Information Systems Research

Personalized Mobile Targeting with User Engagement Stages: Combining Structural Hidden Markov Model and Field Experiment

Yingjie Zhang
Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080, yingjie.zhang@utdallas.edu,

Beibei Li
Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, beibeili@andrew.cmu.edu,

Xueming Luo
Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122, luoxm@temple.edu,

Xiaoyi Wang
School of Management, Zhejiang University, Hangzhou, China 310058, kevinwxy@zju.edu.cn,

Low engagement rates and high attrition rates have been formidable challenges to mobile apps and their long-term success, especially for those whose revenues derive mainly from in-app purchases. To date, little is known about how companies can scientifically detect user engagement stages and optimize corresponding personalized-targeting promotion strategies so as to improve business revenues. This paper proposes a new structural forward-looking Hidden Markov Model (FHMM) as combined with a randomized field experiment on app notification promotions. Our model can recover consumer latent engagement stages by accounting for both the time-varying nature of users' engagement and their forward-looking consumption behavior. Although app users in most of the engagement stages are likely to become less dynamically engaged, this slippery slope of user engagement can be alleviated by randomized treatments of app promotions. The structural estimates from the FHMM with the field-experimental data also enable us to identify heterogeneity in the treatment effects, specifically in terms of the causal impact of app promotions on continuous app consumption behavior across different hidden engagement stages. Additionally, we simulate and optimize the revenues of different personalized-targeting promotion strategies with the structural estimates. Personalized dynamic engagement-based targeting based on the FHMM can, compared with non-personalized mass promotion, generate 101.84% more revenue for the price promotion and 72.46% more revenue for the free content promotion. It also can generate substantially higher revenues than the experience-based targeting strategy applied by current industry practices and targeting strategies based on alternative customer segmentation models such as k-means or myopic HMM. Overall, the novel feature of our paper is its proposal of a new personalized-targeting approach combining the FHMM with a field experiment to tackle the challenge of low engagement with mobile apps.

Key words: User engagement, Mobile content consumption, App platforms, Hidden Markov Model, Forward-looking behavior, Structural econometric model, Field experiment

1. Introduction

Despite the increasing popularity of mobile technologies, a managerially significant problem persists: low user engagement with mobile apps. Consumers today spend significant amounts of time on mobile apps every day. A recent study (ComScore (2016)) showed that in 2016, mobile users spent approximately 73.8 hours per month on smartphone apps, compared to just 22.6 hours on tablets. The average mobile app usage time among the young (i.e., age 18-24) was even higher, approximately 93.5 hours per month. As indicated by the extant literature, researchers thus far have been attracted to the mobile market (e.g., Andrews et al. (2015), Ghose and Han (2014), Han et al. (2015), Luo et al. (2013), Xu et al. (2017)). Notwithstanding this growing trend, however, the in-app conversion rate remains stubbornly low. It has been reported that in February 2016, only 1.9% of all players paid for in-game content, and half of the revenues from all mobile game apps were contributed by only 0.19% of all players. In 2014, the average mobile app conversion rate was less than 2% in the U.S. Moreover, the attrition rate was high, 19% of mobile apps having been opened just once in 2015.

Indeed, such low engagement and high attrition have been major challenges to the long-term success of mobile app companies, especially those whose revenues come mainly from in-app purchases. In order to deal with these challenges, most mobile app companies have started to apply a variety of targeting strategies such as freemium (e.g., time-based freemium, feature-based freemium, seat-limited freemium) in adversely selecting consumers into different types according to their business values. Prior work has shown that providing a free version does, in fact, improve the sales of paid versions at the aggregate level (Ghose and Han (2014)). Besides, some mobile apps offering various plans or sales promotions encourage users to commit to multiple purchases over the long run. However, most of those targeting strategies are not tailored to individual mobile users but rather, are designed to be identical for all. In other words, all users are presented with, and face, the same products, prices and plans over time. Such non-personalized strategies are problematic, especially considering the significant variance of the app user population (e.g., active vs. non-active) over time. Thus, it is essential to tailor personalized-targeting strategies in order to effectively cope with the problem of low user engagement with mobile apps.

Against this background, the novel feature of this paper is its proposal of a new personalized-targeting approach to tackle the challenge of low engagement with mobile apps by combining a structural Hidden Markov model and a field experiment. The present study followed *3-step research design: 1. collect field-experimental data on targeting treatments; 2. develop a structural Hidden Markov model to detect heterogeneous treatment effects of targeting under different user engagement stages; 3. recommend personalized targeting to counter the trend of low user engagement with mobile apps and to increase sales revenues for the mobile app market.* More specifically, first, we conducted a randomized field experiment with two pre-designed targeting strategies. In the data obtained, we are able to identify exogenous promotion treatments' average causal sales impacts on mobile user reading behavior. Second, to further decompose the underlying incentives and mechanism of user behavior, we developed and applied the Forward-looking Hidden Markov Model (FHMM) to recover consumer latent engagement stages by accounting for both the time-varying nature of engagement and consumers' forward-looking consumption behavior. We estimated our model using the experimental data, the randomization of which enabled clear identification without worrying about the potential targeting endogeneity introduced by the user engagement stages (e.g., any potential self-selection bias due to unobserved user behavior will likely cancel out across the three randomized experimental groups). Based on our estimates, we detected the heterogeneous treatment effects and explained the heterogeneity with lifts in users' engagement state transactions. Finally, combining both the structural model and the randomized field experiment, we evaluated the optimal causal effects of engagement-based personalization in targeting strategies. Additionally, we compared our proposed personalization strategies with state-of-art targeting strategies. We found significant improvement in our engagement-based personalization.

Our empirical analyses yielded some interesting findings. First, our FHMM detects four user engagement stages, at each of which, users show different behavioral patterns. Second, without any extra policy interventions, users in most of the engagement stages are likely to become less engaged and leave the app; however, promotions can help alleviate this downward trend. Targeted promotions tailored to user engagement stages are even more effective. Third, our empirical analysis provides strong evidence on the heterogeneous treatment effects of different promotions on users at different engagement stages. We found that aware users, who are the least familiar with the app, prefer price promotion, whereas addicted users, who are the most engaged with the app, show more interest in free content promotion. This finding strongly suggests the importance of designing personalized promotions for different user engagement stages. Fourth, our policy simulation showed that compared with non-tailored mass promotion, our proposed dynamic engagement-based targeting can generate 101.84% more revenue for the price promotion and 72.46% more revenue for the free-content promotion. It can also engender substantially higher revenues than the experience-based targeting strategy applied by current industry practices and semi-dynamic engagement-based targeting with only one-period forward-looking modeling.

Overall, these findings from the combination of the FHMM with a field experiment are nontrivial. They suggest the high potential for revenue improvement in the mobile app market, particularly with respect to the roles of user engagement modeling and personalized targeting. Indeed, the structural model helps decompose heterogeneous treatment effects by engagement segment, which in turns, empowers businesses to target the most efficient users to effectively meet the challenge of low engagement with mobile apps.

2. Literature Review

2.1. User Engagement

Recently, the term "engagement" has been increasingly applied within the academic marketing field. Brodie et al. (2011) performed an exploratory analysis of its theoretical meaning and foundations. Kim et al. (2013) conducted a survey of mobile users' engagement stages and the reasons for their continually engaging with mobile activities. They found that engagement is the product of utilitarian, hedonic, and social motivations. Such studies' psychological findings, albeit interesting, are difficult to apply in the real world. Other researchers have proposed Hidden Markov models for detection of consumers' stages (Abhishek et al. (2012), Netzer et al. (2008)). In these studies, consumers were clustered into different "hidden" stage strata based on their behavioral patterns and willingness to pay. However, the investigative focus was only on consumers' one-time purchasing behavior. Such insights, unfortunately, cannot be generalized to the context of mobile apps, whose success often is a function of consumers' repeated purchasing and long-time loyalty. Further, the extant literature on the modeling of hidden stages and purchase decisions often assumes that consumers are myopic (Abhishek et al. (2012), Montgomery et al. (2004), Netzer et al. (2008)). This assumption can be unrealistic. To retain long-term consumers then, it is essential for apps to understand users' forward-looking behavior in order to predict their current and future decisions and, therefore, proactively tailor targeting strategies for improved user engagement, better experience, and higher satisfaction. Moreover, prior studies on observation data have been challenged by the potential of consumers' self-selection of engagement activities. For example, consumers who have stronger inherent preferences for the products or services on the app platform are more likely to become highly engaged, and also, meanwhile, to make purchases. Therefore, simply observing a positive relationship between engagement levels and purchase activities does not suggest a causal impact, nor does it indicate that companies should target consumers in the high-engagement stage to increase purchase rates. In our study, we automatically detected hidden engagement states using "hard" historical behavior data rather than "soft" survey perceptions. And we also show that the detection of user engagement is effective in designing personalized-targeting strategies.

2.2. Personalized Targeting

Personalized targeting or discrimination has been widely studied in the literature (Bakos and Brynjolfsson (1999), Choudhary (2010), Choudhary et al. (2005), Shiller et al. (2016), Dubé et al. (2017b), Fudenberg and Villas-Boas (2006)). The literature has shown the effects of such personalized targeting from the mobile marketing perspective, using competitive third-degree price discrimination. Prior studies have proposed several effective methods, including geo-based (Fong et al. (2015)), past-consumer-behavior-based (Fudenberg and Villas-Boas (2006)), and demographic-feature-based (Shiller et al. (2016)) approaches. Specifically, Fong et al. (2015) analyzed the causal effects of locational targeting by sending different levels of promotions to three different locations. They found that competitive locational targeting produced increasing returns. Fudenberg and Villas-Boas (2006) presented an analytical model wherein consumers' behavior in the last period affects their current valuation of a product. They pointed out that "As firms get better at processing this large amount of information, the effects of customer recognition are going to become more and more important." We extend these prior studies by applying the concepts of user engagement to the design of presonalized targeting. This allows us to utilize the individual's behavior sequence as well as to propose an easy method of consumer segmentation. Our policy simulation reveals the significant values of engagement-based targeting strategies relative to state-of-art methods. Additionally, we contribute to the literature by finding the heterogeneous effects (i.e., we observe that aware users, who are the least familiar with the app, prefer price promotion, while addicted users show more interest in free-content promotion), and by evaluating the effects with FHMM modeling.

2.3. HMM (Hidden Markov Model) in Marketing

Hidden Markov Model (HMM) is a stochastic process model in which unobserved states can affect the observed outcome. The HMM, widely utilized in the machine-learning field (e.g., Punera and Merugu (2010), Laxman et al. (2008)), recently has been introduced into the marketing field (e.g., Abhishek et al. (2012), Montgomery et al. (2004), Netzer et al. (2008), Kumar et al. (2011)). We summarize the literature in the related fields in Table 1. In general, a choice model is embedded in HMM; however, HMM, without consideration of consumers forward-looking behaviors, is nonetheless myopic. Recently, Arcidiacono and Miller (2011) proposed a strategy to account for unobserved heterogeneity in dynamic discrete choice models. We distinguish our approach from theirs in that we specify the transition of unobserved states using HMM, which allows us to identify the effects of observed features on unobserved engagement stages. This, in turn, would help us design proper mobile-app targeting strategies. The current study applied such a combination as the major framework in developing the novel FHMM. Identification and estimation issues have been proved theoretically in a recent working paper (Connault (2014)).

The prior studies above listed rely on observational data, which potentially incurs endogeneity issues. As we discussed above, a relationship between the engagement stage and the purchase decision does not necessarily indicate a causal impact from either direction. This can be easily solved with randomized field experiments. Recently, several studies have attempted to combine field experiments with the structural modeling (e.g., Dubé et al. (2017a,b)). In many ways, these two approaches are complementary. For example, structural model analysis can provide underlying mechanisms by which to explain findings from experiments (Dubé et al. (2017b)); on the other hand, to support assumptions required by structural models, field experiments offer exogenous shock, which is hard to satisfy in observational data (Dubé et al. (2017a)). In this paper, we propose a structural framework, namely a combination of single-agent dynamic discrete-choice structural models (Hotz et al. (1994), Miller (1984), Rust (1987)) and HMM, to identify the heterogeneous treatment effects as well as to design and evaluate better personalization strategies by combining the structural model with a randomized field experiment.

Also, we alter the perspective of this stream of literature on heterogeneous treatment results from a static one (e.g., quantile treatment effects and causal random-forest-based targeting) to a dynamic one (i.e.,full vs. semi dynamic personalization-based targeting). Static heterogeneous treatment effects have been widely studied, a typical example being the quantile treatment effect (QTE) (Qiu and Kumar (2017), Chernozhukov and Hansen (2005), Firpo (2007)). These papers developed and strengthened quantile regressions to identify the heterogeneous impacts of different variables. Meanwhile, there have been great efforts made to analyze heterogeneous treatment effects using machine-learning methods, including random-forest-based (Wager and Athey (2018)), lasso-based (Weisberg and Pontes (2015)), and Bayesian nonparametric (Bhattacharya and Dupas (2012)) approaches. The model we devised and proposed in this paper distinguishes itself from the literature in that it analyzes the heterogeneous treatment effects from a dynamic perspective, meaning that it allows for dynamic monitoring of individual users' real-time records, analyzes heterogeneous treatment effects, and assigns corresponding targeting strategies.

3. Average Treatment Effects of Mobile Promotions

As we discussed in Introduction, we propose a 3-step research design for analysis of the effectiveness of engagement-based personalized targeting. As the first step in the present study, we exploited data from a randomized field experiment on a mobile reading app. A clean field-experiment design allowed us to understand the average causal effects of different promotion designs. In this section, we first discuss the background of this reading app, and then we provide a detailed description of the setting of our field experiment.

3.1. Research Context

We conducted our empirical analysis on a Chinese top mobile reading app that offers more than 400,000 mobile books to over 130 million users per month. This mobile-app provides products very similar to Amazon Kindle but with specialized mobile-platform services.

This app can be easily and freely downloaded from app stores. Mobile phone users can then freely sign up for it using their phone numbers. Every time the user finishes reading a content unit, the app jumps to the next content unit automatically. If the user chooses not to read the given content unit, the app will show her a new book. In each book, the first several content units are free for any users. After that, to continue reading, users need to either pay per content or subscribe to the app to access all content provided on the platform. At the beginning of a new calendar month, the subscription contract continues automatically unless the user chooses to quit it, which means that the subscription will end from the next calendar month.

3.2. Field Experiment Design

To first understand how individual mobile users behave and react to typical marketing promotions, we conducted a field experiment on this mobile reading app. In our experiment, the pre-treatment period was from September 28 to October 27, 2015; the treatment period was from October 28 to November 8, 2015; and the post-treatment period was from November 9 to December 12, 2015. We randomly assigned users to three groups: two treatment groups with price-discount promotion (hereafter "price promotion") and free-content promotion respectively, and one control group.

Note that users here were those who registered for an account before the experiment and who might or might not have made a previous purchase. In price promotion group, users were provided with discount vouchers (total value: 0.60 RMB) for reading any content unit in the app; In the free-content promotion group, users were provided with five content-unit vouchers (total value: 0.60 RMB) for reading any content unit; in the control group, users were not provided with any promotion information but rather with placebo reminder notification messages (i.e., non-pricing advertisements). The comparison among the three groups indicated whether pricing promotion or non-pricing advertisement could achieve better performance in terms of sales lift. At the same time, the two treatments were designed to test whether users showed more attention to money or to products in marketing promotions.

The app offers *tapstream* data with individual behavioral trails on the app platform through finger taps. In this data, each record includes the following fields: user ID, time stamp, content information (e.g., name of content unit, book name, and book genre), and the user's choice of payment option (i.e., free content, pay-per-use, or subscription). As a check of randomization, Table 2 provides descriptive statistics of our experimental data for the *pre-treatment* period.

3.3. Experimental Results

To analyze the average effects of the different promotions on users' mobile reading app behavior, we used the difference-in-differences (DID) approach, applying, in a panel data structure, the equation:

$$\begin{aligned} Y_{it} = \alpha_0 &+ \alpha_1 \mathtt{Test}_t + \alpha_2 \mathtt{Treat1}_i \times \mathtt{Test}_t + \alpha_3 \mathtt{Treat2}_i \times \mathtt{Test}_t + \alpha_4 \mathtt{postTest}_t \\ &+ \alpha_5 \mathtt{Treat1}_i \times \mathtt{postTest}_t + \alpha_6 \mathtt{Treat2}_i \times \mathtt{postTest}_t + \xi_i + \varepsilon_{it}, \end{aligned} \tag{1}$$

where Y_{it} is the outcome measure of user i at time t; $\mathtt{Treat1}_i$ indicates whether user i is in the first treatment with price promotion; $\mathtt{Treat2}_i$ indicates whether user i is in the second treatment with free-content promotion, $\mathtt{Test}_t$ denotes the treatment period, and ξ_i represents the individual-level fixed effects. In our DID analysis, we defined four sets of outcome variables: total number of content units user i read in day t, number of free-content units, number of units with subscription contract, and number of units with the per-content option. Also, we had, additionally to treatment period data, post-treatment period data. To leverage this benefit, we defined a new variable $\texttt{postTreat}_t$ to explore whether the promotions had effects over a relatively long post-treatment period.

Table 3 presents our main regression results for two groups of users: active users (i.e., users who had reading records for the pre-treatment periods) and all users (both active and inactive users). Generally speaking, the two models returned qualitatively consistent results. The experimental results on average treatment effects yielded several interesting findings: Our experimental results, shown in Table 3, yield several interesting findings. *First*, the negative coefficients of the `Test` and $\texttt{postTreat}_t$ variables suggest that over time, the mobile reading app market has a dismal customer attrition picture. In general, however, most of the interaction terms showed estimates in the positive direction, indicating that promotions can alleviate the attrition trend. *Second*, the overall effect of price promotion was slightly better than that of free-content promotion, because with total amount of content as the outcome measure, the estimate of interaction terms in price treatment (i.e., 1.0026) is higher than that in free content treatment (i.e., 0.8152). We also found that the difference is statistically significant. Meanwhile, we also observed that the two promotions, in general, have different effects on different types of consumers. For example, the free-content promotion encourages active mobile users to read more free content. *Third*, in the analysis of the post-treatment data, we observed that the promotion effects can last after the treatment.

Promotion might be costly in practice. Although we did observe positive average treatment effects of both types of promotion, such a mass approach design might not be efficient for all consumers in all periods. For example, as shown in Table 3, whereas free-content promotion was effective in encouraging users' exploration with free content within the treatment period, the long-term effect was not good, especially compared with its effects on subscribers. Therefore, for better understanding of individual users' engagement evolution and decision making on mobile reading apps, and, thus also, for improved personalized-targeting strategy design, in the next section, we propose a new structural framework of mobile user segmentation. Our model, combined with the above field experiment design, will allow us to understand the heterogeneous treatment effects of different consumer segments, which in turn, will provide suggestions as to the design of optimal personalized-targeting strategies.

4. Forward-looking Hidden Markov Model (FHMM)

We constructed our framework by combining the single-agent dynamic discrete choice model (Rust (1987)) and the HMM (MacDonald and Zucchini (1997)). The HMM is a stochastic process model in which the states are unobserved but can affect the observed outcome. In our framework, we modeled individual users' engagement with the reading app as a hidden state in the HMM. A schematization of our proposed framework is presented in Figure 1. Users with diverse reading experience would be at different engagement stages in different phases. The stages will affect their period utility, which is used to form the expectation about future values. Finally, the decision is made based on the lifetime expected utility. A high level of engagement would, similarly to the purchase funnel concept, lead to a high probability of purchasing if other factors remain constant. In the following sub-sections, we will discuss the model in detail. We name this framework as the forward-looking Hidden Markov model (FHMM).

4.1. Model User Decisions

In each period, $t \in \{1, \ldots, T_i\}$ (total number of periods T_i varies across users), the mobile reading app shows a new content unit on mobile user i's screen. Then, user $i \in \{1, \ldots, I\}$ has the following three ($n^{\{D\}} = 3$) decision choices:

1. $d_{ijt} = 0$, user i chooses not to read (e.g., give up the current content, or leaves the mobile reading app platform). The corresponding utility is normalized to zero.

2. $d_{ijt} = 1$, user i chooses to read according to the pay-per-content option. She needs to pay per-content fee P_C if the given content unit is charged, and she is not under any subscription contract. Otherwise, the required payment $P_C = 0$.

3. $d_{ijt} = 2$, user i chooses to subscribe to the mobile app with payment P_S. After this, up to the expiration of the subscription contract, she has free access to all available content units.

4.2. Period Utility Function

The mobile user's decision-making process is not based solely on the period utility but also on inter-temporal trade-offs, which means that mobile users behave in a forward-looking manner. The determined part of the utility function consists of two components: utility of money and utility of reading. Utility of money is a linear function of the price the user needs to pay at time t for decision d_{it}. Utility of reading indicates the benefits of reading from the current content unit. We model this part with a user-engagement-specific constant.[1] Mathematically,

$$\begin{aligned} &U(d_{it}=d,\texttt{sub}_{it},F_t,e_{it}=e,\varepsilon_{it};\Theta) \\ &=\alpha\cdot\left(P_C\cdot\mathbb{1}\{d=1\}\cdot\mathbb{1}\{F_{it}=0\}+P_S\cdot\mathbb{1}\{d=2\}\right)\cdot\mathbb{1}\{\texttt{sub}_{it}=0\} \\ &+\tilde{\omega_e}\cdot\left(\mathbb{1}\{\texttt{sub}_{it}=1\}\cdot\mathbb{1}\{d=1\}+\mathbb{1}\{\texttt{sub}_{it}=0\}\cdot\mathbb{1}\{d\neq 0\}\right)+\varepsilon_{it}(d), \end{aligned} \tag{2}$$

where $\texttt{sub}_{it}$ is the subscription indicator, which equals 1 if user i is under subscription contract at time t. Mathematically, $\texttt{sub}_{it}=1$ if and only if $\exists n\in[0,t], d_{i,t-n}=2$. With this indicator, our dynamic modeling framework can capture the fact that mobile users can gain more benefits from reading additional content under subscription, even though their period utility might be lower (i.e., the subscription price is higher than the per-content price) than that with the non-subscription option. F_{it} indicates whether the content unit user i reads at time t is free (i.e., $F_{it}=1$ if the content unit is free). Note that F_{it} indicates whether the reading app company assigns no charge on the content, rather than whether user i needs to pay or not.[2] e_{it} denotes user i's engagement level at time t. We treat it as a hidden state that is used to predict users' probability of purchase. The

[1] Our tapstream data allow us to analyze individual users sequential decisions, which in turn, captures the time-varying factors. In our model, we assume that mobile users behave in a forward-looking manner. Our model, additionally, captures the sequence of decision making for an individual user. That is to say, users' current decision/engagement stages are derived from the previous path of consumption and engagement transition. Therefore, we cannot do a double count by incorporating the users' prior consumption units in modeling their utility or engagement transition function. In essence, the forward-looking component in our FHMM modeling framework is consistent with the notion that current consumption can be a function future consumption among rational users (Kwon et al. (2016)). Also, we extend the paper in that we introduce a 3-step research design by combining a structural model framework with a randomized field experiment to alleviate the potential endogeneity in analyzing user engagement.

[2] Through our communication with the reading app company, we found that a large majority of books (over 95%) offered on the app platforms have exactly 20 free chapters, and that this information is public to all users. Under such a situation, we believe that the potential effect from the free-content limit is unlikely to bias our findings, for the following reason: In our study, because almost all users were systematically facing the same free content limit, such a systematic fixed-effect was likely to cancel out when we compared across the control and treatment groups. Furthermore, even with those 5% of users who might encounter more or less than 20 free chapters during their reading experiences, because our users were randomly assigned to the three experimental groups, when we compare across the control and treatment groups, any remaining effect would have further canceled out due to randomization. Note that in this paper, our main goal was to understand and design heterogeneous targeting strategies based on the different user engagement stages. From this perspective, the effect of the free-content limit would be unlikely to have biased our estimation of treatment effects based on the randomized field experimental setting.

number of engagement stages ($n^{\{E\}}$) will be empirically tested and discussed in the results section. Θ is the parameter set Specifically, α is the price coefficient, which is identical across users. Based on the engagement stages, we define $\tilde{\omega_e}$ as an engagement-specific parameter vector (Netzer et al. (2008)). Due to identification concerns, we do not include a constant term in the utility function; otherwise, we cannot simultaneously estimate and the constant term. The utility form suggests that the mean utility of outside goods is normalized to zero. ε_{it} is the idiosyncratic choice-specific shock, which is assumed to independently and identically follow the Type I Extreme Value distribution. This stochastic term brings uncertainty to the model and captures unobserved factors that would affect users' utility. Examples of unobserved factors include promotion or advertisement of outside goods, social influence from friends, and so on.

To ensure identification of hidden state e, we assume the choice probability to be non-decreasing with an increasing engagement state value. Mathematically, this assumption is operationalized as

$$
\begin{aligned}
&\tilde{\omega_1} = \omega_1, \\
&\tilde{\omega_2} = \tilde{\omega_1} + exp(\omega_2), \\
&\dots \\
&\omega_{\tilde{n^{\{E\}}}} = \omega_{\tilde{n^{\{E\}}-1}} + exp(\omega_{n^{\{E\}}}),
\end{aligned}
$$

where ω is estimated from data. This assumption ($\tilde{\omega_1} < \tilde{\omega_2} < \cdots < \omega_{\tilde{n^{\{E\}}}}$) is commonly used in HMM-related work (Abhishek et al. (2012), Ascarza and Hardie (2013), Netzer et al. (2008)).

4.3. State Evolution

The state space is denoted as S. Our utility function, defined in Equation (2), contains three state variables: $(e_{it}, \texttt{sub}_{it}, F_{it})$.[3] Among these, $\texttt{sub}_{it}$ and F_{it} are observable from data, while e_{it} is the hidden stage. Below, we separately define their transition probabilities.

[3] Note that price is not included, because both per-content price and subscription price are constant over time.

First, regarding the subscription option, mobile users can benefit from it within the subscription contract period.[4] Mathematically,

$$\mathtt{sub}' = \begin{cases} 1 & \text{if } \mathtt{sub} = 1 \text{ or } d = 2 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Second, according to the app's marketing strategy, the first N content units of each book are free, whereby N is determined by the app company case by case. In most scenarios, a mobile user does not have any ex-ante knowledge for N. To capture this, we assume that the conditional transition probability of F is fixed and empirically inferreable from the data.

The above discussion suggests that both the transition probability of the free-content indicator (i.e., $f_F(F'|S,d)$) and that of the subscription indicator (i.e., $f_{\mathtt{sub}}(\mathtt{sub}'|S,d)$) are determined once the states and actions are fixed.

Next, the user engagement stage, e_{it}, is hidden (i.e., not observed from data). Like Netzer et al. (2008), we model transitions among engagement stages as a threshold model, wherein a discrete transition occurs if the corresponding transition propensity passes a threshold level. To compute the transition propensity, we model it as a function of the content features ($\mathtt{CF}_{it}$) user i has at time t. In other words, $\mathtt{CF}_{it}$ forms the transition matrix of the hidden engagement stage of user i at time t. For example, if a content unit from a popular book shows on user i's phone screen, her transition propensity is likely to be shifted above the threshold to a higher state; otherwise, her engagement is transited to a lower state, because the estimated transition propensity is below the threshold. We also allow the preferences towards $\mathtt{CF}_{it}$ to be heterogeneous when measuring the transition probability. That is, even with the same content units, users at different stages show diverse transition probabilities. In addition to the content features, how users came to the unit (i.e., self-selected, or recommended by app) would also affect their engagement evolution. To address

[4] In our model, we do not consider users' un-subscribing decision. On the mobile app platform, subscription contracts will continue by default without any extra action. Also, as we will discuss later in the data section, tapstream data do not include users' unsubscribing action. However, our model captures this action by assuming that users chose outside options and furthermore, had no choice on the app.

this, we add a search proxy (denoted as $\texttt{SP}_{it}$) to the engagement transition probability function. Our data do not explicitly include users' searching path on the reading app, but we can use the browsing data on free-content units to generate a proxy for search behavior. For example, if a mobile user just browses 2 (vs. 20) free units, that is a shallow (vs. deep) search or sampling of the book; also if she just browses the free content in 2 books or 2 (vs. 20) categories, that is a narrow (vs. broader) search. Formally, with the unobserved shock following the Type I Extreme Value distribution (i.i.d.), we define the non-homogeneous transition probabilities as the following ordered logit model (note that we have one constraint, $\sum_{e'=1}^{n^{\{E\}}} f_e(e'|S) = 1$):

$$
\begin{aligned}
&f_e(e'=1|S) = \frac{\exp(h(1,e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})}{1+\exp(h(1,e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})} \\
&f_e(e'|S) = \frac{\exp(h(e',e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})}{1+\exp(h(e',e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})} - \frac{\exp(h(e'-1,e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})}{1+\exp(h(e'-1,e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})}, e' \in \{2,\ldots,n^{\{E\}}-1\} \\
&f_e(e'=n^{\{E\}}|S) = 1 - \frac{\exp(h(n^{\{E\}}-1,e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})}{1+\exp(h(n^{\{E\}}-1,e) - \boldsymbol{\delta_e}\texttt{CF} - \boldsymbol{\gamma_e}\texttt{SP})},
\end{aligned}
\tag{4}
$$

where $\boldsymbol{\delta_e}$ and $\boldsymbol{\gamma_e}$ are the engagement-specific coefficients, and $h(e',e)$ is the e'-ordered logit threshold in state e.[5] Regarding the transition probabilities of content feature CF and search proxy SP, we empirically estimate from the tapstream data. In sum, our state space S contains five elements: $S = (e, \texttt{sub}, F, \texttt{CF}, \texttt{SP})$. We assume that all of these elements are independent from each other as conditional on given state values and decisions; therefore, the state transition probability f_S can be expressed as a multiplication of the five elements' transition probabilities:

$$
f_S(S'|S,d) = f_e(e'|S,d)\cdot f_F(F'|S,d)\cdot f_{\texttt{sub}}(\texttt{sub}'|S,d)\cdot f_{\texttt{CF}}(\texttt{CF'}|S,d)\cdot f_{\texttt{SP}}(\texttt{SP'}|S,d). \tag{5}
$$

4.4. Dynamics in Mobile Users' Decisions

Due to the availability of the subscription option, mobile users behave in a forward-looking manner. That is, they make decisions by maximizing the sum of discounted future period utilities:

$$
\max_{D_i=\{d_{i1},d_{i2},\ldots,d_{iT_i}\}} E[\sum_{t=1}^{\infty}\beta^{t-1}U(d_{it},S_{it},\varepsilon_{it};\Theta)|S_0,\varepsilon_{i0}], \tag{6}
$$

[5] To guarantee all the transition probabilities are within the range $[0,1]$, we need to include an additional assumption: $h(e',e) \geq h(e'-1,e), \forall e' = 2,\ldots,n^{\{E\}}$.

where β is the discounted factor in the lifetime utility function.

In detail, we define period t as each mobile user's tap. In the dynamic model specified in Equation (6), we assume that users evaluate their utility within infinite periods, which is a typical assumption of the forward-looking model (Rust (1987)).[6]

Another advantage of our FHMM is the availability of capturing users' expectations about future consumption. For example, whether the following content units of the same would be free or not, could somehow guide a user's decision on reading/subscribing or not. Our forward-looking model can capture such specification using a free-content indicator. Mobile users need to form their expectation about whether the content they might read in the future would be free or not. This expectation determines her evaluation of future utility and, finally, determines her decision.

The solution to the dynamic programming problem, as specified in Equation (6), is the same as that to the Bellman equation. We first rewrite the utility of choosing $d_t = k, k \in \{0,1,2\}$ in state S_t, with $\varepsilon_t(d_t) = \varepsilon_{kt}$ as the additive structure: $U(S_t, d_t = k) = u_k(S_t) + \varepsilon_{kt}$.

Then, by following the typical assumptions in Rust (1987), and assuming that ε_{kt} is i.i.d. across actions and time periods, we obtain the associated value function

$$\begin{aligned}\nu(S,\varepsilon) &= \max_{k\in\{0,1,2\}}\{u_k(S) + \varepsilon_k + \beta E[\nu(S',\varepsilon')|S, d = k]\} \\ &= \max_{k\in\{0,1,2\}}\{u_k(S) + \varepsilon_k + \beta \int \nu(S',\varepsilon') f_S(S'|S,k) d(S',\varepsilon')\}.\end{aligned} \tag{7}$$

A summary of all variables and notations is presented in Table 4.

With the state-specific value function defined in Equation (7), we can derive the conditional choice probability using the derived action-specific value function$V_k = u_k + \beta Q_k V$, where Q_k is the action-specific transition matrix. To estimate the model, we follow the nested pseudo maximum likelihood procedure (used in Aguirregabiria and Mira (2007) and Huang et al. (2015)) to solve the hidden-state single-agent dynamic problem. The detailed estimation procedure is provided in Appendix A.

[6] In our model, we ignore users unsubscribing behavior. Hence, when they choose to subscribe or not, they are evaluating their future frequencies of reading.

5. Heterogeneous Treatment Effects

In this section, we first discuss the detailed process of variable extraction. Then, we provide estimation results based on our FHMM proposed in Section 4. At the end of this section, we reveal and discuss our results regarding the heterogeneous treatment effects with user engagement detection.

5.1. Variable Extraction

Our tapstream data recorded fine-grained information on mobile users' behavioral trails.[7] We first extracted the search proxy, which is one of the factors affecting users' engagement evolution. We considered two types of proxies: the breadth search indicator and the depth search indicator. The breadth search indicator suggests whether users freely search a broad range of books/book-types; and the depth search indicator suggests whether users explore enough information of the same book prior to purchase. In order to recover these search proxies, we use users' behavior in reading free content. Empirically, the breadth search indicator is measured as the number of books and book types a user read in searching (i.e., with free content). Note that for a single user, the number of books with consumed free content is non-decreasing over time. We re-define this indicator as the number of books/types within one impression. The impression is a time series when the user *continuously* uses the app. The depth search indicator is measured as the number of free-content units a user has read in the current book. The two indicators are assumed to be binary, and we use the mean value across all users as the thresholds for setting of the binary values of each of the two variables.

In raw tapstream data, books are pre-classified into 247 types by the app company. We group these types into three genres: fiction, casual, and practical (Li (2015)). Types within the same genre have similar reading purposes. For example, casual books mostly serve entertainment purposes, while practical books require in-depth reading or even note-taking. Additionally, we count the total

[7] Note that in our analysis, we only considered active users (i.e., users with reading records defined from the pre-treatment period). Obviously, these active users were randomly assigned to the three groups and thus of no self-selection issue ex ante. The reason we used only active users, rather than the entire user base, is that we need records (i.e., observed choices) to estimate users preferences and classify them into different engagement groups. We acknowledge this limitation and future research can consider using additional information (e.g., population-wise demographics) to incorporate inactive users as well.

number of records for each book, according to the mean number of records, divide all books into two groups: ordinary and popular.[8] We include the above two covariates (i.e., book popularity and genre indicators) as content features in modeling user engagement evolution.

In our empirical analysis, the distributions of observed states (i.e., two search proxies, content features, and the free-content indicator) were empirically estimated from our data across all users over time. We assumed that if the time gap between two consecutive records is longer than 10 minutes, the user chose outside goods.[9] Otherwise, the time gap was treated as the reading time for the given chapter. The descriptive statistics on the key model variables are presented in Table 5.

5.2. Identification and Estimation Strategy

Our proposed model can theoretically identify the transition probability of unobserved states as well as the conditional probability of outcomes given state values, as described in An et al. (2013). According to the identification theorem discussed in Magnac and Thesmar (2002), given the error term distribution (assumed as a Type I Extreme Value distribution), the transition matrix (identified from the above assumption), discount factor (fixed to 0.99[10]) and utility of outside options normalized to 0, the utility function can be identified for all states and all decision choices.

Empirically, as shown in Equation (2), ω_e is the engagement-specific coefficient of reading utility. The repeated reading activities of the same user or the reading activities of similar users at the same engagement stage, conditional on the same price (i.e., the same free content after subscription, the same pay-per-content unit price before subscription, or the same bundling price upon subscription decision), can help us to identify the reading utility coefficient at each engagement stage. More specifically, if we observe that users choose to read the content rather than switch to outside options, we can infer, conditional on the same engagement stage and price, that the reading utility coefficient ω_e is high. Conditional on the reading utility, we can then identify the price coefficient α

[8] We have tried various other definitions based on the median value, or the top/bottom 25 percentiles as robustness tests. The results show consistency.

[9] We also tried 5-min, 15-min and 20-min as different extraction criteria for reading time, and the results were qualitatively robust.

[10] The qualitative nature of the results is robust to several other values of discounted factors.

through users' repeated purchases (reading activities). Specifically, a high magnitude of α indicates a higher reading frequency in the post-subscription period, meaning that users are more price sensitive and that more reading activities can minimize the waste of money on subscriptions. Moreover, if there is subscription behavior, we observe the user's repeated reading activities before and after the subscription. On that basis, we can also identify the price coefficient with the change in the frequency of repeated reading activities from the same user before and after subscription. For example, if we observe a significant increase in the frequency of repeated reading activities from users after subscription, this could indicate that they are relatively price sensitive (with a high magnitude of α). Additionally, the dynamics in the mobile users' forward-looking behavior also can help us identify the price elasticity. For example, if we observe a high frequency of repeated pay-per-content reading activities from users with no subscription, this might indicate that those users are less price elastic (with a low magnitude of α).

Further, to estimate this structural model, we applied our experimental data. The randomized setting in our field experiment would help us to address some potential endogeneity issues. Specifically, our field experiment considered two types of promotion: price promotion and content promotion. Therefore, we assigned three sets of parameters (in both utility and transition functions) to capture users' diversity among three cases: without promotion, with price promotion, and with content promotion.[11] With our randomized field experiment, for users in the two treated groups, we assumed that their reaction (e.g., their preference in measuring utility, and their forward-looking expectation) would change once the promotion started. Notwithstanding the fact of the user self-selection issue or that the recommendation strategies provided by the mobile reading app might bias our detection of user engagement, the randomized setting would allow us to tease out this effect across the three groups. Therefore, our model and estimation results can

[11] In our estimation process, we first estimated the parameters based on all of the data from the pre-treatment period. In other words, the pre-treatment parameters were commonly applicable for both the control group users and the treatment group users before treatment. Thus, at the beginning of the treatment, the four stages were identified using the common pre-treatment parameters, and too, were aligned among the control and treatment groups. Then, we followed the corresponding users' reading records in the post-treatment period in order to estimate the post-treatment parameters separately for the control and treatment groups.

still provide meaningful managerial implications regarding the heterogeneous treatment effects in the mobile users' engagement evolution.

In sum, the combination of our structural framework with the randomized field experiment has several advantages: on the one hand, the randomization in the experimental data allows us to eliminate the potential endogeneity issues in modeling mobile user engagement; on the other hand, the structural framework of user behavior complements the field experiment setting by identifying the heterogeneous treatment effects as well as evaluating the performance of potential personalized-targeting strategies.

5.3. Heterogeneous Treatment Effects on Engagement Transition

The first step in estimating our model was to determine the number of engagement stages (i.e., the hidden state). We compared several alternative models with different numbers of engagement stages (varying from 2 to 6). With respect to the AIC and BIC criteria, the results, shown in Table 6, indicate that the best solution is to identify four engagement stages. We herein label the four stages "aware", "exploring", "active", and "addicted".

We first present our estimated engagement state transition in Table 7.[12] Since we model engagement transition as a function of multiple covariates, capturing content features and users' search behavior, we compute the transition matrix (shown in Table 7) with the mean values of the covariates. We report all three matrices in the same table for better comparison. We first discuss the transition probability in the control group's without-promotion case (the baseline case without intervention). On the whole, the matrix shows that most of engagement stages are highly likely to switch to the lowest stage. This finding suggests that the mobile reading app could lose its mobile users without any additional intervention. In other words, there is a downward trend or slippery slope of user engagement: users in most of the engagement stages are, dynamically, likely to become less engaged. This finding is consistent with the current industry reality of the app market, wherein

[12] The detailed estimation results and discussions of coefficients in the utility function and transition function are provided in Appendix A.

low engagement and high attrition rates have been major challenges to companies as we discussed in the Introduction.

This downward trend of user engagement, however, can be improved with either price or free-content promotion in the following ways. (1) With promotions, the downward trend from all current stages (e) to the future stages (e') becomes smaller. Mobile users at all current stages e except the "aware" stage have, when compared with the without-promotion cases, a higher probability of moving up to the highest stage in the future stages, and a lower probability of moving down to the lowest stage in the future stages (e'). (2) We also observe, when promotions are available versus the control, a significant increase in the transition probability from the current "exploring and active" stages (i.e., $e = 2, 3$) to future higher stages (i.e., $e' = 3, 4$). The two types of promotions we considered here are related to mobile users' quantitative benefits. Intuitively, the users at the exploring stage are becoming familiar with the app by exploring the content within it. More free content or more available coupons allow users to explore more without extra cost. (3) We also see a significant increase in the probability of users' staying in the highest "addicted" stages (i.e., $e = 4$), i.e., the highest customer retention rates, when promotions are available. These trends are consistent with our reduced-form analysis, shown in Table 3, where we find that promotions do work in encouraging users to consume on the app. (4) Interestingly, the comparison between price promotion and free-content promotion highlights the *heterogeneous* treatment effects in terms of engagement switching probability. For example, we observed that free-content promotion is more effective in encouraging users to transit from a lower stage (e.g., "exploring" and "active" stages) to a higher stage (e.g., "active" and "addictive" stages, respectively). In contrast, price promotion is more effective in keeping users in their current stages (i.e., with more "addictive" users staying in the same stage). The potential reasons behind this is that with the free-content promotion design, more content would be directly available to users to encourage more content consumption and platform exploration, while the price promotion can reduce the cost for addicted users, which in turn, might potentially stimulate them to subscribe to the app and stay on the platform for longer than usual. Overall, these transition probability results for the treatment and control groups from the FHMM estimates help identify each mobile user's engagement stage so as to understand users' dynamic behavioral path, i.e., how they made consumption decisions, and change their engagement levels with the app over time. Next, we use the FHMM estimates to shed more light on the heterogeneous treatment effects among the different engagement stages.

5.4. Heterogeneous Treatment Effects on User Reading Behavior

In Table 7, we present the heterogeneous treatment effects on users' engagement evolution. We further explored, similarly to our reduced-form analysis shown in Table 3, the heterogeneity using individual users' outcome decision as our measure. Specifically, we used structural estimates to identify each mobile user's engagement stage at the beginning of the treatments. Then, we divided all of the users into four segments based on their stages.

Similarly to what we did in Section 3, we used the same panel DID approach and defined the outcome measure as the total number of units a user read per day. We then examined the effects within each engagement segment from the FHMM. The results are shown in Table 8. Interestingly, the treatment effects varied across the four engagement stages. On the whole, we found that the treatment effects were driven mainly by "aware" (i.e., $e = 1$) and "addicted" ($e = 4$) users. Also, the optimal promotions differed. Specifically, price promotion could lead to a higher probability of purchase and higher revenues for aware users, who are the least familiar with the app. On the other hand, addicted users (e.g., loyal users) have a higher probability to read more content units when free-content promotion is available. The intuition behind this difference might be as follows: money matters for aware users because they are new or unfamiliar to the app. Compared with users in other stages, aware users care more about their actual expense on an unfamiliar app. On the other hand, the better performance of content promotion for addicted users implies that they show more loyalty to the app services and that they care more about the service content itself. Two managerial implications of this finding is that the app company should focus more on choosing the right promotion strategies to target unfamiliar opposed to addicted users, and that they should avoid using the same promotion strategies for users at different stages.

Overall, these findings, of the combined structural-model/field-experiment analysis, suggest that the effects of app notifications are dependent on the right mix of data analytics (user engagement modeling) and app notification creativity (promotions emphasizing free content or price discounts). Because over 50% of app users find app notifications annoying (Localytics (2016)), firms should tap into their user engagement analytics to create personalized notifications. Such personalized app notifications are what specific user segments want to receive, and they can generate substantially larger sales impacts than non-personalized broadcasting app notifications. Next, we used the structural FHMM estimates to simulate and identify the optimal personalized-targeting strategies.

6. Targeting-Strategy Design

Thus far, our structural estimates have demonstrated the diverse reactions to promotions among different engagement stages (as shown in Table 8). This provides us with a potentially valuable approach to the segmentation of mobile users in designing targeting strategies. Here, we examine how effective a dynamic personalized engagement-specific targeting strategy corresponding to users' engagement stages would be. To evaluate the performance of this targeting, we compared it with other strategies including mass promotions, historical-purchase-based personalized promotions, and semi-dynamic engagement-based personalized promotions. In our simulation, the user base was the 4586 control group users in our field experiment. To measure the effects of policy intervention, we used users' total payment (i.e., subscription and per-content payment) *per period.*[13] Note that the period we defined is the time period within which a new content appears on mobile users' phone screen. Thus, the total payment per period is used to evaluate the total expected payment per decision.

Specifically, we first computed users' decision probability at each period given the state variable values; then, we calculated the expected payment amount using the decision probability and

[13] The simulated decision probability was computed for each period. In each period, we observed mobile users' past reading behavior as their state variable values. The simulated decision probability was still based on users' forward-looking behavioral patterns and the same state variables, but with different policy interventions (e.g., different prices). An alternative computation method is to compute users' complete decision sequences from the first period using the initial state variable values. This would require us to compute the sequential state variable values as well, which can incur more uncertainty in the predicted decision probability. All the above process was based on our structural estimates, shown in Table 10, Appendix B.

the average of the expected payment per period; finally, we aggregated all of the users in order to compute the overall expected revenue for one single period. In the simulation,[14] we assumed that the promotions started at the same time as our field experiment, and in all of the simulated cases with promotions, users received five free-content coupons or money of equal value (i.e., 0.6 RMB). In Table 9, we present the simulated revenues for the following six cases. (1) Baseline: mass promotion. We assume that app managers do not have any personal information on their users, and so all users receive the same price promotion or free-content promotion. (2) Experienced-based personalized promotion. This strategy is similar to the industry's current effort wherein an app company monitors users' past-purchase records. In this simulation, we used users' cumulative purchase amount before promotion to divide them into four quantiles, and free coupons were provided only to users in certain quantiles. (3) K-means-based personalized promotion. K-means is a popular clustering method in the machine-learning field. Similarly to case 2, we used K-means to divide users into segments before the start of promotions. (4) Myopic-HMM-based personalized promotion. Our FHMM considers mobile users' forward-looking behavior. To further examine whether this is necessary, we considered an alternative myopic HMM model under the assumption that users make decisions in a myopic manner. Again, similarly to case 2, the HMM allowed us to divide users into groups, and we applied to them different promotion designs at the beginning of both promotions. Note that to make the model comparable with the alternatives, we allowed the same utility function as in our FHMM and consider the same features in modeling the transition probability. The only difference was that mobile users' decisions are based on current utility values without consideration of future discounted utility. Note that in Table 9, we decided the number of segments/groups derived from Case (3) and (4) using the one that would generate the highest revenue. Specifically, we have 3 segments with K-means clustering algorithm and 4 groups with the myopic HMM method. (5) Semi-dynamic engagement-based personalized promotion. This, again,

[14] In general, the policy simulation approach might be constrained to examine hypothetical policy changes that are near to the domain in which the model was specified and estimated.

was similar to case 2, though we defined user segments based on users' engagement stages immediately prior to the promotion. Thus, in this case, the app managers were semi-dynamic, which is to say that the FHMM was implemented only before the field experiment, not after it. (6) Dynamic engagement-based personalized promotion. Compared with case 5, where we had only one-period forward-looking modeling, this case allowed the reading app to dynamically monitor users' engagement stages in real time. Thus, in case 6, the app managers were fully dynamic; that is, the FHMM was implemented both before and after the field experiment and over the entire period. And in our simulations, when the promotion started, they were available only to users at certain engagement stages. One single user could get up to five coupons during the promotion period. Note that from case 2 to case 6, to determine which user segments should be targeted, we chose the one with the highest total payment per period within each scenario, so as to compare the optimal effect within each targeting strategy. Mathematically, we simulated the revenues for the four cases using the equation

$$\tilde{\texttt{Revenue}} = \begin{cases} \sum_i \frac{1}{T_i}\left[\sum_{t<\tau^i_{\texttt{prmt}}} R(S_{it};\Theta^{\texttt{N}}) + \sum_{t<\tau^i_{\texttt{prmt}}} R(S_{it};\Theta^{\texttt{prmt}}) - \texttt{Cost}_i\right]; \texttt{Case 1} \\ \max_q \sum_i \frac{1}{T_i}\left[\sum_{t<\tau^i_{\texttt{prmt}} \cup Q^i_{\texttt{Case}_\texttt{n}} \neq q} R(S_{it};\Theta^{\texttt{N}}) + \sum_{t<\tau^i_{\texttt{prmt}} \cap Q^i_{\texttt{Case}_\texttt{n}} = q} R(S_{it};\Theta^{\texttt{prmt}}) - \texttt{Cost}_i\right]; \texttt{Case 2-5}, \\ \max_q \sum_i \frac{1}{T_i}\left[\sum_{t<\tau^i_{\texttt{prmt}}} R(S_{it};\Theta^{\texttt{N}}) + \sum_{t<\tau^i_{\texttt{prmt}} \cap Q^{it}_{\texttt{exp}} = q} R(S_{it};\Theta^{\texttt{prmt}}) - \texttt{Cost}_i\right]; \texttt{Case 6}, \end{cases} \tag{8}$$

where

$$R(S_{it};\Theta^{\texttt{N}}) = \underset{\texttt{sub}}{Pr}(S_{it};\Theta^{\texttt{N}})P_S + \underset{\texttt{perC}}{Pr}(S_{it};\Theta^{\texttt{N}})P_C,$$

$$R(S_{it};\Theta^{\texttt{prmt}}) = \underset{\texttt{sub}}{Pr}(S_{it};\Theta^{\texttt{prmt}})P_S + \underset{\texttt{perC}}{Pr}(S_{it};\Theta^{\texttt{prmt}})P_C, \tag{9}$$

$$\texttt{For case } n = \{2,3,4,5\}, \texttt{Case}_\texttt{n} = \{\texttt{exp}, \texttt{KMeans}, \texttt{HMM}, \texttt{eng}\}$$

where $\tilde{\texttt{Revenue}}$ is the simulated revenue, $\underset{\texttt{sub}}{Pr}(S_{it};\Theta^{\texttt{prmt}})$ and $\underset{\texttt{perC}}{Pr}(S_{it};\Theta^{\texttt{prmt}})$ denote user i's probability of choosing a subscription or per-content option given the state value S_{it} at time t, $\Theta^{\texttt{N}}$ means that the probability is calculated based on without-promotion parameters, while $\Theta^{\texttt{prmt}}$ indicates the probability with promotions. In our simulations, we used both price-promotion and free-content-promotion parameters, with P_S and P_C as the corresponding subscription and per-content prices. The total time period of user i is T_i, and $\tau^i_{\texttt{prmt}}$ denotes the starting period of promotion for user i. $Q^i_{\texttt{exp}}$, $Q^i_{\texttt{KMeans}}$, $Q^i_{\texttt{HMM}}$, and $Q^i_{\texttt{eng}}$ are user quantiles based on either past experience, K-means, myopic HMM, or FHMM at the beginning of promotions. In case 6, the quantile was computed in real-time; accordingly, quantile $Q^{it}_{\texttt{eng}}$ has superscript t. With price promotion, the mobile app company would pay up to 0.6 RMB back to the treated users; and with free-content promotion, the mobile app company would not charge the treated users for the first five contents. These two scenarios would generate $\texttt{Cost}_i$ in the above equation.

As shown in Table 9, the results suggest that under both price and free-content promotion settings, all of the cases from 2 to 6 show a positive increase in revenue, thereby indicating the effectiveness of applying personalized-targeting strategies. Second, we showed larger increases from cases 1 & 2 to cases 5 & 6, implying that engagement from the FHMM is an important factor in predicting users' reaction to promotions, particularly when compared with traditional approaches with past-activity-based personalization, which is commonly used in the promotion design of the current mobile app markets. Third, we also compared our FHMM-based personalization with other popular machine-learning-based personalizations, including the K-Means clustering approach (case 3) and the basic (with modeling of users' myopic behavior) HMM method (case 4). The results showed that these machine-learning approaches are helpful in consumer segmentation (relative to mass promotion) but that our FHMM can do better. Interestingly, we observed that the myopic-HMM-based personalized targeting performs better than the K-means algorithm and the experience-based quantile approach. This indicates the advantage of incorporating HMM in modeling mobile user behavior. Fourth, we found a giant improvement with dynamic engagement-based promotion, i.e., with fully implemented FHMM modeling both before and after the field experiment over the entire period. As compared with the non-tailored mass promotion, the tailored optimal dynamic engagement-based targeting from fully implemented FHMM could generate 101.84% more revenue for the price promotion and 72.46% more revenue for the free-content promotion. Finally, the heterogeneity between the two promotions also implies the importance of understanding users' engagement. Specifically, in case 2, our result shows that if we apply the current industry's practice in using users' past purchasing behavior to design the targeting strategy, we cannot detect the heterogeneity effects under different promotions. On the other hand, our results based on the engagement-based strategies (i.e., cases 5 and 6) showed that with different promotions, the most effective targeting strategy varies as well. For example, if we use fully dynamic engagement-based promotion, it would be optimal to use price promotion to target "aware" and "exploring" users because money matters the most for users who are not familiar with the mobile app. Meanwhile, the optimal targeting strategy with free-content promotion would be to target "addicted" users, who show loyalty to the app and care more about the product than the price itself.

Therefore, overall, the above-reported transition probability results, heterogeneous treatment effects of hidden engagement stages, and simulation results with structural estimates both before and after the field experiment demonstrate the potency of combing the FHMM with a randomized field experiment, which is to say, the value of understanding users' dynamic behavioral paths, revealing the economic value of modeling user engagement, and crafting optimal personalized dynamic engagement-based targeting strategies.

7. Conclusion

This study provides empirical evidence on the importance of detecting user engagement, as well as the effectiveness of designing engagement-based targeting strategies, using a methodological combination of a mobile field experiment with the FHMM. We base our research on an analysis of individual mobile users' continuous reading records on a mobile reading app. Our model can recover consumer latent engagement stages by accounting for both the time-varying nature of user engagement and forward-looking consumption behavior. The structural estimates from FHMM with the field-experimental data enable us to identify heterogeneity in the average treatment effects, in terms of the causal impact of app promotions on (1) the underlying mechanism of user engagement evaluation, and (2) continuous app consumption behavior across different hidden engagement stages. Moreover, our methodology also allows us to identify dynamic heterogeneous treatment effects (DHTE). While prior studies have evaluated heterogeneous treatment effects in terms of how they vary across time-invariant covariates, our simulation case examines how the personalized-targeting promotion effects vary across time-varying hidden stages (via structural estimates from FHMM). Furthermore, we show the effectiveness of leveraging engagement knowledge in personalized-targeting strategies. Compared with non-personalized mass promotion, personalized optimal dynamic engagement-based targeting based on FHMM can generate 101.84% more revenue for price promotion and 72.46% more revenue for free-content promotion.

Our study demonstrates the methodological strengths of combining a structural model and a field experiment, thus, revealing the crucial role of modeling user engagement and optimizing personalized dynamic targeting for potential revenue improvements in the mobile app market. Our proposed method can also be applied to other digital infrastructures and platforms in similar marketing settings to help platforms to identify their users' behavior and adjust their targeting strategies and improve their business models accordingly. In addition to the digital platform market, our method can also provide guidelines for other durable products, the demand for which comes from consumers' forward-looking behaviors.

Our paper has a few limitations that nonetheless provide interesting opportunities for future research. Our current utility model considers the price and users' engagement stages as two main factors. The current consumption, however, would be influenced by the past consumption. To account for such effect, we applied two search proxies to approximate the users' past consumption in measuring users' engagement transitions. Future studies can consider directly incorporating the actual consumption in users' utility function. This would help understand the sequence of consumers' decisions. In addition, there might exist an engagement hierarchy in terms of book genres, books, and chapters (books within a genre, chapters within a book); also engagement evolution and utility preference vary with multiple factors, including cross-device behavior, time-of-day, weather, and others (Li et al. (2017), Xu et al. (2017)). For example, people in metropolitan cities typically spend more time on public transportation, which might allow them to be more engaged in the app during their commuter time. Further study respecting these issues might be more interesting and practical to mobile market managers. Our model, with the necessary adjustments, offers the potential for incorporating such factors.

References

Abhishek, Vibhanshu, Peter Fader, Kartik Hosanagar. 2012. Media exposure through the funnel: A model of multi-stage attribution Avaliable at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2158421.

Aguirregabiria, Victor, Pedro Mira. 2007. Sequential estimation of dynamic discrete games. *Econometrica* **75**(1) 1–53.

An, Yonghong, Yingyao Hu, Johns Hopkins, Matt Shum. 2013. Identifiability and inference of hidden markov models. Tech. rep., Technical report.

Andrews, Michelle, Xueming Luo, Zheng Fang, Anindya Ghose. 2015. Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science* **35**(2) 218–233.

Arcidiacono, Peter, Robert A Miller. 2011. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* **79**(6) 1823–1867.

Ascarza, Eva, Bruce GS Hardie. 2013. A joint model of usage and churn in contractual settings. *Marketing Science* **32**(4) 570–590.

Bakos, Yannis, Erik Brynjolfsson. 1999. Bundling information goods: Pricing, profits, and efficiency. *Management science* **45**(12) 1613–1630.

Bhattacharya, Debopam, Pascaline Dupas. 2012. Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics* **167**(1) 168–196.

Brodie, Roderick J, Linda D Hollebeek, Biljana Jurić, Ana Ilić. 2011. Customer engagement: Conceptual domain, fundamental propositions, and implications for research. *Journal of service research* **14**(3) 252–271.

Chernozhukov, Victor, Christian Hansen. 2005. An iv model of quantile treatment effects. *Econometrica* **73**(1) 245–261.

Choudhary, Vidyanand. 2010. Use of pricing schemes for differentiating information goods. *Information Systems Research* **21**(1) 78–92.

Choudhary, Vidyanand, Anindya Ghose, Tridas Mukhopadhyay, Uday Rajan. 2005. Personalized pricing and quality differentiation. *Management Science* **51**(7) 1120–1130.

ComScore. 2016. The 2016 u.s. mobile app report .

Connault, Benjamin. 2014. Hidden rust models Available at http://economics.mit.edu/files/10373.

Dubé, Jean-Pierre, Zheng Fang, Nathan Fong, Xueming Luo. 2017a. Competitive price targeting with smartphone coupons. *Marketing Science* **36**(6) 944–975.

Dubé, Jean-Pierre, Xueming Luo, Zheng Fang. 2017b. Self-signaling and prosocial behavior: A cause marketing experiment. *Marketing Science* **36**(2) 161–186.

Firpo, Sergio. 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* **75**(1) 259–276.

Fong, Nathan M, Zheng Fang, Xueming Luo. 2015. Geo-conquesting: Competitive locational targeting of mobile promotions. *Journal of Marketing Research* **52**(5) 726–735.

Fudenberg, Drew, J Miguel Villas-Boas. 2006. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems* **1** 377–436.

Ghose, Anindya, Sang Pil Han. 2014. Estimating demand for mobile applications in the new economy. *Management Science* **60**(6) 1470–1488.

Han, Sang Pil, Sungho Park, Wonseok Oh. 2015. Mobile app analytics: A multiple discrete-continuous choice framework. *Management Information Systems Quarterly (MISQ)* **40**(4) 983–1008.

Hotz, V Joseph, Robert A Miller, Seth Sanders, Jeffrey Smith. 1994. A simulation estimator for dynamic models of discrete choice. *The Review of Economic Studies* **61**(2) 265–289.

Huang, Yan, Param Vir Singh, Anindya Ghose. 2015. A structural model of employee behavioral dynamics in enterprise social media. *Management Science* **61**(12) 2825–2844.

Kim, Young Hoon, Dan J Kim, Kathy Wachter. 2013. A study of mobile user engagement (moen): Engagement motivations, perceived value, satisfaction, and continued engagement intention. *Decision Support Systems* **56** 361–370.

Kumar, V, S Sriram, Anita Luo, Pradeep K Chintagunta. 2011. Assessing the effect of marketing investments in a business marketing context. *Marketing Science* **30**(5) 924–940.

Kwon, Hyeokkoo Eric, Hyunji So, Sang Pil Han, Wonseok Oh. 2016. Excessive dependence on mobile social apps: A rational addiction perspective. *Information Systems Research* **27**(4) 919–939.

Laxman, Srivatsan, Vikram Tankasali, Ryen W White. 2008. Stream prediction using a generative model based on frequent episodes in event sequences. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 453–461.

Li, Chenxi, Xueming Luo, Cheng Zhang, Xiaoyi Wang. 2017. Sunny, rainy, and cloudy with a chance of mobile promotion effectiveness. *Marketing Science* **36**(5) 762–779.

Li, Hui. 2015. Intertemporal price discrimination with complementary products: E-books and e-readers Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2735092`.

Localytics. 2016. An even better recipe for perfect push notifications .

Luo, Xueming, Michelle Andrews, Zheng Fang, Chee Wei Phang. 2013. Mobile targeting. *Management Science* **60**(7) 1738–1756.

MacDonald, Iain L, Walter Zucchini. 1997. *Hidden Markov and other models for discrete-valued time series*, vol. 110. CRC Press.

Magnac, Thierry, David Thesmar. 2002. Identifying dynamic discrete decision processes. *Econometrica* **70**(2) 801–816.

Miller, Robert A. 1984. Job matching and occupational choice. *The Journal of Political Economy* 1086–1120.

Montgomery, Alan L, Shibo Li, Kannan Srinivasan, John C Liechty. 2004. Modeling online browsing and path analysis using clickstream data. *Marketing Science* **23**(4) 579–595.

Netzer, Oded, James M Lattin, Vikram Srinivasan. 2008. A hidden markov model of customer relationship dynamics. *Marketing Science* **27**(2) 185–204.

Punera, Kunal, Srujana Merugu. 2010. The anatomy of a click: modeling user behavior on web information systems. *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 989–998.

Qiu, Liangfei, Subodha Kumar. 2017. Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. *Information Systems Research* **28**(3) 529–546.

Rust, John. 1987. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* 999–1033.

Shiller, Ben, et al. 2016. Personalized price discrimination using big data. Tech. rep.

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

Weisberg, Herbert I, Victor P Pontes. 2015. Post hoc subgroups in clinical trials: Anathema or analytics? *Clinical trials* **12**(4) 357–364.

Xu, Kaiquan, Jason Chan, Anindya Ghose, Sang Pil Han. 2017. Battle of the channels: The impact of tablets on digital commerce. *Management Science* **63**(5) 1469–1492.

Table 1 Literature on HMM

Study	Model Specification	Hidden Stages	Objective	Data
Abhishek et al. (2012)	Myopic	Consumer states in a conversion funnel	Advertising attribution	Observational
Montgomery et al. (2004)	Dynamic multinomial probit model	Browsing states	Online browsing behavior	Observational
Netzer et al. (2008)	Myopic	Relationship states	Customer relationships	Observational
Arcidiacono and Miller (2011)	Dynamic, forward-looking	Unobserved heterogeneity	Dynamic Optimization problems	Observational
Our paper	Dynamic, forward-looking	User engagement	Optimized targeting	Field experiment

Table 2 Descriptive Statistics in Pre-treatment Period

	Treatment 1 Price Promotion	Treatment 2 Free Content	Control	p-value of ANOVA
Number of users	14,352	14,460	14,159	
Number of active users	4,594	4,661	4,586	0.4765 (Chi-square stat)
Daily # of records	2.3837 (6.4517)	2.2659 (6.6859)	2.3987 (6.9293)	0.1783
Daily # of books	0.6372 (1.3807)	0.6515 (1.5473)	0.6582 (1.5372)	0.4713
Daily # of genres	0.5354 (1.0762)	0.5216 (1.0740)	0.5250 (1.0684)	0.5207

Note: The standard deviations are shown in parentheses.

Table 3 Field Experiment Analysis

Y_{it}	With active users only				With all users			
	# of units		# of free units		# of units		# of free units	
`Treat1`	1.0026*	0.9435*	0.1306	0.1117	1.0811***	1.0811***	0.2375***	0.2375***
×`Test`	(0.4492)	(0.4663)	(0.1374)	(0.1352)	(0.0964)	(0.0956)	(0.0328)	(0.0305)
`Treat2`	0.8152*	0.7839*	0.2543*	0.2130*	0.3280***	0.3280***	0.0888**	0.0888*
×`Test`	(0.4294)	(0.4453)	(0.1314)	(0.1291)	(0.1004)	(0.0995)	(0.0341)	(0.0318)
`Test`	-1.1543***	-1.2993***	-0.4301***	-0.3597***	-0.3833***	-0.3833***	-0.1786***	-0.1786*
	(0.3346)	(0.3469)	(0.1024)	(0.1006)	(0.0738)	(0.0731)	(0.0251)	(0.0234)
`Treat1`		1.7882***		0.0120		1.4968***		0.2287***
×`postTreat`		(0.3842)		(0.1114)		(0.0703)		(0.0225)
`Treat2`		1.8568***		0.0418		0.2482***		0.0012
×`postTreat`		(0.3670)		(0.1064)		(0.0732)		(0.0234)
`postTreat`		-2.3748***		-0.3070***		-1.4124***		-0.5674***
		(0.2890)		(0.0838)		(0.0538)		(0.0172)
Observations	322,328	569,696	322,328	569,696	1,193,680	2,109,760	1,193,680	2,109,760

Notes: ***p<0.001; **p<0.01; *p<0.05.

Clustered standard errors are shown in parentheses.

Unit: book chapter. `Treat1`: price promotion; `Treat2`: free-content promotion;

`Test`: test period indicator; `postTreat`: post-treatment period indicator.

Table 4 Summary of Notations

Notation	Description
i, t	Indices of mobile user and period
I, T_i	Total numbers of users and periods (for user i)
$n^{\{S\}}, n^{\{E\}}, n^{\{D\}}$	Total numbers of states, engagement stages, and decision choices
d_{it}	User i's choice at time t
P_C, P_S	Per-content price (0.12 CNY) and subscription price (5 CNY)
$\mathtt{sub}_{it}, F_t$	Indicator of subscription, and free content unit
e_{it}	User i's engagement stage at time t
α	Price coefficient in utility function
$\tilde{\omega}_e$	Engagement-specific coefficient of reading in period utility function
$(f_e, f_F, f_{\mathtt{sub}}, f_{\mathtt{CF}}, f_{\mathtt{SP}})$	Transition probabilities
$\mathtt{CF}$	Feature vector of content units read by user i at time t
$\mathtt{SP}$	Search proxy vector of user i at time t
$h(e', e)$	e'-ordered logit threshold in state e
$\boldsymbol{\delta_e}, \boldsymbol{\gamma_e}$	Vector of engagement-specific response coefficients of $\mathtt{CF}_{\boldsymbol{it}}, \mathtt{SP}_{\boldsymbol{it}}$
β	Discount factors in lifetime utility function (assumed as 0.99)

Table 5 Descriptive Statistics of Extracted Variables in FHMM Model Estimation

Var.	Description	Mean	Minimum	Maximum	Std.dev.
T_i	Number of decision peiords	482.0497	2	8866	725.1177
$\mathtt{CF}_{\mathtt{Pop}}$	Popularity indicator	0.9527	0	1	0.2122
$\mathtt{CF}_{\mathtt{fiction}}$	Fiction genre indicator	0.9278	0	1	0.2587
$\mathtt{CF}_{\mathtt{casual}}$	Casual genre indicator	0.0500	0	1	0.2180
$\mathtt{SP}_{\mathtt{breadth}}$	Breadth search indicator	0.0442	0	1	0.2056
$\mathtt{SP}_{\mathtt{depth}}$	Depth search indicator	0.5065	0	1	0.4999
$\mathtt{Sub}$	Subscription indicator	0.5071	0	1	0.4999
F	Free-content indicator	0.8284	0	1	0.3770
Y	Decision indicator	0.8653	0	1	0.3430

Table 6 Comparison of HMM Models

Model	# of states	Log-likelihood	AIC	BIC	# of Variables
FHMM	2	-68,4801	-68,510.1	-68,645.3	15
	3	-48,310.5	-48,360.5	-48,535.9	25
	4	**-20,079.7**	**-20,153.7**	**--20,413.2**	37
	5	-28,101.9	-28,203.9	-28,561.6	51
	6	-38,308.7	-38,442.7	-38,912.7	67

Note: The best model in each column is shown in bold.

Table 7 Estimated Transition Matrix of Engagement Stages

	$f(e'\|e,\bar{\mathtt{CF}},\bar{\mathtt{SP}})$	$e'=1$ (aware)	$e'=2$ (exploring)	$e'=3$ (active)	$e'=4$ (addicted)
	$e=1$	0.9993	0.0002	0.0005	0.0000
Control:	$e=2$	0.9771	0.0024	0.0080	0.0125
without promotion	$e=3$	0.6677	0.0071	0.2645	0.0607
	$e=4$	0.3429	0.1773	0.2580	0.2218
	$e=1$	1.0000	0.0000	0.0000	0.0000
Treatment 1:	$e=2$	0.7685	0.0875	0.0040	0.1400
price promotion	$e=3$	0.2847	0.7122	0.0018	0.0013
	$e=4$	0.1195	0.0565	0.2234	0.6007
	$e=1$	0.9997	0.0003	0.0000	0.0000
Treatment 2: free-	$e=2$	0.5326	0.3053	0.0428	0.1194
content promotion	$e=3$	0.2901	0.0286	0.1259	0.5554
	$e=4$	0.1925	0.1249	0.3819	0.2965

Note: $\bar{\mathtt{CF}}$ and $\bar{\mathtt{SP}}$ mean the mean value of the covariables in the engagement transition function: content features and search proxies.

Table 8 Field Experiment Analysis by Segment

Engagement	**Without post-treatment period**				**With post-treatment period**			
Stage	$e=1$	$e=2$	$e=3$	$e=4$	$e=1$	$e=2$	$e=3$	$e=4$
`Treat1`	1.9754*	2.8907	2.6116	6.7371*	1.9832*	2.8562	2.7984	6.2227*
×`Test`	(1.0115)	(4.1255)	(5.9846)	(3.2803)	(1.0090)	(3.9856)	(5.5721)	(3.2435)
`Treat2`	1.0968*	4.0066	4.0339	7.7964*	1.2421	3.8507	4.1725	7.0252*
×`Test`	(1.0529)	(4.0760)	(5.9526)	(3.3015)	(1.0609)	(3.9359)	(5.5356)	(3.2553)
`Test`	-1.4397**	4.4666	-5.0889	-7.5515**	-1.5865**	-4.2040	-5.5133	-7.3530**
	(0.6057)	(3.9705)	(5.8600)	(3.1431)	(0.6123)	(3.8276)	(5.4367)	(3.1033)
`Treat1`					2.9279*	4.4983	-0.6426	1.6779
×`postTreat`					(1.2983)	(4.9389)	(2.8639)	(2.0801)
`Treat2`					3.4285**	5.56678	-0.8956	2.0980
×`postTreat`					(1.1628)	(4.9058)	(2.8279)	(2.1053)
`postTreat`					-3.0413***	-4.4841	-2.3989	-1.7587
					(0.6813)	(4.8072)	(2.6000)	(1.8171)
Observations	158,928	56,330	58,265	54,524	280,896	99,560	102,980	96,368

Notes: ***p<0.001; **p<0.01; *p<0.05. Clustered standard errors are shown in parentheses. DV: Number of all units read per day. `Treat1`: price promotion; `Treat2`: free-content promotion; `Test`: test period indicator; `postTreat`: post-treatment period indicator.

The t-test shows that the differences in treatment effects between the two treatment groups were significant when $e=1\&e=4$.

Table 9 Comparison of Simulated Targeting Strategies

Targeting Strategy	**Price Promotion**		**Free-Content Promotion**	
	Target Group(s)	Revenue per period (RMB)	Target Group(s)	Revenue per period (RMB)
Case 1: mass promotion		427.6223		456.9403
Case 2: experience-based personalized promotion	the 4^{th} (highest) quantile	483.4055 (13.04%))	the 4^{th} (highest) quantile	484.0798 (5.94%))
Case 3: KMeans-based personalized promotion	the 3^{rd} quantile	492.8214 (15.24%))	the 3^{rd} and 4^{th} quantile	492.8617 (7.86%))
Case 4: myopic-HMM-based personalized promotion	$e=3$ & $e=4$	495.6522 (15.91%))	$e=2$ & $e=3$ &$e=4$	496.8221 (8.73%))
Case 5: semi-dynamic engagement-based promotion	$e=3$ & $e=4$	499.9547 (16.92%))	$e=2$ & $e=3$ &$e=4$	501.7523 (9.70%))
Case 6: dynamic engagement-based promotion	$e=1$ & $e=2$	863.0918 (101.84%))	$e=4$	788.0202 (72.46%))

Notes: % compared to mass promotion (case 1) is shown in parentheses.

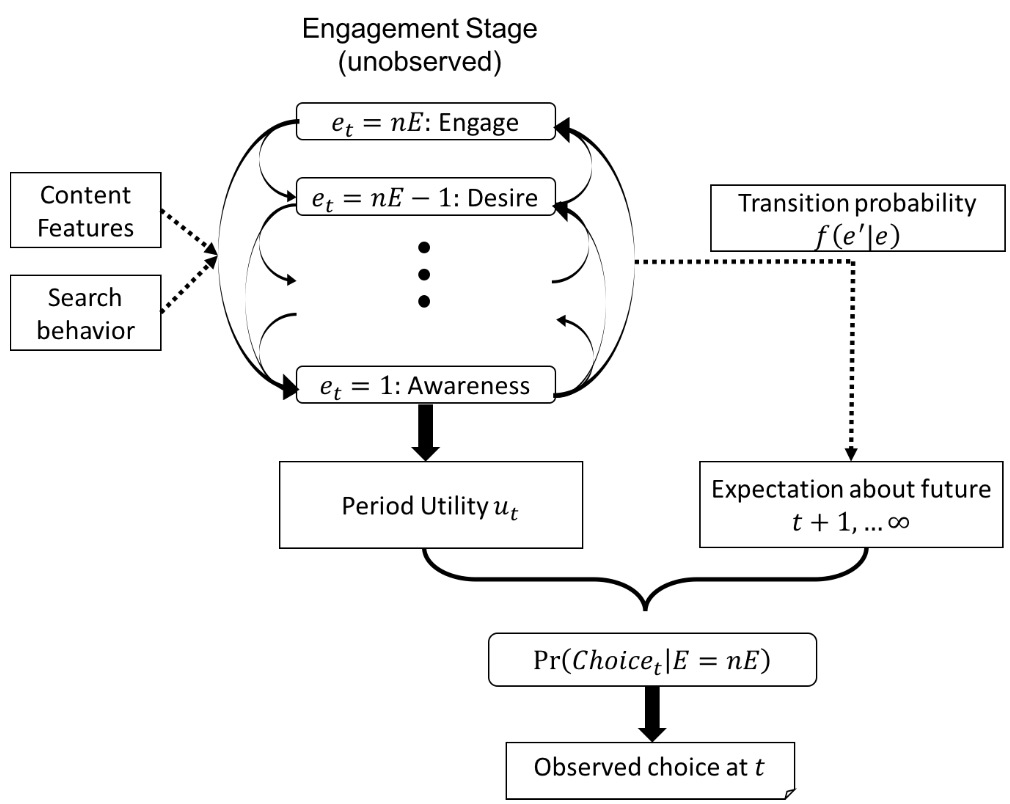

Figure 1 FHMM Framework

Online Appendix: Personalized Mobile Targeting with User Engagement Stages: Combining Structural Hidden Markov Model and Field Experiment

Appendix A: Identification and Estimation

The structural parameters in our model are $\{\alpha, \boldsymbol{\omega}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{h}, \beta\}$. The three elements in bold are parameter vectors of engagement-specific coefficients: for example, $\boldsymbol{\omega} = \{\omega_1, ..., \omega_{nE}\}$. Due to the identification issue of discount factor, we assume $\beta = 0.99$. Thus, the model primitives that need to be estimated are: $\Pi = \{\alpha, \boldsymbol{\omega}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{h}\}$.

A.1. Identification

Our model combines both HMM model and single-agent dynamic discrete choice model. We will first discuss the identification issue from the theoretical perspective using several existing theorems in the literature, and then discuss empirically the identification issue from our data set.

A.1.1. Theoretical Identification An et al. (2013) proves that to identify the transition probability of unobserved states and conditional probability of outcome given states, we need four assumptions: (1) The time series process of unobserved states and outcome should be strictly stationary and ergodic; (2) The probability matrix $S_{Y_t, Y_{t-1}}$ has full rank; (3) The conditional probability should be different with different value of unobserved states; (4) There exists a functional $F(\cdot)$ such that $F(f_{Y_t|S_t})$ is monotonic in S_t (Y_t is outcome variable and is the unobserved state). We argue that our model satisfies the above four assumptions:

(1) We allow mobile users to move upward and downward in the engagement funnel, which satisfies the ergodicity assumption; To meet the stationarity assumption, we need to assume the initial probability of state $\boldsymbol{\Phi_i}$ is stationary. To achieve that, we follow Netzer et al. (2008) and Huang et al. (2015), by calculate the stationary $\boldsymbol{\Phi_i}$ from the equation: $\boldsymbol{\Phi_i} = \boldsymbol{\Phi_i} \tilde{\boldsymbol{Q}}_{\boldsymbol{i}}$, where $\tilde{\boldsymbol{Q}}_{\boldsymbol{i}}$ is the transition matrix calculated at the mean value of covariates for individual i.

(2) Assumption 2 requires that $P(Y|S = j)$ is not a linear function of $P(Y|S = k), \forall k \neq j$. This can be easily satisfied from our model setting.

(3) In our model, different values of unobserved state are corresponding with different coefficients in the utility function, which will lead to different value functions and probability of decision choices.

(4) One easy way to interpret assumption 4 is that given Y, different values of unobserved state have different conditional probability, so that the conditional probabilities can be correctly ordered. Similar to (2), our model setting meets this requirement.

According to the identification theorem discussed in Magnac and Thesmar (2002), given the distribution of error term (assumed as Type I Extreme Value distribution), the transition matrix (identified from the above assumption), discount factor (fixed to 0.99) and utility of outside option normalized to 0, the utility function can be identified for all states and all decision choices.

A.1.2. Empirical Identification As shown in Equation (2), ω_e is the engagement-specific coefficient of reading utility. The repeated reading activities of the same user or reading activities from similar users at the same engagement stage, conditional on the same price (i.e., same free content after subscription, same pay-per-content unit price before subscription, or same bundling price upon subscription decision), can help us identify the reading utility coefficient at each engagement stage. More specifically, conditional on the same engagement stage and price, if we observe that users choose to read the content rather than switch to outside options, we can infer the reading utility coefficient ω_e is high.

Conditional on the reading utility, we can then identify the price coefficient, α. Note that due to the unique setting of our context, we only observe three levels of price in the data: free, pay-per-content price, and subscription price. Nevertheless, we observe users' repeated purchases (reading activities). We are able to identify the price elasticity from users' frequency of repeated reading activities after subscription. Specifically, a high magnitude of indicates higher reading frequency in post-subscription period, meaning that the users are more price sensitive and more reading activities can eliminate the waste of money on subscription. Moreover, if there is a subscription behavior, we observe the user's repeated reading activities before and after the subscription. Then we can also identify the price coefficient with the change in the frequency of repeated reading activities from the same user before and after subscription. For example, if we observe a significant increase in the frequency of repeated reading activities from the users after subscription, this may indicate that the users are relatively price sensitive (with high magnitude of α). In addition, the dynamics in the mobile users' forward-looking behavior can also help us identify the price elasticity. For example, if we observe high frequency of repeated pay-per-content reading activities from the users with no subscription, this may indicate the users are less price-elastic (with low magnitude of α).

A.2. Estimation Procedure

We follow the nested pseudo maximum likelihood procedure that was used in Aguirregabiria and Mira (2007) and Huang et al. (2015), to solve the hidden-state single-agent dynamic problem.

Step 1: Estimate $f_{\boldsymbol{X}}(\boldsymbol{X'}|S,d,\boldsymbol{\lambda_1})$.

In the observed reading experience variable vectors $\boldsymbol{X}$, we consider two vectors: (1) content features $\texttt{CF}$; and (2) search proxies $\texttt{SP}$. As we discussed before, the transition probabilities of these observed variables are a function of previous state variables. We assume the function to be a logit function and use data to estimate the coefficients vector $\boldsymbol{\lambda_1}$ out of the iteration loop. Then we will get $\hat{f}_{\boldsymbol{X}}(\boldsymbol{X'}|S,d,\boldsymbol{\lambda_1})$.

Step 2: Conduct the joint likelihood function.

Let $D_i = \{d_{i1}, d_{i2}, \ldots, d_{iT_i}\}$ represent the choice sequence for individual i, $S_i = \{S_{i1}, S_{i2}, \ldots, S_{iT_i}\}$ represent the state sequence, and $\boldsymbol{\Phi_i}(1 \times n^{\{S\}})$ represent the initial state distribution. Due to the identification concern, to compute the initial state probability $\boldsymbol{\Phi_i}$, we need to solve the equation $\boldsymbol{\Phi_i} = \boldsymbol{\Phi_i}\tilde{\boldsymbol{Q}}_{\boldsymbol{i}}$ (under the constraint $\sum^{n^{\{S\}}} \pi_{iS} = 1$) where $\tilde{\boldsymbol{Q}}_{\boldsymbol{i}}$ is the transition matrix calculated at the mean value of covariates for individual i. For simplicity, we use O_i to represent the observed state sequences (i.e., $O_i = \{S_i \setminus e_i\}$). Then, the probability (i.e., individual likelihood function) of the observed outcome sequence D_i and observed state sequence O_i is:

$$l(D_i, O_i) = P(D_i, O_i|\boldsymbol{\lambda_2}) = \sum_{e_i} \big(P(D_i|O_i, e_i) \times P(e_i|O_i, \boldsymbol{\lambda_2}) \times P(O_i|\boldsymbol{\lambda_1})\big). \tag{A.1}$$

Step 3: Calculate the parameters in the transition matrix and utility function.

Let $\texttt{CCP}(n^{\{S\}} \times n^{\{D\}})$ be a matrix of conditional choice probabilities, where each element $\texttt{CCP}_{jk}$ is the probability of choosing decision k given the state j. Let Q_d represent the action-specific state transition matrix. Then the unconditional state transition matrix $Q = \sum_d \widetilde{\texttt{CCP}(\cdot,d)} \odot Q_d$, where $\widetilde{\texttt{CCP}(\cdot,d)}$ is a $n^{\{S\}}$-times copy of the d^{th} column of $\texttt{CCP}$ (i.e., $\texttt{CCP}(\cdot,d)$) and its dimension is $n^{\{S\}} \times n^{\{S\}}$. We use $\odot$ to denote an element-by-element multiplication. By following Aguirregabiria and Mira (2007), we can compute the state-specific value function V as:

$$V = (I - \beta Q)^{-1}[\sum_d \texttt{CCP}(\cdot,d) \odot (\boldsymbol{u_d}(\alpha,\boldsymbol{\omega}) + \boldsymbol{\epsilon_d})], \tag{A.2}$$

where I is an $n^{\{S\}} \times n^{\{S\}}$ identity matrix, $\boldsymbol{u_d}$ is an $n^{\{S\}} \times 1$ action-specific period utility absent the random shock, and $\boldsymbol{\epsilon_d}$ is an $n^{\{S\}} \times 1$ action-specific random shock vector with a closed form under the logit assumption: $\boldsymbol{\epsilon_d} = \text{Euler's constant} - log(\texttt{CCP}(\cdot,d))$.

The action-specific value function $V_d(n^{\{S\}} \times 1)$ is:

$$V_d = \boldsymbol{u_d}(\alpha,\boldsymbol{\omega}) + \beta \cdot Q_d \cdot V. \tag{A.3}$$

Then the conditional choice probability is:

$$\texttt{CCP}_{jk} = P(d=k|S=j) = \frac{\exp(V_d(S))}{\sum_{k\in\{0,1,2\}} \exp(V_k(S))}. \tag{A.4}$$

With the calculated CCP matrix, we can use Equation (A.1) to conduct the overall likelihood function $L = \sum_i l(D_i, O_i)$ and estimate $(\alpha, \boldsymbol{\omega}, \boldsymbol{\lambda_2})$ by maximizing L. With the estimated $(\hat{\alpha}, \hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\lambda_2}})$, we can update $\texttt{CCP}, Q_d, Q$ and then redo the above procedure. We will stop the iteration until $\|b^{(n+1)} - b^{(n)}\| < \varepsilon_Q$, where n is the n^{th} iteration, $b = (\alpha, \boldsymbol{\omega}, \boldsymbol{\lambda_2})$.

To better illustrate our algorithm, we demonstrate the pseudo codes as follows.[15]

Algorithm 1 FHMM Model Estimation

1: **Inputs:**
  Observed states O_i and decisions D_i
2: **Initialize:**
  $\texttt{CCP}$, threshold δ, parameter set b
3: **while** $b^{\texttt{new}} - b > \delta$ **do**
4:   $b \leftarrow b^{\texttt{new}}$
5:   Compute Q_d given b
6:   $Q = \sum_d \widetilde{\texttt{CCP}}(\cdot, d) \odot Q_d$
7:   $V = (I - \beta Q)^{-1}[\sum_d \texttt{CCP}(\cdot, d) \odot (\boldsymbol{u_d}(\alpha, \boldsymbol{\omega}) + \boldsymbol{\epsilon_d})]$
8:   $V_d = \boldsymbol{u_d}(\alpha, \boldsymbol{\omega}) + \beta \cdot Q_d \cdot V$
9:   Update $\texttt{CCP}$ given V_d
10:   Compute likelihood function $l(D_i, O_i)$
11:   Solve $b^{\texttt{new}}$ to maximize $l(D_i, O_i)$
12: **end while**

Appendix B: Detailed Structural Model Estimation Results

Table 10 presents our estimation results based on the best model with the four engagement stages. For comparison purposes, we list all three sets of parameters (i.e., control, treatment with price promotion, and treatment with free content promotion) in the same table. In our utility function, the price coefficient is a user-inherent attribute and common across all stages. The with-promotion estimates are smaller than the without-promotion one in absolute value, indicating that consumers show smaller price sensitivity when promotions are available. The reading utility coefficient, by contrast, is stage-specific. With the assumption of non-decreasing reading utility values, we convert

[15] A sample code for demonstration purpose can be found here.

the estimated ω_e to the true reading utility coefficient $\tilde{\omega_e}$. In Table 10, we also report the estimated parameters in the engagement transition function. The threshold $h(e', e)$ is the e'-ordered logit threshold in state e. It means that if a user's propensity from reading experience is above the high threshold, she is very likely to move forward to a higher engagement stage. Among the four stages, we observe that thresholds in the aware stage (i.e., $e = 1$), on average, are higher than those in the other three stages, meaning that "aware" users are more stable and more likely to stay in their current stage. In the mobile app context, aware users are mostly new users, who might only randomly visit the app without much time investment. On the contrary, when users enter the "exploring" or higher stage, they are more familiar with the app and its content, which makes them more likely to frequently change engagement stages. Additionally, the stage-specific estimates in the search proxies also yield several interesting findings. For example, we observe that with respect to search behavior, both promotions have consistently positive effects on users in the exploring stage. In our design, both promotions allow mobile users to read more free-content units. For those users who have some interest in this app and are willing to explore more in it, these promotions provide free opportunities to do the search both broadly and deeply, which in turn, encourages them to become more engaged. The above finding, to some degree, verifies the four labels we use in the engagement stages. For example, we show users at the second (i.e., exploring) stage are affected more in their search traces.

Table 10: Structural Model Estimation Results

Variable	$e=1$ **(aware)**			$e=2$ **(exploring)**			$e=3$ **(active)**			$e=4$ **(addicted)**		
	C	**T1**	**T2**	**C**	**T1**	**T2**	**C**	**T1**	**T2**	**C**	**T1**	**T2**
Utility function												
Price coef. β	**-3.2037**	**-2.1508**	**-1.2828**									
	(0.7281)	(0.4285)	(0.8038)									
Reading coef.	**1.9302**	-0.3254	1.5528	-0.9887	**3.5839**	1.0035	2.4003	**-0.9543**	**-1.9730**	1.6145	**0.7396**	4.2936
ω_e	(0.2735)	(0.3768)	(2.2509)	(25.2055)	(0.6664)	(18.4537)	(2.8650)	(0.1781)	(0.4354)	(37.5083)	(17.7046)	(7.3071)
$\tilde{\omega_e}$	1.9302	-0.3254	1.5528	2.3022	35.6885	4.2806	13.3282	36.0736	4.4196	18.3535	38.1687	77.6456
Transition function												
Threshold 1	2.6234	**6.9377**	**2.2121**	**1.0053**	0.3395	**-0.8708**	-0.3030	**1.1272**	-0.7970	-0.6872	**-2.0898**	0.4658
(lowest)$h(1,e)$	(6.0706)	(1.0298)	(0.8036)	(0.5499)	(0.3814)	(0.5584)	(1.1666)	(0.7437)	(1.2343)	(1.1117)	(0.8740)	(1.1014)
Threshold 2	**2.8805**	**10.0570**	**6.2276**	**1.1199**	**0.9224**	**0.6427**	**-0.2710**	**7.8292**	**-0.6612**	0.0451	**-1.6359**	**1.1424**
(lowest)$h(2,e)$	(0.6271)	(1.8142)	(0.5529)	(0.0918)	(0.2694)	(0.4583)	(0.1230)	(0.8722)	(0.0044)	(0.1873)	(0.9463)	(0.2305)
Threshold 3	**6.1125**	**11.0255**	**9.5982**	**1.6274**	0.9551	**0.9989**	1.7408	**8.6835**	**-0.1209**	1.2213	**-0.4996**	**2.7855**
(lowest)$h(3,e)$	(0.7393)	(0.8566)	(0.5949)	(0.0249)	(1.3133)	(0.2738)	(2.0615)	(1.2384)	(0.0975)	(2.4374)	(0.3907)	(0.5361)
Popularity	-1.5863	**-3.5964**	**-4.0482**	-0.5405	**-1.3404**	-0.5405	-0.8935	**-0.5054**	**1.9529**	0.7988	**0.5755**	-0.1166
indicator $\mathtt{CF}_{\mathtt{Pop}}$	(4.1106)	(0.0208)	(0.8971)	(2.3454)	(0.1794)	(2.3454)	(0.9686)	(0.3972)	(1.1341)	(1.1278)	(0.4595)	(0.4168)
Fiction genre	-2.0593	**-1.5266**	**-2.2133**	**-1.9353**	**-0.1870**	-0.4899	**0.7910**	1.3902	**-1.3293**	**-0.5341**	**-0.2848**	**1.4096**
$\mathtt{CF}_{\mathtt{fiction}}$	(4.9272)	(0.2257)	(1.2648)	(0.1571)	(0.1536)	(1.7098)	(0.2806)	(1.8586)	(0.0029)	(0.0624)	(0.1328)	(0.2852)
Casual genre	-1.8086	**-3.3966**	**-2.5629**	**-1.0894**	**-2.5921**	**0.6402**	**-0.8833**	**1.3576**	**2.0758**	**-1.3952**	**-1.8131**	**-1.5115**
$\mathtt{CF}_{\mathtt{casual}}$	(2.6087)	(0.7750)	(1.2648)	(0.0419)	(0.6402)	(0.2570)	(0.3731)	(0.9071)	(0.9420)	(0.6238)	(0.3010)	(0.1336)
Depth search	-2.2873	**-5.0584**	-0.0964	-0.6191	**1.0612**	**-0.7925**	**-1.7369**	**2.2592**	**-0.8754**	0.4484	-0.8106	**1.3069**
indicator $\mathtt{SP}_{\mathtt{depth}}$	(3.3259)	(0.5520)	(0.2012)	(1.7392)	(0.3665)	(0.0962)	(0.5339)	(1.2827)	(0.0529)	(1.3038)	(0.8607)	(1.0942)
Breadth search	-0.1213	-0.4960	0.1351	**-1.5228**	0.5714	**2.7889**	0.1971	**-0.7638**	**0.2973**	**-0.6645**	1.4618	**-1.1439**
indicator$\mathtt{SP}_{\mathtt{breadth}}$	(1.6255)	(0.7533)	(0.8879)	(1.1173)	(1.8954)	(0.8726)	(0.3208)	(0.3062)	(1.7529)	(0.3386)	(1.5670)	(1.0992)

Note: standard errors are shown in parentheses.

Significant (95%) estimates are in bold.

C: control set without promotion; T1: treatment with price promotion; T2: treatment with free-content promotion

Price coef. β is not engagement-specific. We present only the estimates in $e=1$ cells.