of click being the reward function. Using our generic formulation of an MDP in Equation (2), we can write the publisher's optimization problem specific for a click-maximizing objective as follows:

$$\operatorname*{argmax}_{a \in \mathcal{A}_i} \left[R_t(a; S_{i,t}) + \beta \mathbb{E}_{S_{i,t+1}|S_{i,t},a} V_{t+1}(S_{i,t+1}) \right], \tag{16}$$

where the value function for any state variable at exposure number t can be written as follows:

$$V_t(S_{i,t}) = \max_{a \in \mathcal{A}_i} \left[R_t(a; S_{i,t}) + \beta \mathbb{E}_{S_{i,t+1}|S_{i,t},a} V_{t+1}(S_{i,t+1}) \right] \tag{17}$$

This policy is both adaptive and forward-looking, i.e., it takes the session-level information (adaptive) and future information (forward-looking) into account. We also call this policy fully dynamic sequencing policy.

- **Adaptive myopic sequencing policy:** This sequencing policy does not take into account the

expected future rewards when making the decision at any point. This is equivalent to the adaptive forward-looking sequencing with $\beta = 0$ that turns off the weight on the future rewards. Thus, we can write the objective function for adaptive myopic sequencing as follows:

$$\underset{a \in \mathcal{A}_i}{\operatorname{argmax}} R_t(a; S_{i,t}) \tag{18}$$

In this policy, the publisher selects the ad that maximizes CTR in the current period. It is worth noting that this policy is adaptive, as it uses the session-level information that is time-varying. However, it is myopic in the sense that it ignores future information. This case reflects the common practice of using contextual bandits in the industry.

- **Non-adaptive single-ad policy:** This policy only uses the pre-session information. Since it does not use adaptive information, this policy allocates all the impressions to a single ad that has the highest average CTR. This is similar to the practice of using a fixed ad slot where the whole session is allocated to one ad. The objective in this case is the same as Equation (18) only for $t = 1$.[9]

  This policy provides some insight into the ad sequencing problem because it has two distinct features. First, it captures the potential gains from using a short-lived ad slot as compared to the fixed ad slot. Second, it demonstrates the value of adaptive session-level information.

- **Random sequencing policy:** In this sequencing policy, the publisher randomly selects ads from the ad inventory. While this is a naive policy, it can serve as a benchmark showing how well we can do without any model.

6.2 Empirical Evaluation

We now explain how we use our empirical estimates for both click and leave outcomes to develop and evaluate all the policies presented in §6.1. We start with the solution concept for the baseline policies in §6.2.1. These policies are more straightforward and easier to derive. We then present the solution concept for the adaptive forward-looking sequencing policy in §6.2.2. We finally discuss how we can evaluate the performance of these policies.

6.2.1 Solution Concept for Baseline Policies

We now describe how we obtain the baseline policies described in §6.1: (1) adaptive myopic sequencing policy, (2) non-adaptive single-ad policy, and (3) random sequencing policy. Since none

---

[9]One could argue that the optimal single-ad that is selected for the entire session may be different from the optimal ad for the first exposure. We acknowledge this issue and check the robustness of our results by using a dynamic optimization constrained by a single ad to be shown for the entire session. In the main text, however, we use the more straightforward approach of allocating the entire session to the ad with the highest CTR in the first exposure.

of these policies incorporate expected future rewards, we need to mainly rely on our estimates for the reward function with the click-maximizing objective, i.e., $\hat{y}_{i,t}(a; S_{i,t})$. We present our empirical approach to solve for these policies below:

- **Adaptive myopic sequencing policy:** As presented in Equation (18), this policy selects the ad with the highest CTR at any point. Therefore, we can write $a_t^m(S_{i,t}) = \operatorname{argmax}_{a \in \mathcal{A}_i} \hat{y}_{i,t}(a; S_{i,t})$, where $a_t^m(S_{i,t})$ indicates the ad to be shown under adaptive myopic sequencing in state $S_{i,t}$. We can characterize the best policy under adaptive myopic sequencing as follows:

$$\hat{\pi}^m(a \mid S_{i,t}) = \begin{cases} 1 & a = a_t^m(S_{i,t}) \\ 0 & a \neq a_t^m(S_{i,t}) \end{cases} \tag{19}$$

- **Non-adaptive single-ad policy:** This case is the solution for the adaptive myopic sequencing policy at $t = 1$. As such, we have $a_t^s(S_{i,t}) = a_t^m(S_{i,1})$ for any t in session i. Thus, the non-adaptive single-ad policy for each session i can be characterized as follows:

$$\hat{\pi}^s(a \mid S_{i,t}) = \begin{cases} 1 & a = a_t^s(S_{i,t}) \\ 0 & a \neq a_t^s(S_{i,t}) \end{cases} \tag{20}$$

- **Random sequencing policy:** This policy basically gives the same chance to all the ads in the inventory. Thus, we can write the random sequencing policy as follows:

$$\hat{\pi}^r(a \mid S_{i,t}) = \begin{cases} \frac{1}{|\mathcal{A}_i|} & a \in \mathcal{A}_i \\ 0 & a \notin \mathcal{A}_i \end{cases}, \tag{21}$$

where $\mathcal{A}_i$ is the set of ads competing in session i.

6.2.2 Solution Concept for the Adaptive Forward-looking Sequencing Policy

Solving for the best policy in an MDP can become a daunting task when the state space is high dimensional. Given the set of features we use for our estimation tasks, we need to store the full history within the session. For example, to update session-level features such as $Click_{i,a,t}$ or $Breadth_{i,t}$, we need to know the entire sequence of actions and outcomes within the session up until exposure number t, which can be computationally burdensome for large t.

The session-level history that we need to store for the transition function contains the sequence of all ads and the click outcomes. That is, for exposure number t in session i, this history is defined as $G_{i,t} = \langle A_{i,1}, Y_{i,1}, ..., A_{i,t-1}, Y_{i,t-1} \rangle$. This set can take $(2|\mathcal{A}|)^{t-1}$ unique values for each t, indicating

that it grows exponentially in t. Thus, one way to resolve this problem is to consider a finite horizon case with a reasonable T that is sufficiently large to let us exploit the dynamic framework, while reasonably small to help us avoid computational issues. We argue that that $T = 6$ satisfies both these conditions since the user is quite likely not to get to the seventh exposure at all given the results we show in Figure 5a: around 75% of the sessions have shown at most six exposures. Thus, optimizing the exposure after that point has only a marginal effect on the performance of the click-maximizing policy.

To empirically derive the adaptive forward-looking sequencing policy, we need two key estimands – $\hat{y}_{i,t}(a; S_{i,t})$ and $\hat{l}_{i,t}(a; S_{i,t})$. The former affects both the reward function and state transitions, whereas the latter only affects the state transitions. For notational convenience and brevity, let $\tilde{V}_t(a, S_{i,t})$ denote our estimate for the sum of both current period reward and expected future rewards given action a being taken in state $S_{i,t}$. Namely, it is the estimated objective in Equation (16). Using our empirical estimates of expected click and leave probability, we can write $\tilde{V}_t(a, S_{i,t})$ as follows:

$$
\begin{aligned}
\tilde{V}_t(a, S_{i,t}) =& \hat{y}_{i,t}(a; S_{i,t}) + \left(1 - \hat{l}_{i,t}(a; S_{i,t})\right) \hat{y}_{i,t}(a; S_{i,t}) V_{t+1}\left(\langle S_{i,t}, a, Y_{i,t} = 1\rangle\right) \\
& + \left(1 - \hat{l}_{i,t}(a; S_{i,t})\right) (1 - \hat{y}_{i,t}(a; S_{i,t})) V_{t+1}\left(\langle S_{i,t}, a, Y_{i,t} = 0\rangle\right),
\end{aligned}
\tag{22}
$$

where $\langle S_{i,t}, a, Y_{i,t} = 1\rangle$ and $\langle S_{i,t}, a, Y_{i,t} = 0\rangle$ denote the state variables in the next period. The equation above shows that we can easily break the expected future rewards into two deterministic pieces. We now propose a backward induction solution concept for this case as described below:

1. We start from the last period, T. We assume that this is the last exposure number within the session. Hence, the problem is static in that period. We can write:

$$
\hat{V}_T(S_{i,T}) = \max_{a \in \mathcal{A}_i} \hat{y}_{i,T}(a; S_{i,T})
\tag{23}
$$

$$
\hat{a}_T^d(S_{i,T}) = \operatorname*{argmax}_{a \in \mathcal{A}_i} \hat{y}_{i,T}(a; S_{i,T})
\tag{24}
$$

That is the maximum value the publisher can extract from this particular state variable in the last time period.

2. For any $t < T$, we can write:

$$
\hat{V}_t(S_{i,t}) = \max_{a \in \mathcal{A}_i} \tilde{V}_{i,t}(a, S_{i,T})
\tag{25}
$$

$$
\hat{a}_t^d(S_{i,t}) = \operatorname*{argmax}_{a \in \mathcal{A}_i} \tilde{V}_{i,t}(a, S_{i,T})
\tag{26}
$$

Now, if we go backward and solve for the value functions, everything on the right-hand side of Equation (25) is known because we have already solved for the value functions in the next period. Therefore, we can find the value function for all the states in time period and continue this process until exposure number 1.

Once we have estimated the value function and best ad for all the states, we can simply characterize the adaptive forward-looking sequencing policy as follows:

$$\pi^d(a \mid S_{i,t}) = \begin{cases} 1 & a = a_t^d(S_{i,t}) \\ 0 & a \neq a_t^d(S_{i,t}) \end{cases} \tag{27}$$

We use backward induction as our main approach to determine the adaptive forward-looking sequencing policy. The only limitation of this approach is that we have to set an endpoint for the session, which does not allow us to fully exploit exposures with $t > T$. As a robustness check, we also use an infinite horizon with more memory efficient state variables, where only we consider state variables that do not require storing the entire session-level history.

6.3 Evaluation

One of our main goals in this paper is to examine to what extent the dynamic framework helps publishers achieve better user engagement with ads. As such, we need to use an evaluation metric that allows us to evaluate and compare different policies. For any exposure t, we denote a t-step trajectory by g_t and define it as the sequence of states, actions, and rewards in all the steps as follows:

$$g_t = \langle s_1, a_1, r_1^c, \dots, s_{t-1}, a_{t-1}, r_{t-1}^c, s_t \rangle,$$

where any s_k is determined by the sequence prior to that exposure k and the distribution of transitions, a_k is determined given the policy, and r_k^c is the reward for the pair of s_k and a_k with the click-maximizing objective. Let τ denote the distribution of transitions and π denote any given policy. The joint distribution (τ, π) then determines the probability of each sequence g_t. Let $\rho_T(\pi; S_{i,1})$ denote the expected number of clicks generated in session i for the horizon length T, when using policy π. We can write:

$$\rho_T(\pi; S_{i,1}) = \mathbb{E}_{g_t \sim (\tau, \pi)} \left[\sum_{t=1}^{T} \beta^{t-1} r_t^c \right] \tag{28}$$

Now, we can use our estimates for the distribution of transitions τ and estimate $\rho_T(\pi; S_{i,1})$ for any policy π as follows:

$$\hat{\rho}_T(\pi; S_{i,1}) = \sum_{t=1}^{T} \sum_{g_t \in \mathcal{G}_T} \sum_{a \in \mathcal{A}_i} \pi(a \mid S_{i,t}) \hat{R}_t^c(a; S_{i,t}) P(g_t \mid \tau, \pi), \tag{29}$$

where $\mathcal{G}_T$ is the set of all possible trajectories and g_t denotes the trajectory in the first t periods. The last component in Equation (29) is the probability that a specific trajectory happens given the policy and distribution of transitions.

The main reason why we employ on this approach for evaluation is that it gives us session-level performance metrics on each policy. However, for robustness, we employ other approaches such as importance sampling and doubly robust method.

6.4 Results from the Click-Maximizing Policy

We now present the results on the performance of different sequencing policies when the main objective is to maximize the expected number of clicks per session. In §6.4.1, we illustrate session-level outcomes for different sequencing policies and examine the gains from adopting an adaptive forward-looking policy. In §6.4.2, we explore the heterogeneity in the gains from adopting an adaptive forward-looking sequencing policy across sessions. In §6.4.3, we use the distribution of counterfactual sequencing policies to explain the differences in them.

6.4.1 Gains from the Adaptive Forward-looking Sequencing Policy

We start by demonstrating the gains from the adaptive forward-looking sequencing policy compared to other policies described in §6.1. We use the direct method presented in §6.3 to evaluate the performance of these sequencing policies. As such, we only focus on the first six exposures and do not evaluate sessions after the sixth exposure. We draw a random sample of 1000 users from our test data that gives us 12,136 unique sessions. We estimate the optimal policy under all the sequencing policies defined above and present the results in Table 5.

As expected, all model-based sequencing policies lead to major improvements in the expected number of clicks per session over random allocation of ads. Adopting an adaptive forward-looking sequencing policy results in 0.1772 clicks per session which translates to an 80.36% improvement over the random sequencing, 7.87% over the non-adaptive single-ad sequencing, and 1.50% over the adaptive myopic sequencing. These gains demonstrate the value of each piece of information and modeling paradigm. Our results indicate that both session-level and future information play an important role in ad allocation problem.

Next, we focus on the gap between the adaptive forward-looking and non-adaptive single-ad

	Sequencing Policies			
	Fully Dynamic	Adaptive Myopic	Single-Ad	Random
Expected No. of Clicks	2150.60	2118.81	1993.75	1192.40
Expected No. of Clicks Per Session	0.1772	0.1745	0.1642	0.0983
Expected No. of Impressions	39365.30	39460.26.37	38925.97	38859.06
Expected Session Length	3.24	3.25	3.21	3.20
Expected CTR	5.46%	5.37%	5.12%	3.07%
% Click Increase over Random	80.36%	77.69%	67.21%	0.00%
No. of Users	1000	1000	1000	1000
No. of Sessions	12,136	12,136	12,136	12,136

Table 5: Performance of different sequencing policies for a sequence size of 6 with a click-maximizing objective

policy. This comparison relates to a broader question on whether the publisher should use a short-lived ad slot. Our results suggest that the use of dynamic ad slot leads to considerable gains and value creation compared to the fixed ad slot, justifying the current trend of using short-lived ad slots in the industry.[10]

We then examine the performance of adaptive forward-looking sequencing compared to that of adaptive myopic sequencing policy. Our results suggest that there are gains from adopting a forward-looking objective in adaptive ad sequencing. It also provides evidence regarding the interdependence of ads shown within the sequence. It is worth noting that 1.50% improvement is a lower bound for what the publisher can achieve by adopting a forward-looking objective. This is because we only focus on the first six exposures for computational reasons. The sequencing effects likely grow in later exposures. Further, we only focus on a limited ad inventory, with 15 ads. We expect the gains from the adaptive forward-looking ad sequencing to improve as a result of expanding ad inventory.

One likely explanation for the gains from an adaptive forward-looking sequencing policy is that it is the only model that takes users' leave decision into account. As such, one channel through which this policy might improve the outcome is to increase the session length, thereby enhancing the total number of clicks. To see whether this is the channel, we estimate the expected session length for all the policies. We find no significant difference between our policies in terms of the session length. Rather, the adaptive forward-looking sequencing policy generates a higher CTR, which suggests that users are more engaged with ads as a result of adaptive forward-looking ad

---

[10]However, the caveat here is that our approach in single-ad policy is not the equivalent of the fixed ad slot, as there are very short interruptions in every one minute this ad is being shown, whereas the fixed ad slot keeps the ad immobile throughout the session.

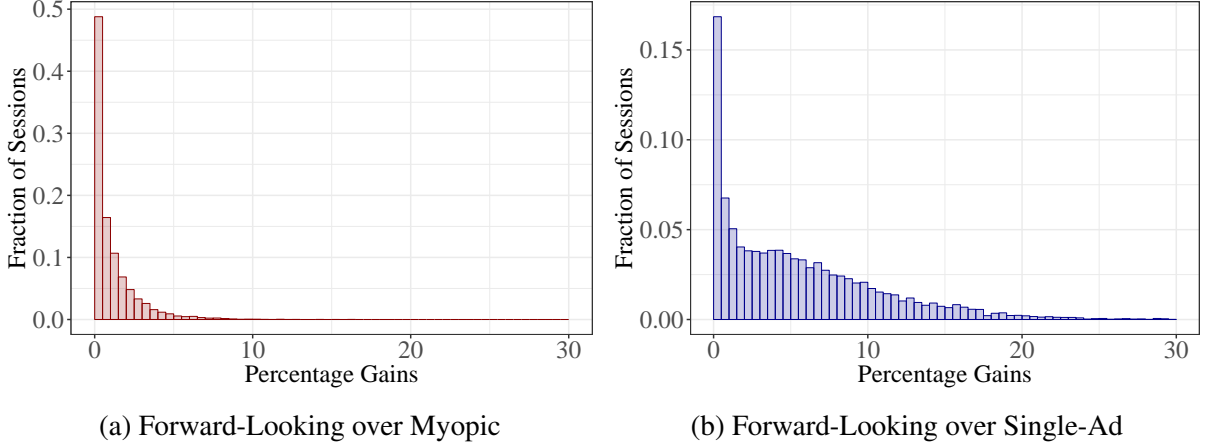(a) Forward-Looking over Myopic   (b) Forward-Looking over Single-Ad

Figure 12: Distribution of session-level gains from the adaptive forward-looking over both adaptive myopic and single-ad sequencing policies

sequencing.

6.4.2 Heterogeneity in Gains from the Dynamic Framework

While the results in Table 5 establish the average gains in clicks from adopting the adaptive forward-looking sequencing policy, they do not explain how these gains vary across sessions. In this section, we are interested in identifying the sessions for which the adaptive forward-looking sequencing is most helpful. As such, we focus on the session-level gains from the adaptive forward-looking compared to the adaptive myopic and non-adaptive single-ad policies and explore the heterogeneity across sessions. Let $Gain_m^f$ and $Gain_s^f$ denote the percentage gains from the adaptive forward-looking over adaptive myopic and single-ad sequencing respectively. In Figure 12, we show the distribution of these two session-level gains. Both figures show significant heterogeneity in the session-level gains from the adaptive forward-looking sequencing policy over adaptive myopic and single-ad policies.

To further explore this heterogeneity across user history, we regress both $Gain_m^f$ and $Gain_s^f$ on a set of historical variables, controlling for user fixed effects. We also control for the number of ads competing in a session as it is an important factor determining the gains from different sequencing policies. The results are presented in Table 6. The first two columns show the results using the gains from the adaptive forward-looking over the adaptive myopic sequencing policy as the dependent variable. In column 1, we examine how the relative gains from the adaptive forward-looking over adaptive myopic sequencing changes as a user participates in more sessions. On the one hand, as the user participates in more sessions, the publisher has more history and data on him, and it may improve the gains from adopting a forward-looking objective. On the other hand, as the user

becomes more experienced in the platform, the relative effectiveness of sequencing and temporal interventions may shrink (Rafieian and Yoganarasimhan, 2020a). As shown in column 1, we find some evidence for the latter: the relative gains from the adaptive forward-looking over adaptive myopic sequencing reduces as the user becomes more experienced.

	Dependent variable			
	$Gain_m^d$	$Gain_m^d$	$Gain_s^d$	$Gain_s^d$
Session Number$_{i,u}$	-0.000032***	0.000028**	-0.000447***	-0.000276***
	(-4.51)	(-1.95)	(-17.29)	(-5.41)
No. of Competitors$_i$	0.000032	-0.000253***	0.004188***	0.003070***
	(0.27)	(-2.17)	(9.90)	(7.31)
Imp$_{i,u}$		0.000019***		0.000056***
		(11.79)		(9.66)
Breadth$_{i,u}$		-0.000401***		-0.001681***
		(-15.61)		(-18.13)
User FE	✓	✓	✓	✓
No. of Obs.	12,136	12,136	12,136	12,136
R^2	0.3021	0.3223	0.2864	0.3097
Adjusted R^2	0.2469	0.2686	0.2300	0.2550
Note:				*$p<0.05$; **$p<0.01$; ***$p<0.001$

Table 6: OLS estimates of session-level gains from the adaptive forward-looking sequencing policy across user history

In column 2, we include two historical variables to further explore which parts of user's prior experience are accountable for the findings in column 1 – total number of impressions ($Imp_{i,u}$), and the total number of distinct ads ($Breadth_{i,u}$) the user has seen prior to the session. With these controls, we find that the number of prior impressions has a positive correlation with the outcome, implying that having more data on a user benefits the adaptive forward-looking sequencing more than the myopic one. However, note that the number of distinct ads the user has seen that makes adaptive forward-looking sequencing less effective. This is probably because the sequencing effects mostly come from users' information processing and differentiation of ads in a sequence. Once the user has processed many different ads in prior sessions, it is less likely that the optimal sequencing of them largely affects her decision.

In columns 3 and 4, we show the results with another dependent variable – gains from the adaptive forward-looking over non-adaptive single-ad sequencing. Again, we find that users' tenure in the platform makes adaptive forward-looking sequencing less effective, and this is because users with more experience have seen more diverse ads, and it is harder to influence their decision by

dynamic sequencing of ads. It is worth noting that the results in all columns must be interpreted in the relative terms. That is, a negative coefficient does not mean that adaptive myopic or non-adaptive single-ad policies perform better than the adaptive forward-looking sequencing policy. Rather, it identifies where adaptive forward-looking sequencing performs relatively worse.

6.4.3 Distribution of Micro-Interventions

In the previous sections, we established the gains from the adaptive forward-looking sequencing policy over other sequencing policies and showed how these gains vary across different sessions based on user-level history. These gains mainly come from the active selection of session-level features through the selection of ads in the adaptive forward-looking sequencing policy. In this section, we first illustrate how these sequencing policies are different in micro-interventions at any point and how that results in different session-level features. We then aim to uncover the reasons behind such differences, given the goal of each sequencing policy.

To make sense of how sequences evolve under each policy, we focus on binary micro-interventions as defined in §3.3.3. There are two key points about these micro-interventions. First, these binary micro-interventions together constitute our session-level features. For example, breadth-increasing change at any point results in a higher breadth of variety in a session, whereas repeat increases the number of prior exposures of an ad. Second, as shown in the prior research, all these three micro-interventions can be effective policies from a broader perspective. Recent experimental papers document the positive effects of multiple exposures on sales of an ad (Sahni, 2015a; Johnson et al., 2016b). While confirming the positive effects of multiple exposures of an ad within a session, Rafieian and Yoganarasimhan (2020a) show that both breadth-constant and breadth-increasing changes will increase the likelihood of clicking on the next ad. The fact that these micro-interventions are mutually exclusive makes the decision very challenging from a dynamic point-of-view, as using each comes with trade-offs. Thus, comparing distributions of these micro-interventions under different sequencing policies can help in understanding the decision-making process for each policy.

We present the average values for these micro-interventions at different time periods under each sequencing policy in Figure 13. Since these are binary variables, the y-axis reports the percentage of making any micro-intervention. As shown in these figures, both dynamic and adaptive myopic sequencing policies employ a mix of these micro-interventions. As such, the lines for them lie somewhere between the non-adaptive single-ad and random sequencing policies.

We further explore the difference between the adaptive forward-looking and adaptive myopic sequencing to reflect how these micro-interventions change if the publisher takes expected future rewards into account as in the adaptive forward-looking sequencing. We find that the adaptive

(a) Micro-intervention: repeat

(b) Micro-intervention: change

(c) Micro-intervention: breadth-increasing change
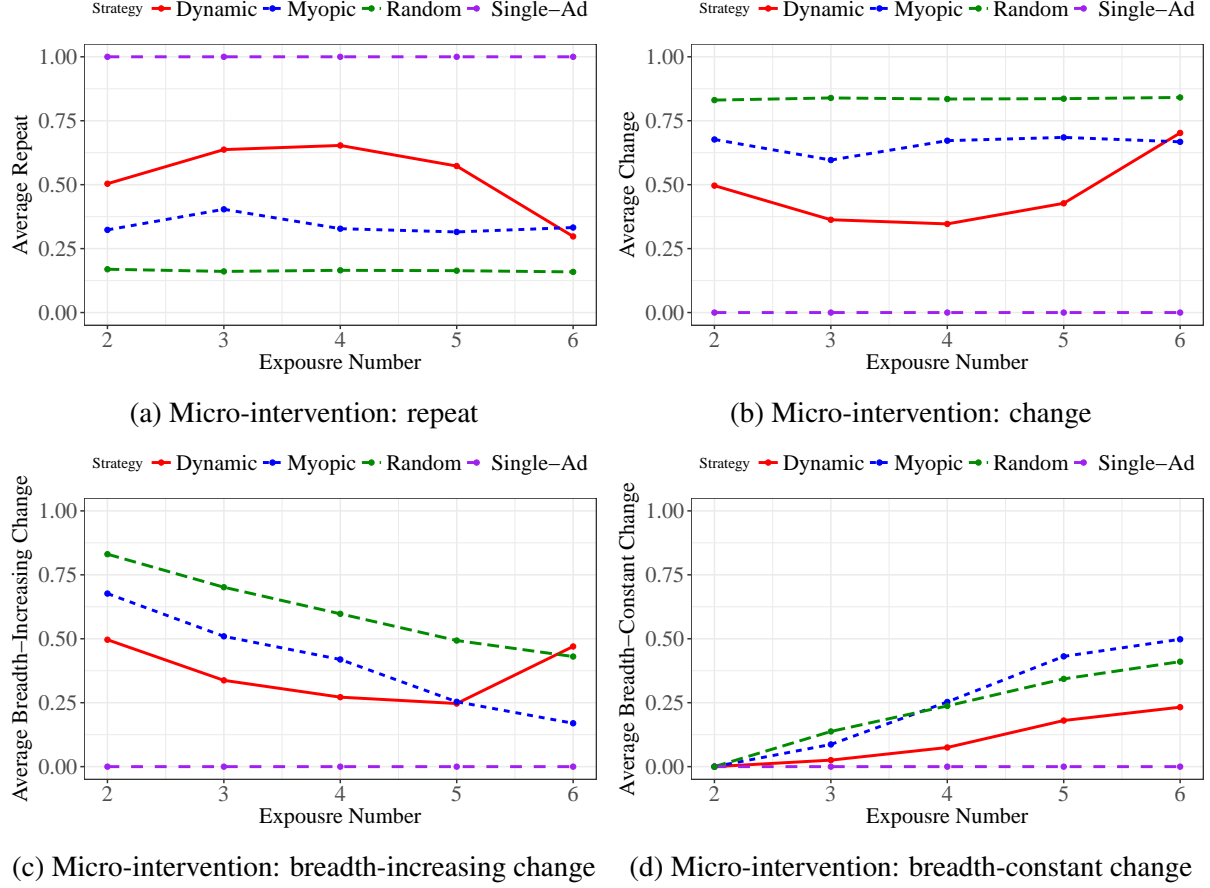
(d) Micro-intervention: breadth-constant change

Figure 13: Distribution of micro-interventions at different time periods under different sequencing policies

forward-looking sequencing policy tends to repeat ads in the earlier time periods, whereas change is a more popular decision in the adaptive myopic sequencing across all time periods. This is likely because repeating the same ad is a sub-optimal decision at the moment, but will help transition to a state where we can reinforce an ad with prior exposures. This is why the adaptive forward-looking sequencing policy takes advantage of repeating in the beginning, while in the last period, it tends to act similarly to the adaptive myopic sequencing.

Next, we look into the difference in the distribution of breadth-increasing and breadth-constant changes at different time periods. Since more prior exposures and changing the last ad helps at the moment, the adaptive myopic sequencing policy employs breadth-constant changes even more than the random sequencing. The adaptive forward-looking sequencing policy, however, seems to have a mix of both micro-interventions.

Overall, Figure 13 illustrates the differences in sequencing policies at a micro level. Our findings

demonstrate the trade-offs between these micro-interventions by comparing different sequencing policies. It is generally hard to offer a unifying explanation for the optimal policy in a Markov Decision Process, as it uses an elaborate objective function. Thus, it is important to notice that these findings are descriptive and must be interpreted cautiously.

7 Implications

Our findings in this paper have several implications for marketing practitioners and policy makers. The most direct set of implications is for publishers and ad-networks. We examine the opportunity cost of using a fixed ad slot, which is the current practice in many digital ad platforms. We document considerable loss as a result of running a fixed ad slot, in terms of the value created by making a better match between ads and users. This finding has implications for publishers and ad-networks that want to design their ad format. However, these results must be interpreted with caution.

More importantly, we establish the gains from the adaptive forward-looking sequencing of ads compared to adaptive myopic sequencing. While it significantly adds to the computational complexity of the problem, our findings indicate that inter-temporal trade-offs in ad allocation problem play an important role in optimal policy design. As such, publishers can create value by dynamically sequencing ads.

While advertisers are not the main target of the implications in this paper, our findings offer them some new insights. Given the gains from the adaptive forward-looking sequencing of ads, it can be beneficial for advertisers to incorporate session-level information while targeting their ads. Further, our paper adds to the understanding of short-lived ad formats, which, in turn, can provide some insights for advertisers with regards to their banner design.

More broadly, our general framework can be extended to any context with adaptive interventions. For example, in the context of mobile health, a growing body of work focuses Just-In-Time Adaptive Interventions (JITAI) in mobile apps and studies the impact of them in shaping consumers' health behavior including physical fitness and activity, smoking, alcohol use, and mental illness (Nahum-Shani et al., 2017). Similarly, in any context that adaptive interventions can be used to educate people, we can specify a dynamic framework that helps us achieve better outcomes (Mandel et al., 2014). These showcases can serve as motivation for the public sector to use these tools in cases where collective action is required, such as environmental protection and political participation.

8 Conclusion

Mobile in-app advertising has grown exponentially over the last years. The ability to exploit the time-varying information about a user to personalize ad interventions over time is a key factor in the growth of in-app advertising. Despite the dynamic nature of the information, publishers

often use myopic decision-making frameworks to select ads. In this paper, we examine whether a dynamic decision-making framework benefits the publisher in terms of the user engagement with ads, as measured by the number of clicks generated per session. Our dynamic framework has two main components: (1) a theoretical framework that specifies the domain structure such that it captures inter-temporal trade-offs in the decision to show what ad at any time period and (2) an empirical framework that breaks the policy design problem into a combination of machine learning tasks. We apply our framework to large-scale data from the leading in-app ad-network of an Asian country. Our results indicate that the adaptive forward-looking sequencing of ads results in significant gains in the expected number of clicks per session, compared to a set of benchmark policies. Next, we document heterogeneity in gains across sessions and show that adopting an adaptive forward-looking policy is most effective when users are new to the platform. Finally, we illustrate the differences in sequencing policies using a descriptive approach.

Our paper makes three contributions to the literature. First, from a methodological point-of-view, we develop a unified dynamic framework that starts with a theoretical framework that specifies the domain structure in mobile in-app advertising, and an empirical framework that breaks the problem into tasks that can be solved using a combination of machine learning methods and causal inference tools. Second, from a substantive standpoint, we are the first to document the gains from adopting an adaptive forward-looking sequencing policy as compared to the adaptive myopic sequencing policy. This comparison is of particular importance as the adaptive sequencing is the current approach in the industry. Third, we establish the gains from using a short-lived ad slot as compared to a fixed ad slot. The answer to this question informs the publisher's decision to use which type of ad slot.

Nevertheless, there are some limitations in our study that serve as excellent avenues for future research. First, our counterfactual policy evaluation is predicated on the assumption that users do not change their behavior in response to sequencing policies. While we exploit randomization to obtain our counterfactual estimate, it would be important to validate these findings in a field experiment. Further, we use the training data offline to learn counterfactual estimates for click and leave outcomes. Extension of our framework to an online setting that captures exploration/exploitation trade-off is important as offline evaluation may be costly. Finally, we use the entire within-session history to update state variables. Future research can look into more parsimonious frameworks that can be scalable to longer time horizons.

References

A. Ansari and C. F. Mela. E-customization. *Journal of marketing research*, 40(2):131–145, 2003.

N. Arnosti, M. Beck, and P. Milgrom. Adverse selection and auction design for internet display advertising. *American Economic Review*, 106(10):2852–66, 2016.

N. Arora and T. Henderson. Embedded premium promotion: Why it works and how to make it more effective. *Marketing Science*, 26(4):514–531, 2007.

R. Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM, 2016.

J.-P. Dubé and S. Misra. Scalable price targeting. Technical report, National Bureau of Economic Research, 2017.

D. Dzyabura and H. Yoganarasimhan. Machine Learning and Marketing. In *Handbook of Marketing Analytics*. Edward Elgar Publishing, 2018.

eMarketer. Mobile In-App Ad Spending , 2018. URL `https://forecasts-na1.emarketer.com/584b26021403070290f93a5c/5851918a0626310a2c186a5e`.

eMarketer. Time Spent with Mobile, US , 2019a. URL `https://forecasts-na1.emarketer.com/584b26021403070290f93a96/5851918b0626310a2c186b39`.

eMarketer. Digital Ad Spending 2019 , 2019b. URL `https://content-na1.emarketer.com/us-digital-ad-spending-2019`.

J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

S. Han, J. Jung, and D. Wetherall. A study of third-party tracking by mobile apps in the wild. *Univ. Washington, Tech. Rep. UW-CSE-12-03-01*, 2012.

J. R. Hauser, O. Toubia, T. Evgeniou, R. Befurt, and D. Dzyabura. Disjunctions of Conjunctions, Cognitive simplicity, and Consideration Sets. *Journal of Marketing Research*, 47(3):485–496, 2010.

G. J. Hitsch and S. Misra. Heterogeneous treatment effects and optimal targeting policy evaluation. *Available at SSRN 3111957*, 2018.

G. A. Johnson, R. A. Lewis, and E. I. Nubbemeyer. The online display ad effectiveness funnel & carryover: A meta-study of predicted ghost ad experiments. 2016a.

G. A. Johnson, R. A. Lewis, and D. Reiley. Location, location, location: repetition and proximity increase advertising effectiveness. 2016b.

W. Kar, V. Swaminathan, and P. Albuquerque. Selection and ordering of linear online video ads. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pages 203–210, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3692-5. doi: 10.1145/2792838.2800194. URL `http://doi.acm.org/10.1145/2792838.2800194`.

H. Li and P. Kannan. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51(1):40–56, 2014.

L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM, 2010.

T. Lu, D. Pál, and M. Pál. Contextual multi-armed bandits. In *Proceedings of the Thirteenth international conference on Artificial Intelligence and Statistics*, pages 485–492, 2010.

P. Manchanda, J.-P. Dubé, K. Y. Goh, and P. K. Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1):98–108, 2006.

T. Mandel, Y.-E. Liu, S. Levine, E. Brunskill, and Z. Popovic. Offline policy evaluation across representations

with applications to educational games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1077–1084. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

S. Mullainathan and J. Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, and S. A. Murphy. Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6):446–462, 2017.

O. Rafieian and H. Yoganarasimhan. How does variety of previous ads influence consumer's ad response? *Available at SSRN 3647373*, 2020a.

O. Rafieian and H. Yoganarasimhan. Targeting and privacy in mobile advertising, 2020b.

O. J. Rutz and R. E. Bucklin. From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research*, 48(1):87–102, 2011.

N. S. Sahni. Effect of temporal spacing between advertising exposures: evidence from online field experiments. *Quantitative Marketing and Economics*, 13(3):203–247, 2015a.

N. S. Sahni. Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quantitative Marketing and Economics*, 13(3):203–247, 2015b.

N. S. Sahni. Advertising spillovers: Evidence from online field experiments and implications for returns on advertising. *Journal of Marketing Research*, 53(4):459–478, 2016.

E. M. Schwartz, E. T. Bradlow, and P. S. Fader. Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4):500–522, 2017.

E. H. Simpson. Measurement of Diversity. *Nature*, 1949.

Z. Sun, M. Dawande, G. Janakiraman, and V. Mookerjee. Not just a fad: Optimal sequencing in mobile in-app advertising. *Information Systems Research*, 28(3):511–528, 2017.

G. Theocharous, P. S. Thomas, and M. Ghavamzadeh. Personalized ad recommendation systems for life-time value optimization with guarantees. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

O. Toubia, T. Evgeniou, and J. Hauser. Optimization-based and Machine-learning Methods for Conjoint Analysis: Estimation and Question Design. In A. Gustafsso, A. Herrmann, and F. Huber, editors, *Conjoint Measurement: Methods and Applications*, pages 231–258. Springer, 4 edition, 2007.

G. L. Urban, G. Liberali, E. MacDonald, R. Bordley, and J. R. Hauser. Morphing banner advertising. *Marketing Science*, 33(1):27–46, 2013.

J. Yi, Y. Chen, J. Li, S. Sett, and T. W. Yan. Predictive model performance: Offline and online evaluations. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1294–1302. ACM, 2013.

H. Yoganarasimhan. Search personalization using machine learning. *Available at SSRN 2590020*, 2018.