Demand Dynamics in the “Rental-by-Mail” Business Model

June 2015

Abstract

We develop an integrated model for consumption and purchase decisions in the Rental-by-Mail (RBM) business model, which features closed loop delivery and quota-based subscription pricing. The model captures two distinct types of consumer dynamics: short-term dynamics in consumption decisions, and long-term dynamics in purchase decisions. Applying this model to a unique panel data provided by an RBM firm, we find that consumers trade off immediate and future consumption opportunities, accounting for the nontrivial turnaround time entailed by mail delivery. Furthermore, consumers trade off a higher payment and a larger set of consumption opportunities. The Bayesian estimation framework uncovers substantial heterogeneity in consumer preferences, such as price sensitivities and switching costs. We further conduct counterfactual analyses to investigate how consumers react to alternative product design and speed of delivery (or turnaround time) and the implications for firm profitability. Investigating the case of digital delivery, we find that firm profitability can be significantly reduced.

Keywords: Rental-by-Mail Business Model; Dynamic Consumer Model; Bayesian Estimation.
1 Introduction

The Rental-by-Mail, or RBM model is used for a wide array of rental products including movies (e.g., Netflix), books (e.g., BooksFree), art works (e.g., TurningArt), video games (e.g., GameFly) and toys (e.g., BabyPlays). Consumers have the convenience of receiving and returning rental products in the mail (versus traveling to physical stores), a deeper selection of rental products,\(^1\) with the tradeoff being a delay in obtaining the product. Adopters of the RBM model have seen great successes. Netflix alone serves over 20 million consumers in the U.S., and has delivered more than 2 billion DVDs via mail (Schonfeld, 2011). A prominent recent adopter of the RBM model is Rent The Runway, a fashion firm that grew to 5 million customers from 2009 to 2014 with its “unlimited” designer dress rental service. Table 1 presents the product offerings across a range of RBM services. In this paper, we develop an empirical framework to model consumer decision-making in the RBM setting, use data from a representative RBM service to recover consumer preferences, and examine the impact of a number of firm strategies in the setting through counterfactuals.

Table 1: Plans Offered by Representative RBM Services

<table>
<thead>
<tr>
<th>Service</th>
<th>Delivery</th>
<th>Product Selection</th>
<th>Subscription Plan</th>
<th>Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netflix</td>
<td>Mail</td>
<td>100,000 + movies</td>
<td>1 at a time</td>
<td>$7.99 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 at a time</td>
<td>$11.99 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 at a time</td>
<td>$19.99 per month</td>
</tr>
<tr>
<td>GameFly</td>
<td>Mail</td>
<td>8,000 + games and movies</td>
<td>1 at a time</td>
<td>$10 per 2 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 at a time</td>
<td>$20 per 2 months</td>
</tr>
<tr>
<td>Rent the Runway</td>
<td>UPS or courier services</td>
<td>1,000 + dresses</td>
<td>3 at a time</td>
<td>$99 per month</td>
</tr>
<tr>
<td>TurningArt</td>
<td>Mail</td>
<td>1,000 + works of art</td>
<td>1 at a time</td>
<td>$15 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 at a time</td>
<td>$20 per month</td>
</tr>
<tr>
<td>BookFree</td>
<td>Mail</td>
<td>250,000+ titles</td>
<td>2 at a time</td>
<td>$8 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 at a time</td>
<td>$10 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 at a time</td>
<td>$13.50 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9 at a time</td>
<td>$18 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12 at a time</td>
<td>$22.25 per month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15 at a time</td>
<td>$27.50 per month</td>
</tr>
</tbody>
</table>

The RBM model can be best characterized as having two unique features. First, unlike traditional brick-and-mortar rental stores, RBM services use mail or courier services to deliver rental

\(^1\)For example, 100,000 unique titles are available from Netflix, compared with ~8,000 in a typical brick-and-mortar Blockbuster store.
products directly to the consumer’s home and collect the rental products returned by the consumer. This key feature results in a closed loop rental process, in that the firm will ship a “new” product when it receives a product from the consumer. As we elaborate later, the closed loop rental process links consumers’ current and future consumption opportunities. Furthermore, many RBM services that specialize in information (digitizable) goods are also undergoing a switch to digital delivery of products through the Internet. Netflix, for instance, is pursuing a hybrid physical and digital delivery model, whereas others like Amazon offer only digital rentals.

Second, RBM services typically use a subscription-based pricing model rather than renting products a la carte. Specifically, the consumer chooses from a menu of subscription plans for a specified period, typically a month, with each plan featuring a price, and a quota, which is the number of rental products that she can check out at any time. These considerations lead to the following research questions that we examine in this paper.

1. What drives consumers’ consumption decisions in a closed loop RBM system? Do consumers smooth their consumption across time, or is consumption bunched? How do consumers balance current and future consumption?

2. What is the impact of operational characteristics (e.g., speed of delivery) on consumers’ consumption and purchase decisions, and on firm profits? Does the firm always benefit from an improvement in operational efficiency?

3. Given that RBM services usually offer a menu of subscription plans, to what degree do lower-end plans cannibalize higher plans? How should the firm set its menu to minimize such cannibalization? How should the firm set the prices of its plans?

4. If an RBM firm switches to digital delivery, what are the impacts on its revenue and profitability?
The consumer’s journey within the RBM model is detailed in Figure 1. A consumer initially signs up for a service plan (Step 0), after which the firm sends out the rental products to the consumer (Step 1). The consumer then receives the products (Step 2), and holds them for any desired time (Step 3) before eventually returning the products (Step 4). The firm receives the returned products (Step 5) and sends the same number of new products to the consumer. The total number of products in the rental process is determined by the chosen plan (e.g., 3 DVDs or 1 dress at a time) and remains constant during the subscription period (or payment cycle). Thus, the model features a closed loop rental process from Steps 1-5 until the consumer terminates the service. Note that if the consumer upgrades (downgrades) her plan, more (fewer) rental products will be made available to the consumer.

Any RBM consumer makes two types of decisions that have different frequencies. She first makes a purchase decision at the beginning of each payment cycle (e.g., a month), by choosing a subscription plan or exiting the service. Next, the consumer makes short-term (e.g., daily-level) consumption decisions on when and how many products to consume. Both types of decisions have significant implications for the firm. On one hand, the revenue of an RBM service is collected from subscription fees, and thus is completely determined by consumers’ plan purchase decisions. On the other hand, the operating costs (e.g., delivery and processing costs for the returned rentals)
of the service are driven by the consumer’s consumption decisions, but not purchase decisions. Intuitively, a consumer’s willingness to pay for each subscription plan is determined by her expected consumption throughout the payment cycle, linking the consumption and purchase decisions.

Both consumption and purchase decisions are affected by the two unique RBM characteristics discussed above. First, the delivery method entails a *turnaround time*, which consists of both the processing and delivery time. In Figure 1, the turnaround time is the time interval between Step 4 (when the consumer sends out consumed products) and Step 2 (when the consumer receives new products). Thus, while the RBM service gives consumers the convenience of receiving products at home, it also implicitly limits the number of movies they can watch in a subscription cycle based on the turnaround time, which separates the sequence of consumption opportunities for the consumer. Furthermore, consumers’ consumption decisions are intertemporally connected due to the closed loop rental and delivery process of the RBM model, which results in a conceptually distinct and interesting dependence between current consumption decisions and future consumption opportunities. We refer to this effect as *short-run dynamics* since these decisions are made over a short time horizon, i.e., within a payment cycle.

Second, the subscription pricing of RBM implies that the consumer’s purchase decisions affect the consumer’s choice set in the upcoming subscription period, albeit in an indirect way. The purchase decisions are temporally separated from the consumption decisions, because except for the subscription fee, there is typically no direct fee associated with the consumption (“unlimited rental”). Thus, the consumer must choose the best-suited plan, given her expected consumption choices. There are exogenous factors that determine the value consumers derive from consumption (and hence, the plans), e.g., due to a changing product assortment. We refer to the dynamics in consumers’ purchase decisions as *long-run dynamics* because they are made over a longer horizon, i.e., across payment cycles.

To summarize, consumer decisions are affected by the menu of service plans set by the RBM firm, as well as the turnaround time. Thus, the current inventory of products that the consumer holds, and the expected future product availability are significant factors in how consumers value the service, as well as their decision-making. From the firm’s perspective, it has the incentive and the ability to fine-tune its service menu and turnaround time.

To answer the research questions discussed above, we develop an integrated model for consumer decisions in the RBM setting, beginning from micro foundations. This model explicitly characterizes the consumer’s set of available consumption choices at any given time as determined not just by
the plan quota, but also the operational performance characterized by the turnaround time. It also recognizes the intertemporal tradeoffs inherent in the two types of dynamics mentioned above.

First, in the short-run dynamics, the consumer trades off immediate vs. future consumption, given the flexibility in consumption. While using all products immediately (e.g., watching all three movies on the same day) may lead to maximal immediate utility, the consumer will have to wait a full turnaround time for new movies to arrive, and thereby forfeit consumption opportunities in the near future. Second, in the long-run dynamics, the consumer trades off the subscription price and the additional flexibility in consumption (e.g., more movies in the mail), and this decision is impacted by the set of products available from the firm. Our model explicitly characterizes these two types of tradeoffs induced by the plan quota and turnaround time.

We take the model to a unique data set provided by an RBM service specializing in movie rentals, in order to uncover the consumer preferences that underlie the observed consumption and purchase decisions. The data contains a representative sample of consumers, for whom we observe monthly payments made, and detailed, daily-level movie shipping records. The data have a number of interesting patterns, all of which are captured in the framework developed in this paper. First, we find that consumers watch much fewer movies than the plan allows, foregoing immediate consumption opportunities. Second, consumption is more likely to occur when consumers hold a high level of inventory, everything else being equal. Third, consumers exhibit inertia in plan choice, and do not change their plans even when faced with significant exogenous changes that alter the value of the plans. Fourth, we find that the turnaround time has a significant effect on consumers’ consumption. All of these data patterns have significant implications for the revenue and operating costs of the firm.

Estimating the model with heterogeneous consumers brings a number of computational challenges. Since consumers in the data exhibit forward-looking behavior, modeling both the short- and long-run dynamics requires us to characterize a very large state space. Estimating such a model by iteratively solving the Bellman equation for each guess of the parameter values is computationally intractable. However, we adapt the recent estimation framework developed by Imai et al. (2009) (IJC), who suggest performing a single iteration of the Bellman operator for each Markov chain Monte Carlo (MCMC) guess of the parameter values, and storing a set of pseudo value function guesses to be leveraged for later iterations of the MCMC process. A major advantage of this approach is that it allows a flexible, hierarchical Bayesian model of individual heterogeneity.

We estimate multiple model specifications, and find that the heterogeneous forward-looking
consumer model with semiparametric features best fits the data, compared with a myopic or a homogeneous model, or a model with a parsimonious but more restrictive parametric consumption utility. We find evidence for both intertemporal smoothing, as well as bunching by consumers across days. Specifically, we find that consumers prefer to smooth consumption during the weekdays, but their preference for watching more movies is higher during the weekend, as we might expect. Consumers also value a larger content set provided by the firm. Moreover, we find that consumers incur a substantial switching cost in changing their current plans, even though the firm does not impose a fee for doing so. Finally, we uncover substantial heterogeneity across consumers across all utility parameters, e.g., the average switching cost in the estimation sample is about $27, but the individual-level estimates diverge substantially around this average, with a maximum of $75.

Based on the parameter estimates, we conduct several managerially relevant policy experiments as counterfactual scenarios, in order to investigate the potential incremental profit that can accrue from an informed design of the service menu and turnaround time. We find that the firm can consolidate its offerings and induce consumers to self-select into higher-margin plans; resulting in an increase in overall total profit by as much as 19.8%. We find that increasing operational efficiency (i.e., reducing the turnaround time) is a double-edged sword: while it can increase consumers' willingness to pay for the service, it also increases the firm’s operating costs because consumers are now induced to consume more movies. Thus, the net effect of reduced turnaround time is ambiguous. We find that reducing the turnaround time by one day can increase the revenue by 5.7%, yet the marginal cost rises by 9.5%. At the current marginal cost of $2.0, the firm can increase its profit by 3.5%. However, a sensitivity analysis shows that for a higher level of marginal costs, the firm expects a lower operating profit. Finally, we consider digital delivery, an alternative distribution method for delivering digitizable goods (e.g., movies, digital books and video games). In this setting, consumers have instant access to the products, which effectively reduces the turnaround time to zero. We show that with consumers' estimated preferences, a migration to digital delivery can significantly reduce the firm’s profit, and we uncover the mechanism by which this occurs.

The intended contribution of this paper is threefold. First, to the best of our knowledge, we are the first to provide an integrated model for consumers' dynamic consumption and subscription decisions in the Rental-by-Mail (RBM) business model. Second, we examine a number of counterfactual scenarios to investigate how firm profitability in the RBM setting varies as a result of marketing mix choices, and find a mix of product and pricing strategies that enhances overall profitability. Finally, we identify a negative impact of digital distribution on firm profitability and
the mechanism that underpins such an impact.

The rest of the paper is organized as follows. Section 2 presents a brief literature review. Section 3 describes the data, model-free and reduced-form evidence and the outline of the modeling framework. Sections 4 and 5 present the detailed model and discuss the results. Section 6 presents the counterfactual analysis. Section 7 concludes and indicates future research opportunities.

2 Literature Review

Our research is related to two separate streams of research in the literature: modeling demand for RBM products, and consumer dynamics in consumption and inventory for consumer packaged goods (CPG). Our empirical setting and modeling framework integrates aspects for both of these streams, and to our knowledge, is the first empirical study of the RBM model from microfoundations, and we simultaneously endogenize consumption and purchase decisions in the model.

The first stream of literature examines consumers’ demand and return behaviors for rental products, recognized as a key input that informs important strategic operational decisions of the rental firm (e.g., optimal stocking level). Studies have examined the impact of a pre-specified rental duration on optimal stocking, rental allocation among physical retail stores, and more recently, theoretical modeling of aggregate demand in closed loop RBM systems, incorporating design elements such as inventory or deadlines that can be used to avoid excessive peak demand (Bassamboo et al., 2009; Baron et al., 2011).

Our research differs from this stream of research in several ways. First, these studies focus on consumers’ demand of rental products, but do not model consumers’ purchase decisions, which we study. Second, while these studies recognize the importance of consumer demand, they are either theoretical or simulation studies, and rely on parametric assumptions about consumer demand without empirical support, and are made at the aggregate (e.g., store) level. These studies also do not consider the effects of firm’s marketing decisions (e.g., service plans offered) and operational decisions (e.g., delivery speed) on consumers’ demand for rental products. In contrast, we empirically analyze a unique and detailed panel data to derive insights into individual RBM consumers, We explicitly characterize how individual consumer preferences and firm decisions that lead to the observed demand. Finally, since our empirical model is based on micro foundations, we can study the profit impact of various design decisions made by the firm in our counterfactuals.

The second stream involves the extensive marketing literature on CPG goods (e.g., yogurt and
Consumption in both consumables and in our rental setting is subject to the current product inventory held by the consumer. Consumers balance utility from consumption with the costs of carrying inventory and stockout risk, resulting in a dependence of consumption on inventory. Whereas purchases in CPG settings are easily observable and captured, similar data on consumption is not typically available, since that would require monitoring the consumer’s consumption actions. Hence, early work in this stream began with the assumption that consumers have a constant usage rate (Bucklin and Lattin, 1991; Chintagunta, 1993; Tellis and Zufryden, 1995), which was later relaxed to more flexible specifications (Assuncao and Meyer, 1993; Wansink, 1996; Ailawadi and Neslin, 1998; Erdem et al., 2003; Sun, 2005), and the literature features both theoretical and empirical studies. Stockpiling and consumer inventory are known to impact consumption incidence and quantity across a wide range of product categories (Chandon and Wansink, 2002). Forward looking consumers respond to price promotions by stockpiling and increasing consumption across multiple CPG categories (Sun, 2005).[Yacheng: the sentence below is quite repetitive with “The main takeaway from these...” ] The main takeaway from these studies is that when consumption is flexible and intertemporal substitution is possible, it is critical to study consumers’ dynamic consumption and purchase decisions simultaneously in order to get unbiased parameter estimates (Ching and Osborne, 2015). For example, a static model would overestimate price sensitivities of forward-looking consumers (Hendel and Nevo, 2006). We further note that for most RBM services, consumption of rental products (e.g., movies) can be characterized as flexible, and intertemporal substitution is possible.

We model consumers’ consumption, as well as purchase decisions in a unified, utility-maximizing framework. However, there are two crucial differences between this study and the CPG literature. In the CPG literature, the actual consumption was not observed; and instead was inferred based on purchase data, whereas we have two separate data sets on consumption (as measured by products delivered and returned) and purchases (as captured by subscription data). Second, the setting we study is a closed loop, where making a consumption choice not only reduces the consumer’s current inventory, but also adds to future inventory for the consumer, since the firm sends a new product when it receives an old one. In contrast, in typical CPG settings, there is no explicit link between current consumption and future inventory. This closed loop aspect requires us to develop a new modeling framework, which is essential to accurately characterize consumer demand dynamics in the RBM setting.
3 Institutional Setting and Data Description

3.1 Data

An anonymous online movie rental service in the U.S. (the “focal firm”) provided the data on the condition of anonymity. The focal firm operates on the same business model as Netflix. It uses USPS first-class mail to send its subscribers DVD disks, along with a postage-prepaid envelope for the return of just-watched DVDs. The focal firm offers specialized niche content, and, with fewer than 100,000 subscribers, it is relatively small compared with Netflix. It offers four regular subscription plans, with mailing quotas of 1, 2, 3 and 5, respectively. We focus on the three most popular plans with quotas 1, 2 and 3, which jointly accounts for more than 98% of the purchase shares. We refer to these plans as the “Low plan” ($11.95 for a mailing quota of 1 DVD), “Medium plan” ($19.95 for 2 DVDs) and “High plan” ($29.95 for 3 DVDs), respectively. The same set of plans was offered during the entire observation period, and there were no changes in the price levels. The firm’s plan-switching policy states that the consumer may only change her plan choice at the beginning of each monthly billing cycle, but not within the cycle.

The data set contains detailed information about the payment and shipping records for a representative sample of consumers over an observation period of approximately two-and-a-half years. The payment history records, for each individual consumer, when a payment is made, as well as the amount of payment. Matching the sequence of payments made by the consumer and the menu of plans offered, we reconstruct the entire sequence of purchases (plan choice) of the consumer over time. We use the date of the first payment to determine when the consumer signed up for the service, and the last observed payment (if received before the end of the observation period) to determine when the consumer left the service. Each consumer is identified by a firm-assigned ID and a (partial) credit card number. We observe that among the consumers who cancelled their subscriptions, no one returned to sign up again.

The focal firm delivers to consumers nationwide from a single distribution center operated in a mountain state. Thus, there is a natural variation in the turnaround time across consumers, based on the distance between the distribution center and the consumer’s mailing address. Specifically, the firm characterizes the turnaround time as either 5 days for consumers who are geographically close, or 7 days for those who are located further away. The shipping history contains the rental records for every movie rented by a consumer during her subscription. Specifically, for each movie rental, we observe the exact date when it was sent to the consumer, and the date when it was
received by the firm. The rental records show that the focal firm was prompt in processing the
returned movies: in almost all (>99.5%) instances, the firm shipped the same number of movies
to the consumer on the next business day after it received the returned movies. Next, there is a
potential issue that if the consumer has not added a sufficient number of movies to the queue, the
firm would not be able to ship anything when it received a product from the consumer, leading to
the closed loop being broken. However, in practice we find that over 95% of the firm’s customers
(including our entire estimation sample) maintain a sufficiently large number of movies in their
queues, so that the closed loop is maintained.

Because the firm did not directly observe the exact date when the consumer received a movie, we
infer the date based on when a movie was shipped out by the firm and the one-way shipping time.
Similarly, the date of consumption is inferred based on the date when the movie was returned and
the one-way shipping time: this inference is based on the assumption that consumers will return
the movie immediately after watching it, which is typically made in the literature (Milkman et al.,
2009).

Combining both purchase and shipping records, we are able to construct the entire subscription
and rental history for each consumer. For any given day when the consumer is an active subscriber,
we know the plan she subscribes to, the number of days until the next payment, the number of
movies that a consumer has available for immediate consumption (consumer inventory), and the
number of movies in the mailing process, which includes movies the consumer will receive on the
next day, the day after, and so on. This information is summarized in Table 2.

Table 2: Summary Statistics for Data

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purchased mailing quota</td>
<td>1.96</td>
<td>0.36</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Number of consumer movie inventory</td>
<td>1.40</td>
<td>0.78</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Number of movies in the mail</td>
<td>0.56</td>
<td>0.74</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Consumption</td>
<td>0.101</td>
<td>0.350</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Low plan chosen</td>
<td>0.18</td>
<td>0.386</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Medium plan chosen</td>
<td>0.69</td>
<td>0.465</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>High plan chosen</td>
<td>0.13</td>
<td>0.337</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Turnaround time (consumer-level variable)</td>
<td>6.98</td>
<td>0.99</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Tenure in days (consumer-level variable)</td>
<td>281.0</td>
<td>232.1</td>
<td>30</td>
<td>993</td>
</tr>
</tbody>
</table>
We make a few observations in Table 2. First, the average number of movies in a consumer’s inventory is 1.40, or 71.4% of the average purchased quota (1.96), and the rest of the movies are in the mailing process. This shows that the turnaround time imposes a significant restriction on the consumer’s usage. Second, among the three plans, the Medium plan is the most popular, chosen in 69% of all time periods, followed by the Low plan (18% of all time periods) and the High plan (13% of all time periods). Third, the proportion of consumers who have a longer turnaround time of 7 days (52.7%) is about the same as those who have a shorter turnaround time of 5 days (47.3%). Finally, there is substantial variation in the number of days consumers subscribe to the service.

3.2 Reduced-Form Evidence

We now present model-free evidence on consumers’ consumption and purchase decisions over time, and in particular, how these decisions are affected by the mailing quota and turnaround time, the two key design variables of the RBM service.

Drivers for Consumption Decisions

We first examine a number of likely drivers for consumers’ consumption choices: consumer (movie) inventory, turnaround time, seasonality, content set, learning and forward-looking behavior.

Effects of consumer inventory and turnaround time on consumption rates. Figure 2 shows the distribution of the average daily consumption rates, conditional on (1) the number of movies available (consumer inventory) and (2) the individual-specific turnaround time. Note that turnaround times are based on the consumer’s home location and the distance from the distribution center, and can be treated as exogenous. We separate by the three types of plan choices, where the plan quota is 1, 2 and 3, respectively, and find two patterns. First, conditional on the turnaround time, the consumption generally increases with the number of movies in the consumer’s inventory (sub-figure b and c), but at a decreasing rate (sub-figure c). [Yacheng: we cannot say “consumption increases with inventory” for q=1] Second, comparing consumers with different turnaround times, we find that conditional on the plan choice and number of movies available, the longer the consumer needs to wait, i.e., a higher turnaround time, the lower the consumption. The same pattern is observed for all plan choices. These observations demonstrate that both the consumer’s inventory and turnaround time impact consumption decisions.
Effect of seasonality. Additionally, consumption utility can be driven by other extraneous factors, such as seasonality. To investigate whether there is seasonality in consumption, we plot the monthly average consumption for each of the 12 calendar months. As Figure 3 shows, the monthly average consumption is quite stable. Across all 12 months, the average consumption rate is 2.99 DVDs, and the standard deviation is 0.07. Because the standard deviation is small relative to the mean (coefficient of variation =0.024), we do not find seasonality to be an important driver for consumption rates.

Role of Learning. Empirical evidence from the past literature has suggested that consumers
may learn about the service over time, and multiple sources of learning may exist in other empirical contexts (see Ching et al. (2013) for an excellent review). Consumers may learn about the quality of the service, as well as their usage patterns and requirements (Iyengar et al., 2007; Gopalakrishnan et al., 2014). If there is substantial consumer learning, then we would expect consumers to be more likely to switch plans earlier in their tenure. We examine the percentage of consumers who either switched up or down in their plans for each of the first six months of their tenure with the firm. From Figure 4, we do not see a clear decreasing trend in the switching or exit probability across the first six months. Thus, we do not have definitive evidence that learning plays a significant role in this RBM setting. Since the data patterns indicating reduced switching over time are typical variations used to identify the degree of learning when no such reduction is present, this setting would not be able to identify any learning.
Forward-looking Behavior and tradeoff between intertemporal consumption opportunities. The closed loop delivery and the plan quota induce an implicit tradeoff between current and future consumption for the consumer. To see this, note that once a subscription plan is chosen by the consumer, the total number of movies checked out by the consumer is fixed (and equals the plan quota). Conceptually, the quota comprises of movies in the consumer’s inventory, which represents current consumption opportunities, and movies in the mail, which represents future consumption opportunities. It is important to make the distinction between myopic and forward-looking consumers because each has a different set of tradeoffs. A myopic consumer considers only current consumption opportunities, whereas a forward-looking consumer accounts for both current and future consumption opportunities. To examine which assumption is more consistent with the observed consumption decisions in the data, we run the following regression for consumer i’s consumption in period (day) t, $c_{it}$:

$$
c_{it} = \theta_0i + \theta_1 \text{inventory}_{it} + \theta_2 n_{arrivingsoon}_{it} + \theta_3 n_{arrivinglate}_{it} + \theta_4 \text{weekend}_{it} + \theta_5 n_{contentset}_{it} + \theta_6 n_{contentset}_{it}^2 + e_{it}
$$

The dependent variable is current period consumption ($c_{it}$), modeled as a function of the number of movies in the consumer’s possession ($\text{inventory}_{it}$), the number of movies that the consumer will
receive the next day \((n_{arrivingsoon}i_t)\), the number of movies that will arrive only after the full turnaround time \((n_{arrivinglate}i_t)\), and other movies that are in the intermediate stage of the delivery \((n_{arrivingintermediate}i_t)\). Note that unlike \(inventoryi_t\), none of \((n_{arrivingsoon}i_t, n_{arrivinglate}i_t \text{ and } n_{arrivingintermediate}i_t)\) is available for immediate consumption. Thus, these variables should have no effect on a myopic consumer’s consumption decisions. Forward-looking consumers, however, will account for the availability of future movies and will adjust their consumption decisions accordingly. For example, we expect that, holding consumer inventory constant, a consumer with movies arriving soon is more likely to consume, compared to the case where she has movies arriving later. The dummy variable \(weekendi_t\) is 1 if day \(t\) is either Saturday or Sunday, and zero otherwise: it captures the systematic difference in consumers’ consumption utility on weekends vs. weekdays. The \(n_{contentset}i_t\) denotes the number of unique movie titles available to the consumer in period \(t\). This variable and its squared term account for the potential effects of having a larger content set on the consumer’s average consumption.

Table 3: Daily Consumption Regression

<table>
<thead>
<tr>
<th>Dependent Variable: Number of movies watched in period (t)</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>(t)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.41e-2***</td>
<td>8.16e-3</td>
<td>6.63</td>
</tr>
<tr>
<td>Number in consumer’s possession</td>
<td>3.53e-2***</td>
<td>3.03e-3</td>
<td>11.63</td>
</tr>
<tr>
<td>Number arriving soon</td>
<td>2.83e-2***</td>
<td>4.44e-3</td>
<td>6.37</td>
</tr>
<tr>
<td>Number arriving late</td>
<td>-1.16e-2**</td>
<td>4.43e-3</td>
<td>-2.60</td>
</tr>
<tr>
<td>Number arriving intermediate</td>
<td>-6.52e-3*</td>
<td>3.32e-3</td>
<td>-1.97</td>
</tr>
<tr>
<td>Weekend</td>
<td>2.02e-3***</td>
<td>3.89e-4</td>
<td>5.19</td>
</tr>
<tr>
<td>Content set size</td>
<td>9.38e-5***</td>
<td>2.54e-5</td>
<td>3.69</td>
</tr>
<tr>
<td>Square of content set size</td>
<td>4.12e-8</td>
<td>2.21e-8</td>
<td>1.87</td>
</tr>
</tbody>
</table>

Note: Individual consumer fixed effects included

***: \(p<0.001\), **: \(p<0.01\), *: \(p<0.05\).

We find that the consumer movie inventory has a significant positive effect on consumers’ consumption decision \((\theta_1 = 0.035, p < .001)\). Consistent with forward-looking consumers, movies in the mailing process also have significant effects that conform with our intuition. For example, conditional on the same number of movies in one’s possession, a consumer is more likely to consume when she expects movies to be arriving soon \((\theta_2 = 0.028, p < .001)\). Conforming with our expectation, consumers are more likely to consume during weekends \((\theta_4 = 0.002, p < .001)\). The size of
the content set also has a positive effect on the consumption ($\theta_5 = 0.00009$, $p < .001$). However, the quadratic term of the content set size was not statistically significant ($\theta_6 = 4.12e - 8$, $p > .05$).

4 Model

In this section, we develop an integrated model of consumers’ endogenous purchase and consumption decision-making with RBM services. Our model accounts for two institutional facts. First, RBM businesses use subscription pricing, and the payment is made at the beginning of each cycle (typically at the beginning of the month). Thus, the choice set of the consumer differs, depending on whether it is a non-payment day (only the consumption decision is made), or a payment day (both consumption and purchase decisions are made). Second, conditional on the chosen plan (i.e., quota), the consumer’s feasible set for consumption also differs across days. For example, the consumer will not be able to watch any movie if nothing is available to her.

A consumer makes two decisions sequentially. She first makes a subscription decision, i.e., choosing a plan or discontinuing the service at the beginning of each subscription cycle, which is a month for the focal firm. After making her subscription decision, the consumer then makes daily consumption decisions on how many movies to watch throughout the subscription cycle. We denote the set of payment or billing time periods using $T_p$, and the set of weekend time periods (i.e., Saturdays and Sundays) as $T_w$. Consumers in the model are indexed by $i \in I$, plan choices by $q \in Q_t$, consumption choices by $c \in C_t$ and time periods by $t$. Plans in our setting correspond to a quantity indicating the maximum number of movies the consumer can hold at any time. The plan choice set $Q_t$ varies over time. If the consumer is in a payment period (i.e., $t \in T_p$), she can choose to change her plan, or leave the service: $q \in Q_t = \{0, 1, 2, 3\}$, where 0 represents the outside option. In a non-payment period, $t \notin T_p$, consumers cannot change their plan, so $Q_t$ is the empty set. For convenience, we drop the $i$ subscript for the consumer, though the billing periods are different.

The unit of analysis in the model thus incorporates both the plan and consumption choices available to a consumer at any time period.
4.1 Period Utility

The instantaneous period utility for consumer \( i \) in period \( t \) when choosing a decision \( d_{it} = (q_{it}, c_{it}) \) is denoted as \( u(d_{it}) \), given below.

\[
u(d_{it}) = \left( \sum_{k \in C_t} \alpha_k^i I[c_{it} = k] \right) + \alpha_{iw}^i c_{it} I[t \in T_w] + \alpha_{is}^i c_{it} \log(\omega_t) \]

\[
+ I[t \in T_P] \left( \alpha_{ip}^i P_{q_{it}} + \alpha_{iw}^i I[q_{it} \neq q_{i,t-1}] \right) + \epsilon_{idt}
\]  

The first term flexibly captures the utility associated with each consumption choice, e.g., \( k \in \{1, 2, 3\} \) for a consumer who has \( k \) movies in her possession. Thus, when a consumer watches \( k \) movies in period \( t \), she receives a utility of \( \alpha_k^i \).\(^2\) The next term captures the idea that consumers might prefer consuming during the weekend, given the recreational nature of the consumption; thus, we model consumers as receiving different (unrestricted in sign) utility during the weekends, i.e., \( t \in T_w \), where \( \alpha_{iw}^i \) indicates the consumer’s coefficient for weekend consumption. Finally, consumers obtain a higher value of utility from the movies they watch when they are making a choice from a larger content set, which is denoted by \( \omega_t \) and refers to the number of unique movies available from the firm in period \( t \). The specific log functional form allows for increasing utility from more content and has a diminishing marginal impact, which is commonly used in the literature (e.g., Shen et al., 2014). Note that content set utility is interacted with the consumption amount \( c_{it} \), so it only accrues when a consumer watches at least one movie in that period, i.e., \( c_{it} \neq 0 \). Higher positive values of \( \alpha_{is}^i \) would indicate consumers having higher marginal consumption utility when the content set is large, which can occur when a consumer is more likely to find a better match with her preferences from a larger content set. Consumer \( i \)'s choice \( c_{it} \in C_{it} \) is restricted to the choice set \( C_{it} = \{0, \ldots, x_{it}^0\} \), where the consumer can choose to consume a number of movies \( c_{it} \) between a minimum of 0, and a maximum of \( x_{it}^0 \), which denotes the number of movies in the consumer’s current possession. We elaborate \( x_{it}^0 \) in Section 4.2 below.

The next two terms in the instantaneous utility function are only present in the payment periods

\(^2\)We also estimated an alternative functional form with linear-quadratic consumption utility, please see results in \( \S 5 \).
(the beginning of the payment cycle), i.e., when \( t \in T_p \). The consumer chooses a plan \( q_{it} \) and the associated price \( P_{q_{it}} \), which is applicable until the next payment cycle.\(^3\) Thus, the disutility of paying is captured by the fourth term, with \( \alpha^p_i \) being the consumer’s price coefficient. Next, we model the idea that consumers face a switching cost whenever they choose a plan that is different from the previous choice, i.e., \( q_{it} \neq q_{i,t-1} \). This term captures the common feature of subscription services (and also our setting), such that if the consumer does not make an active plan choice change in the billing period, then she retains the previously chosen plan. Consequently, the switching cost could be interpreted as the time and effort cost of logging into the firm’s website in order to switch to a different plan, even though the firm does not explicitly charge a fee to change plans. Note that the consumer may discontinue the service by choosing plan 0, i.e., \( q_{it} = 0 \), and, following the logic above, will incur the switching cost as well. The coefficient \( \alpha^{sw}_i \) then denotes the switching cost of the consumer. Finally, \( \epsilon_{idt} \) is the idiosyncratic shock assumed to be distributed as a Type I Extreme Value, independent across consumers, choices and time periods: it is observable to the consumer, but not to the researcher.

We have examined the static instantaneous (period) utility of a consumer in the above discussion. If consumers were myopic (i.e., static utility maximizers), then they would make consumption and plan choices to maximize the function above. However, this does not account for the fact that the current decisions of consumers have a significant impact on their future utility. Indeed, the purpose for which consumers choose a more expensive plan in the current period is to enable a higher level of future movie consumption. In our setting, since purchases are determined by willingness to pay for a sequence of consumption opportunities, it is especially important to understand how consumption decisions are related to purchase decisions.

### 4.2 Intertemporal Tradeoffs in the Decision Process

Formally, the state for consumer \( i \) in period \( t \) is composed of four components, each with its own state transition process, and is defined as:

\[
s_{it} = (x_{it}, w_{it}, z_{it}, \omega_t)
\]

\(^3\)Other RBM services have used different subscription policies. For example, Netflix allows its subscribers to upgrade their plans at any time, yet only allows subscribers to downgrade their plans at the end of the month. Source: Personal communication with Netflix customer service on May 8, 2015. Our modeling framework can be easily extended to accommodate alternative subscription policies.
The components of the state variable are as follows: \( x_{it} \) indicates the mailing state of the consumer, \( w_{it} \) denotes the day of the week, \( z_{it} \) tracks the time (number of days) to the next payment date, when the consumer is allowed to change plans, and finally, \( \omega_t \) captures the content set available. The first three components of the state space undergo deterministic transitions, whereas the last state variable denotes the content set undergoes monotonically increasing and stochastic transitions. We examine the evolution of each component, in turn.

**Mailing State**

Consider the closed loop delivery process in which consumers watch new movies, return them and obtain new movies after the turnaround time. Formally, we define the turnaround time \( \tau \) as the time it takes for a consumer who mails a movie in period \( t \) to receive a new movie, i.e., in period \( (t + \tau) \). The turnaround time includes both the two-way mailing time and processing time. It is therefore exogenous to the consumer, and partly determined by the delivery service, e.g., the postal or courier service.

We introduce the mailing state \( x_{it} \), which fully characterizes the closed loop rental process and the consumer’s dynamically evolving consumption set. This vector details the state of each product (movie) through the process of obtaining a movie, watching it and returning it to the firm, which in turn processes the returned movie, and mails out the next movie to the consumer. The turnaround time critically determines the state space. We specify the mailing state for consumer \( i \) in period \( t \) as:

\[
x_{it} = \begin{pmatrix}
  x_{it}^0, x_{it}^1, \ldots, x_{it}^s, \ldots, x_{it}^{\tau_i}
\end{pmatrix}
\]  

(4)

where \( \tau_i \) is the turnaround time for consumer \( i \). Note that \( x_{it}^0 \) is the number of movies currently held by the consumer, whereas \( x_{it}^s \) denotes the number of movies that the consumer will receive in the future period \( (t + s) \), where \( s \in \{1, 2, \ldots, \tau_i\} \). Observe that the total number of movies across all of the states \( (x_{it}^0, x_{it}^1, \ldots, x_{it}^{\tau_i}) \) must be equal to the number of movies in the plan:

\[
\sum_{s=0}^{\tau_i} x_{it}^s = q_{it}
\]  

(5)

While the turnaround time is exogenous, the transition process for \( x_{it} \) is *endogenous* and determined by the consumption and plan choices made by the consumer. The state \( x_{it} \) evolves to \( x_{i,t+1} \),
based on the following law of motion when there is no change in plan:

\[
\begin{align*}
    x_{i,t+1}^0 &= x_{it}^0 + x_{it}^1 - c_{it} \\
    x_{i,t+1}^k &= x_{it}^{k+1}, \quad 1 \leq k \leq (\tau - 1) \\
    x_{i,t+1}^\tau &= c_{it}
\end{align*}
\]  

(6)

The first line of equation (6) states that the inventory \( x_{i,t+1}^0 \) that will be available to consumer \( i \) at \( t + 1 \) is the current inventory \( x_{it}^0 \), plus any movies that she will receive \( x_{it}^1 \), less her current period consumption \( c_{it} \), which is shipped back to the firm. The last line of equation (6) means that after the consumer returns the just-watched movies to the firm, she will receive the same number of movies, but only after the full turnaround time of \( \tau \) days. The middle \( (\tau - 1) \) lines of equation (6) have a straightforward interpretation: on day \( t + 1 \), movies in the mail are one day closer to be delivered to the consumer. Note that \( \sum_{k=0}^{\tau} x_{i,t+1}^k = \sum_{k=0}^{\tau} x_{it}^k = q_{it} \): in the closed loop RBM rental process, the total number of movies in the mail, plus the consumer inventory, is always equal to the quota on any given day.

To illustrate the dynamics of the mailing states, consider consumer \( i \) in Table 4. The consumer has a turnaround time of \( \tau = 5 \) days, and subscribes to a plan with 3 movies, i.e., \( q = 3 \). Suppose that the initial mailing state for the consumer is \( x_{it} = (2, 0, 0, 0, 1, 0) \), so that she currently holds 2 movies and expects 1 movie to arrive in 4 days. The rows of Table 4 illustrate how the state \( x_{i,t+1} \) evolves when the consumer chooses either not to watch any movies in the current period, i.e., \( c_{it} = 0 \), or watch one \( (c_{it} = 1) \) or two \( (c_{it} = 2) \) movies. For example, if \( c_{it} = 0 \) (no consumption), the inventory available to consumer \( i \) remains unchanged at 2 on the next day. Meanwhile, she is one day closer to receiving a movie in the mail \( (x_{i,t+1}^3 = 1 \) and \( x_{i,t+1}^4 = 0 \), vs. \( x_{it}^3 = 0 \) and \( x_{it}^4 = 1 \)).

Table 4: Illustration of the Mailing State Transition for \( \tau = 5 \) Days

<table>
<thead>
<tr>
<th>Initial mailing state ( x_{it} )</th>
<th>( x_{it}^0 ) (inventory)</th>
<th>( x_{it}^1 )</th>
<th>( x_{it}^2 )</th>
<th>( x_{it}^3 )</th>
<th>( x_{it}^4 )</th>
<th>( x_{it}^5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_{i,t+1} ) with ( c_{it} = 0 )</td>
<td>( 2 )</td>
<td>( 0 )</td>
<td>( 0 )</td>
<td>( 1 )</td>
<td>( 0 )</td>
<td>( 0 )</td>
</tr>
<tr>
<td>( x_{i,t+1} ) with ( c_{it} = 1 )</td>
<td>( 1 )</td>
<td>( 0 )</td>
<td>( 0 )</td>
<td>( 1 )</td>
<td>( 0 )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>( x_{i,t+1} ) with ( c_{it} = 2 )</td>
<td>( 0 )</td>
<td>( 0 )</td>
<td>( 0 )</td>
<td>( 1 )</td>
<td>( 0 )</td>
<td>( 2 )</td>
</tr>
<tr>
<td>( x_{i,t+1} ) with ( c_{it} &gt; 2 ) (infeasible)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
The above discussions illustrate that the consumption decisions are intertemporally linked due to the consumer’s current inventory level, which imposes two constraints for the consumer. The first and explicit constraint is that the inventory imposes a hard cap on current consumption. The second and less explicit constraint is through the tradeoff consumers must make between current and future consumption opportunities. Given the uncertainty in consumption, the consumer has to plan her consumption in order to optimize the total utilities from both immediate consumption and consumption in the near future. In order to capture consumers’ intertemporal tradeoffs, we need to understand how current choices made by the consumer will impact future choice sets. We conceptualize three different sources of intertemporal tradeoffs, expanded in detail below.

The first intertemporal tradeoff evaluated by the consumer is between watching the available movie(s) now versus later. The intuition becomes apparent as we consider a consumer with the smallest plan, \( q = 1 \) and a turnaround time of \( \tau_i = 5 \) days. When the consumption utility is low due to an idiosyncratic shock, the consumer is likely to derive higher utility from postponing consumption. In other words, waiting provides an option value for the consumer. Such an intertemporal tradeoff is driven by the uncertainty in consumption utility, which is not possible for the consumer to perfectly predict due to uncertain factors such as available time for consumption. Specifically, at time period \( t \), the consumer receives an idiosyncratic shock of \( \epsilon_{idt} \) if \( d = (j, c) \in Q_t \times C_{it} \). If the shock corresponding to \( c_{it} = 1 \) is sufficiently high, e.g., due to unanticipated time availability or a suitable occasion, then the consumer will have the following tradeoff. She could watch the movie in period \( t \), but would have to wait for \( \tau_i = 5 \) days for the next movie to arrive. Thus, even if she has a better occasion (i.e., a higher idiosyncratic shock) to watch movies during the days \( \{t + 1, \ldots, t + 5\} \), she will not have movies to watch. Thus, the consumer would want to wait for a sufficiently high level of shock to choose in order to consume within the current period.

Second, in addition to the above tradeoff, consumers with plans \( q > 1 \) account for the number of movies that are expected to arrive in the near future when making their consumption decisions. Consider a consumer who subscribes to a three-movie plan (\( Q = 3 \)) and two situations: (a) she either has a high current inventory (3 movies), with no movies in the mail or (b) a low current inventory (1 movie), with 2 movies scheduled to arrive in the mail after 5 days. Will she be equally likely to watch one movie in each of the situations (a) and (b)? We note first that she derives the same amount of immediate utility from immediate consumption under both (a) and (b). However, there is a difference due to the intertemporal tradeoff. If she watches a movie in scenario (a), she will still have 2 movies in her inventory if a sufficiently high consumption occasion arises the next
day. On the other hand, if she watches a movie in scenario (b), she will not have any movies to watch for the next 5 days until a movie arrives. Thus, she will likely miss high value consumption opportunities that occur in the near future.

More broadly, her consumption decisions will be different because they lead to different future options. Since consumers value the (discounted) future utility, as well as the immediate utility, it is rational for her to reduce consumption when the current inventory is low, and increase it when inventory is high. The above mechanism extends to consumers across multiple plans, but the level of shocks required to make a current consumption decision will be different across different states, as well as across different plans for the same consumer. Furthermore, the above arguments suggest that a shorter (longer) turnaround time reduces the value of waiting for a forward-looking consumer, and consequently increases (decreases) the likelihood for immediate consumption.

We have focused above on the two dynamic tradeoffs in consumption choices, which capture the short-run dynamics in our setting. The third and final tradeoff is embedded in the long-run dynamics characterizing how consumers choose subscription plans. In choosing a plan, consumers account for the expected sum of discounted utilities for consumption choices that are enabled by a specific plan. Thus, consumers who have a higher consumption utility will, everything else being equal, tend to choose higher plans in the model. Also, consumers who have a higher turnaround time will tend to choose higher plans because they will have more occasions where they will be waiting for movies to arrive and have no movies in their possession. Also, as the size of the content set changes, consumers will be more inclined to upgrade to higher plans, since they might obtain higher utility from consumption. Note that consumers have rational expectations about how the content set evolves over time. Finally, consumers face a switching cost in changing plans, which can cause them to continue with their current plan, even though a higher plan might become more attractive due to a larger content set.

**Weekend State**

We include a state variable that captures the day of the week. It is the day of the week variable \( w_{it} \in \{1, 2, \ldots, 7\} \), beginning with Monday (day 1), and ending with Sunday (day 7). Formally, the transition process for the day of the week is specified as:

\[
w_{i,t+1} = \begin{cases} 
    w_{it} + 1, & w_{it} < 7 \\
    1 & w_{it} = 7 
\end{cases}
\]

(7)
The intra-week dynamics are driven by the exogenous state of weekends versus weekdays. If consumers have a different (e.g. higher) utility for weekend consumption, then we would expect higher consumption during the weekends, compared with weekdays, for two reasons.

First, the effect of higher instantaneous utility during the weekends would lead to higher consumption probability, conditional on all other state variables. Second, consumers would reduce their consumption on weekdays to ensure that they have sufficient movies available to watch during the weekend, which is also impacted by the turnaround time.

**Plan Period State**

The next state variable is the time from the billing period \( z_{it} \), which takes values from the set \( z_{it} \in \{1, \ldots, \mathcal{T} \} \), where \( \mathcal{T} \) is the length of the billing cycle. Recall that \( T_p \) denotes the set of all payment or billing periods. In periods with \( z_{it} \neq \mathcal{T} \), consumers make only consumption choices, while in periods where \( z_{it} = \mathcal{T} \), they make choices on both consumption and the plan.

\[
z_{it} = \begin{cases} 
  z_{i,t-1} - 1 & t \notin T_p \\
  \mathcal{T} & t \in T_p 
\end{cases}
\]

**Content Set State**

The final state variable \( \omega_t \) denotes the size of the content set, or the number of movies available in period \( t \), and from the consumer’s perspective, it evolves exogenously as a random process. Note that \( \omega_t \) is the same for all consumers. Consumers form expectations about the stochastic evolution of the content set, and we specify this state variable by a probability distribution across a discrete number of content set sizes, \( \omega \in \{1, 2, \ldots, N_\omega\} \), and with a \((N_\omega \times N_\omega)\) probability transition matrix \( \Omega \). Consumers have rational expectations and will expect the content set size to evolve according to \( \Omega \).\(^4\) We let the content set directly impact the consumption utility in equation (2), so that consumers obtain an increasing utility with a larger content set, but with diminishing marginal utility.

Overall, we have four components of the state, with one endogenous state \( x_{it} \) and three exogenous states, and all except \( \omega_t \) evolve deterministically. The total dimension of the state space

\(^4\)We estimate \( \Omega \) separately from the data, and incorporate it into consumers’ expectations.
is the product of (1) the DVD mailing state, which characterizes the consumers’ DVD inventory and the number of DVDs at different stages of the mailing process; (2) the day of the week, which affects immediate consumption utility; (3) the day in the payment cycle; and (4) the size of the content set.

4.3 Intertemporal Utility Maximization

A forward-looking consumer solves the following problem:

\[
\max_{d_{\tau} \in \mathcal{Q} \times \mathcal{C}_{\tau}, \forall \tau \geq t} u_{it}(d_{it}, \epsilon_{it}) + E \left[ \sum_{\tau = t + 1}^{\infty} \delta^{\tau-t} u_{it}(d_{\tau}, \epsilon_{it}) \right]
\]

(9)

The dynamic program can be represented in terms of the Bellman equation of the ex-ante value function \( V \) as a function of the state variable \( s_{it} \), which is defined in equation (3).

\[
V(s_{it}) = E_{\epsilon} \left[ \max_{d \in \mathcal{Q} \times \mathcal{C}_{it}} \left( u_{idt}(\epsilon_{idt}) + \beta \mathbb{E}_{s_{it+1}|s_{it}} \left[ V(s_{i,t+1}) \mid s_{it}, d \right] \right) \right]
\]

(10)

Equation (10) embeds two key tradeoffs corresponding to short- and long-run dynamics. The first is the tradeoff between current and future consumption. Myopic consumers might choose to watch all movies when the immediate consumption utility for that option is the highest of all options, whereas forward-looking consumers might wait because they internalize the negative impact of watching all movies immediately, in that they would have nothing in the inventory until the turnaround time has elapsed. They will also make choices based on the mailing state, the weekend state and the content set size. Second, consumers make tradeoffs in choosing between high- and low-quota plans: high-quota plans charge higher prices, yet they not only allow consumers higher utility with more options for immediate consumption, but also diminish the likelihood that consumers will find themselves with no products during time periods with a high idiosyncratic shock. Both immediate utility and greater flexibility in consumption become even more important as the content set increases, and consumers obtain higher consumption utility due to the larger variety in product choices.
4.4 Identification and Endogeneity

The structural parameters in equation (2) jointly capture the value a consumer derives from the RBM service, and they are obtained as a result of our estimation. We now intuitively discuss how these parameters are separately identified and the variation in the data that aids in such identification, beginning with the consumption parameters that can be identified with daily consumption patterns. Note that the identification of the discount factor is well known to be generically confounded in dynamic discrete choice models without an exclusion restriction (Magnac and Thesmar, 2002). We do not attempt to estimate it, and rather set the daily discount factor at $\delta = 0.999$ for all consumers.

The average consumption frequency during periods when the consumer has at least one movie available identifies the coefficient of each of the consumption terms in the utility function corresponding to the number of movies consumed, i.e., $(\alpha_{1i}^1, \alpha_{1i}^2, \alpha_{1i}^3)$. Thus, if we only had consumers with $q = k$ in the data, we would only be able to identify terms up to (and including) $\alpha_{ki}^k$. Intuitively, more bunching in consumption would lead to a more positive coefficient for higher consumption terms, whereas intertemporal smoothing or spreading over time in consumption would lead to a negative value of the coefficient. It is helpful to note that whereas the identification for consumption utility factors is driven by short-term dynamics involving daily consumption choices, the price coefficient and switching costs are identified from long-term dynamics involving purchase decisions.

Next, for the weekend coefficient, $\alpha_w^i$, identification is provided by the difference in the probability of consumption during the weekends, compared with weekdays, with a more positive value of the coefficient, indicating a higher probability difference.

The variation in the frequency of consumption with exogenous changes in the content set provides identification for the coefficient of the content set $\alpha_{cs}^i$. If consumers increase their consumption frequency in response, this would lead to a more positive estimate for this coefficient. The implicit logic is that with a larger content set size, consumers have more choice, which would then lead them to obtain movies that are a better match with their preferences, and they are therefore more likely to consume or watch these movies whenever they have a chance.

The price coefficient is identified by the variation in consumer choice of the plan, relative to their long-term average consumption. For the switching cost coefficient, when we observe an exogenous change in the firm’s content set size, if we observe consumers switching to a higher plan with more movies, we would then find the switching cost to be low. In contrast, if we observe persistence in
the plan choice, even under significant changes in the content set, the switching cost would then be estimated to be higher in magnitude and negative in sign.

Price endogeneity is unlikely to be a first-order concern in our empirical context. First, according to the focal firm, the main rationale of setting price levels is to achieve a “sufficient” (as perceived by the management) segmentation among its subscribers with different usage rates and willingness to pay. Second, and most importantly, our focal firm did not change any of the subscription prices during the entire observation period. More specifically, having a constant set of prices would not lead to correlation or dependence between the idiosyncratic error term and the constant.

4.5 Estimation

The discrete-choice, dynamic structural model developed above captures both the static and intertemporal tradeoffs faced by consumers. The estimation of this model, however, is computationally challenging for three reasons. First, the value function is highly jagged due to the interaction of the short- and long-term dynamics of the RBM setting. Second, the dimension of the payoff-relevant state variables is very large. Third, the value function iteration takes longer to converge when the discount factor is closer to one, as in our case: since our time period is a day, this leads to a high discount factor (0.999), and the problem is further exacerbated.

We plot the value function in Figure 5 below, for a representative consumer (whose preference parameters are fixed at the posterior mean of the individual parameters) at the first day of the payment cycle, which is assumed to be a Monday. There are two subspaces, corresponding to the two levels of turnaround time: states to the left (right) of the first red solid line correspond to $T = 5\,(7)$ days, respectively. The black lines separate the states for the three subscription plans. Observe that the consumers’ value functions are higher for the lower turnaround time, reflecting the fact that the consumers derive less value from a higher turnaround time. Figure 5 clearly demonstrates that the shape of the value function is jagged.

Traditionally, the computational burden associated with estimating dynamic discrete choice models with large state spaces is due to the challenges of iteratively computing the value or policy function (Aguirregabiria and Mira, 2010). Furthermore, the value function corresponding to our state space above is not smooth, mainly due to the complexity of the fully characterized mailing states. There are a number of ways to alleviate the computational burden, specifically via the

---

5 Overall, there are $247 \cdot (N_s) \times 7 \times 30 \times 5 = 259,350$ states.
random grid approach of Rust (1997), or the interpolation method of Keane and Wolpin (1994). These methods, however, are likely to provide highly inaccurate estimates, since they attempt to smooth over the jagged value functions. Similarly, while we could have used two-step methods, most notably the conditional choice probability approach of Hotz and Miller (1993) and Arcidiacono and Miller (2011), again in the first stage non-parametric estimation of choice probabilities, we are not likely to observe data throughout the large state space. This would then necessitate either a kernel regression or interpolation, and would again result in the same type of inaccuracies. Any efforts to capture consumer heterogeneity would make the estimation process even more computationally demanding. For example, even if we use a simple discrete number of mass points to model heterogeneity, the number of parameters will increase threefold.

Figure 5: Value Function

We use a hierarchical Bayes (HB) estimation approach. Although HB methods have been commonly employed in Marketing (Allenby and Rossi, 1998), they are relatively rarely used for models with forward-looking consumers. The primary reason is computational, since in dynamic structural models, we need to typically solve the Bellman equation and obtain the value function by performing value function iteration to convergence for each value taken by the parameters in the estimation process. Bayesian methods typically require thousands of iterations across the
parameter space to achieve convergence, making it challenging to combine them. Recently, Imai et al. (2009) (IJC) proposed a novel and highly practical method to reduce the computational burden by interweaving the MCMC iteration with only one step of the value function iteration. They demonstrated that such an iterative process will converge to the posterior distribution and provide convergence of the value function, as well. The crucial aspect of their method is that value function iteration until convergence is not required at each MCMC iteration, and the algorithm efficiently uses past information to form approximations of the true value function.

Consumers are heterogeneous in their valuation for consumption quantity represented by parameters \((\alpha_1^i, \alpha_2^i, \alpha_3^i)\), as well as the price coefficient \(\alpha_p^i\) and the value for the content set represented by \(\alpha_{cs}^i\). We represent these heterogeneous coefficients by the heterogeneous hierarchical specification:

\[
(\alpha_1^i, \alpha_2^i, \alpha_3^i, \alpha_w^i, \alpha_{cs}^i, \alpha_p^i, \alpha_{sw}^i) \sim \mathcal{N}(\Delta'z_i, V_\alpha)
\]

Thus, the means of the distribution of coefficients can potentially depend on individual specific characteristics \(z_h\) with the coefficient \(\Delta\) and the covariance matrix \(V_\alpha\). The hierarchical priors are specified as \(\text{vec}(\Delta|V_\alpha) \sim N(\delta, A_\alpha)\) where \(A_\alpha = A^{-1} \odot V_\alpha\) and the prior on the covariance matrix is \(V_\alpha \sim \mathcal{IW}(\nu, I_{|\alpha|})\). We specify uninformative priors in our empirical implementation: although the Bayesian framework allows the researcher flexibility to incorporate prior information, we do not have additional information regarding consumer demographics or preferences beyond their consumption and purchase histories. We provide the detailed estimation procedure in Appendix A.

While the IJC algorithm employs smoothing using a kernel regression, it is important to note that this smoothing is across the parameter space, rather than the state space. Therefore, the jagged nature of the value function across the state space will not result in a problematic approximation for IJC, when compared with methods that interpolate the value function over the state space. Although the IJC algorithm substantially alleviates the computational burden and allows us to account for individual-level heterogeneity, the large state space in our research setting nevertheless renders the estimation time-consuming. In order to achieve a practical estimation time, we use a sample of \(N = 200\) consumers for the estimation. The estimation procedure was written in \(R\), and the computationally intensive value iteration and likelihood functions were separately coded in \(C++\) (using the \texttt{RcppArmadillo} package) to speed up the estimation. Eventually, a single iteration of the IJC algorithm requires approximately one minute on a 32-core Amazon cloud computer.\(^6\)

\(^6\)We ran \(R = 10,000\) iterations for all models. Figure 10 in Appendix B plots the pseudo-likelihood over the \(R\)
5 Results

5.1 Parameter Estimates

We estimate four alternative models to see whether our proposed model provides a good framework for the consumers' consumption and purchase behavior. The first model is a myopic model. We use a discount factor of 0— that is, the consumers are assumed to be maximizing utility from immediate consumption, and oblivious to the possible loss of future consumption utility. The second model assumes consumer forward-looking, but fails to acknowledge consumer heterogeneity and treats all consumers as homogenous. Models 3 and 4 incorporate both forward-looking and consumer heterogeneity. While Model 3 uses a semiparametric utility function, Model 4 uses a linear-quadratic formulation for the parametric forms (linear or quadratic) of consumption utility. The linear-quadratic utility function is a popular choice model for consumers' usage decisions in other settings with subscription plans (e.g., Iyengar et al., 2007; Lambrecht et al., 2007).

Following the convention in the literature (e.g., Rossi et al., 2012), we use log-marginal density ($LMR$) as the basis for model comparison. In the last row of Table 5, we report the $LMRs$ of the four competing models. The dynamic, semiparametric model significantly outperforms all three competing models, based on log-marginal densities. In other words, it is important to allow for an appropriate specification of consumption utility, forward-looking consumers and consumer heterogeneity in modeling consumers' joint consumption and purchase decisions for our focal RBM services. Further comparisons of the log-marginal densities suggest that it is most important to account for forward-looking behavior, followed by consumer heterogeneity and the flexible specification of consumption utilities.

We summarize the model estimates in Table 5 below, focusing on the means and standard deviations of the posterior distributions of the parameters.

---

7 The myopic model has the worst fit among all four models ($LMR=-99,886.9$), which is not surprising in light of the reduced-form evidence that suggests consumer forward-looking. The improvement of model fit from the myopic model to the two dynamic specifications suggests that a forward looking model better explains consumers' consumption and subscription decisions. The model with homogeneous consumers has a better fit than the myopic model ($LMR=-35,414.8$). Between the different functional forms for consumption utility, we find that the semiparametric specification ($LMR=-28,665.5$) fits better than the linear-quadratic model ($LMR=-34,599.7$).
Table 5: Parameter Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Myopic</td>
<td>Homogeneous</td>
<td>Dynamic (Semiparametric)</td>
<td>Dynamic (Linear-Quadratic)</td>
</tr>
<tr>
<td>Consumption ($\alpha^1$)</td>
<td>-3.44</td>
<td>-1.87</td>
<td>-2.36</td>
<td>-2.89</td>
</tr>
<tr>
<td></td>
<td>(1.29)</td>
<td>(0.24)</td>
<td>(2.77)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Consumption ($\alpha^2$)</td>
<td>-4.79</td>
<td>-2.35**</td>
<td>-2.51</td>
<td>-2.51</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(0.42)</td>
<td>(5.95)</td>
<td>(5.95)</td>
</tr>
<tr>
<td>Consumption ($\alpha^3$)</td>
<td>0.37</td>
<td>1.19</td>
<td>1.24</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>(13.22)</td>
<td>(10.7)</td>
<td>(36.62)</td>
<td>(36.62)</td>
</tr>
<tr>
<td>Linear consumption ($\alpha^l$)</td>
<td></td>
<td></td>
<td>-2.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.68)</td>
<td></td>
</tr>
<tr>
<td>Quadratic consumption ($\alpha^q$)</td>
<td></td>
<td></td>
<td>0.544</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.32)</td>
<td></td>
</tr>
<tr>
<td>Price ($\alpha^p$)</td>
<td>-1.01</td>
<td>-0.99</td>
<td>-1.138</td>
<td>-0.472</td>
</tr>
<tr>
<td></td>
<td>(0.74)</td>
<td>(0.38)</td>
<td>(1.04)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Weekend ($\alpha^w$)</td>
<td>0.126</td>
<td>0.135</td>
<td>0.208</td>
<td>0.159</td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td>(0.05)</td>
<td>(0.65)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>Content set ($\alpha^{cs}$)</td>
<td>0.006</td>
<td>0.0238</td>
<td>0.161</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.02)</td>
<td>(0.95)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>Switching cost ($\alpha^{sw}$)</td>
<td>-40.56</td>
<td>-24.52</td>
<td>-36.69</td>
<td>-32.51</td>
</tr>
<tr>
<td></td>
<td>(16.82)</td>
<td>(7.55)</td>
<td>(26.68)</td>
<td>(31.72)</td>
</tr>
<tr>
<td>Log marginal density</td>
<td>-99,886.9</td>
<td>-35,414.8</td>
<td>-28,665.5</td>
<td>-34,599.7</td>
</tr>
</tbody>
</table>

*Note: For heterogeneous parameters, it is the posterior mean of the hierarchical (population) parameter.*

Examining the consumption parameters ($\alpha^1, \alpha^2, \alpha^3$), it is apparent that the consumers have divergent preferences in the utilities derived from the three levels of consumption. It is interesting to note that each of the population averages of the consumption coefficients ($\alpha^1, \alpha^2$) is negative. While the negative signs may initially appear surprising, it is consistent with the frequency of consumption choices. We model consumption at the daily level, and negative parameters reflect the data that in most periods (days), consumers chose not to watch movies, even when they have movies in their inventory. This pattern is due to the higher utility they place on the outside option, which incorporates alternative, non-movie-watching activities. Thus, the consumer would only choose to consume in periods when they receive a high enough positive utility shock. The population average of the third coefficient, $\alpha^3$, is positive, and the range is much larger than those of ($\alpha^1, \alpha^2$), likely due to the relatively low frequency of occasions where consuming three movies is possible (i.e., the consumer subscribed to the High plan, and has all three movies in inventory). Finally, it is interesting to note that the myopic model on average produces smaller estimates of consumption.
utilities at all three consumption levels, compared with the dynamic model. Intuitively, the myopic model ignores the effect of current consumption on the future consumption set, and attributes the observed low consumption rates exclusively to low consumption utility, whereas the forward-looking model would internalize the additional cost of not having future inventory due to current consumption.

The average weekend effect is positive, suggesting that consumers on average prefer to consume on weekends, compared with weekdays. This is consistent with prior expectations, yet we notice that such a coefficient is negative for many consumers, who possibly have the flexibility to watch movies during the weekdays, yet prefer to engage in other types of activities during the weekend. Combining the parameter estimates for the consumption coefficients discussed above, an interesting point to note is that the positive weekend consumption effect partially offsets the average negative consumption utility. Specifically, consumers during weekday periods do not necessarily obtain a higher utility from watching more movies, i.e., the coefficients are ordered as $\alpha_3 > \alpha_1 > \alpha_2$, adding further support to the value of the outside option, i.e., non-movie-watching activities. However, the utility ordering between one and two movies is reversed during weekends, when watching two movies is preferred to watching one, provided that the consumer has multiple movies in her possession ($\alpha_3 + 3\alpha_w > \alpha_2 + 2\alpha_w > \alpha_1 + \alpha_w$).

The average effect of content size is positive, suggesting that the average consumer derives additional value from a larger content set, as one might expect, due to a better match of content to her preferences. The price coefficient is identified from the variation in the expected sum of a stream of consumption utilities, compared across the different plans available, and the population price coefficient is estimated to be negative.

The switching cost is also estimated to be negative and significant, indicating that consumers face a nontrivial cost (not monetary) when choosing to change plans. To better understand the magnitude of switching costs, we combine estimates for price sensitivities and switching costs to compute the monetary equivalence of the switching cost by normalizing it with respect to the price coefficient ($\frac{\alpha_{sw}}{\alpha_p}$), shown in Figure 6. Among the consumers in the estimation sample, the average switching cost is $27.80, and the highest switching cost is $74.70. The magnitude of switching cost is high, yet in line with previous research has demonstrated that high switching costs can be attributed to consumers’ inertia, which can occur due to a number of factors, including expectation of transaction costs, uncertainty in the effort required to switch, etc. (Goettler and Clay, 2011).

Figure 11 of Appendix B illustrates the distributions of the individual-level posterior means
for each of the seven parameters of the dynamic, semiparametric model, with both the population mean (solid red line) and the median (dotted blue line) shown. We observe that overall, there is significant individual-level heterogeneity, which demonstrates the value of the IJC method in the sense that such heterogeneity cannot be captured easily by alternative ways of modeling consumer heterogeneity, e.g., with discrete segments. Distributions of the weekend and content coefficients are both bell-shaped. Figure 11 also shows that the distribution of the price coefficient is left-skewed, indicating a small percentage of highly price-sensitive consumers. All individual-level estimates of price coefficients are negative, even though we did not place any constraints on the signs of the price coefficient (i.e., they are estimated freely).

5.2 Demand Elasticities

Since RBM plans routinely offer a menu of service plans (e.g., Table 1), it is key to understand how consumers consider these plans simultaneously. A meaningful starting point is to examine the purchase elasticities at the current levels of prices. Using the parameter estimates from the dynamic semiparametric model, we compute the change in purchases of all three subscription plans with respect to a small change in the subscription price of each plan. Table 6 details the elasticities, where the first row indicates the change in purchases with respect to a price increase of the Low
plan, and so on.

<table>
<thead>
<tr>
<th></th>
<th>Low plan</th>
<th>Medium plan</th>
<th>High plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low plan</td>
<td>-0.047</td>
<td>0.107</td>
<td>-0.137</td>
</tr>
<tr>
<td>Medium plan</td>
<td>0.023</td>
<td>-0.225</td>
<td>0.015</td>
</tr>
<tr>
<td>High plan</td>
<td>-0.000</td>
<td>0.180</td>
<td>-0.182</td>
</tr>
</tbody>
</table>

There are a few noteworthy observations here. First, we find that all own-elasticities are negative, confirming our expectation that consumers reduce their purchases due to a small price increase for a plan. Second, own-price elasticity is larger in magnitude for the Medium plan, compared to the High plan, consistent with consumer heterogeneity and sorting, i.e., consumers who are less price sensitive choose the High plan.

Third, when we examine the cross-elasticities, we find that some of the plans exhibit positive cross elasticities, whereas others show negative effects. We find that surprisingly, the price increase of the Low plan has a positive impact on the Medium plan, but a negative impact on quantities of the High plan. These cross-elasticities suggest that consumers view the Low and Medium plans as substitutes. However, for the High plan, the reasoning is more nuanced. With a price increase for the Low plan, few consumers switch to the High plan, but more consumers exit the RBM service. If these consumers had not exited, some proportion of them would go on to upgrade to the High plan over time, as the content set increases. Since they are now “lost,” the impact on the High plan due to an increase in the price of the Low plan turns out to be negative. Thus, our framework is able to capture nuances of dynamics in consumer behavior, which is not apparent without careful modeling.

Next, we examine the impact of a price increase of the Medium plan, the cross price effects are positive, which reflects the typical substitution pattern. Finally, for a price increase in the High plan, we find that it has a negligible impact on the quantity of the Low, but a positive impact on the Medium plan. Overall, these effects reflect the multiple mechanisms that underlie consumer choices across the plans, and suggest that the firm could benefit from careful design of the plans.

6 Counterfactual Analysis

Results from the estimation of the proposed model above provides insights into consumers’ consumption and subscription decisions under the RBM business model. We show that both decisions
are not only driven by levels of subscription fees, but also by the turnaround time, as well as switching costs and the content set. In this section, we examine a number of counterfactual scenarios. First, we evaluate alternative pricing and product mix decision of the RBM firm. Given our findings that it is not appropriate to treat the different plans as being primarily substitutes in the presence of switching cost, we examine how to optimally set prices for the different plans, as well as determine whether it is best for the firm to offer all three plans for long-term profits.

Next, we examine how consumers respond when the turnaround time is reduced, and its impact on firm profitability. Intuitively, a longer turnaround time implies a longer wait for the rental products, which lowers the value of the service for the consumers. This intuition is confirmed in the model-free evidence in Figure 2, and is formally modeled to explicitly recognize that a longer turnaround time imposes a stricter constraint on the consumption sets available to the consumers. This dependence on turnaround time is also evident from the comparison between the value functions in the two sub-state-spaces corresponding to shorter and longer turnaround times, illustrated in Figure 5.

Finally, in the third counterfacutal, we examine how consumers might respond in the case of digital distribution of movies, or streaming online. Many firms like Amazon and Apple offer consumers the option to rent a movie that is digitally delivered. We investigate how our focal firm’s consumers would respond if the firm is able to switch to digital distribution.8

6.1 Counterfactual 1: Price Optimization

We first examine the degree to which the focal firm can increase the profit by changing the current subscription prices.9 The elasticities obtained above suggest that the Low and Medium plans are likely substitutes for the consumers, and from the firm’s viewpoint, result in cannibalization, i.e., consumers who might have been willing to pay for a higher priced plan might switch to a lower priced plan.

To obtain the optimal menu of plans for the firm, we focus on firm profits as a function of the plan design. Management of the focal firm indicated that it costs the firm a marginal cost of $MC = 2.0$ for each DVD sent to and collected from its subscribers. To validate and generalize this cost estimate for DVD rental businesses, we obtained additional information from public sources,

---

8 An important caveat of these analyses is that we do not explicitly account for the effect of potential change in the marketing mix on customer acquisition; given the directional change in prices and turnaround time in our counterfactuals, we are likely to understate the advantage of reduced prices.

9 The prices are set at $11.95 (Low plan), $19.95 (Medium plan) and $29.95 (High plan)
and find it to be consistent with the estimated marginal cost of Netflix.\textsuperscript{10} Subsequently, we compute the profit for the firm based on the costs and consumption behavior.

The counterfactual exercise is implemented as follows. For each of these consumers in the estimation sample, we simulate her consumption and subscription decisions iteratively. For each time period, we observe the consumer’s current state variables \((w_{it}, z_{it}, x_{it})\). We use the individual-level parameters to compute the expected utilities for each of the options. We then apply the multi-logit formula to compute the probability for each of the choices, as well as the outside option, where consumption is fixed at zero. These choice probabilities allow us to simulate her consumption and purchase decisions (if in a payment period).\textsuperscript{11}

The implementation of this counterfactual is computation-intensive, because the value functions must be re-computed for each set of alternative prices and for each consumer. Thus, we break down the search for optimal prices in two steps. In the first stage, we conduct a grid-search within the neighborhood of the current price vector. Figure 7 shows the results: each of the sub-graphs holds one price fixed at the optimal (from a grid search) level, while varying the other two prices. In doing so, we choose a relatively wide range of prices (from 50% to 150% of the current plan prices) for each of the three plans. The first sub-figure in Figure 7 shows that when the price of the Low plan is fixed and the prices for the Medium and High plans are varied, the highest profit is achieved at a pair of prices that are strictly within the search regions - in other words, we have a pair of interior solutions. The second sub-figure shows that when the price of the High plan is fixed, the profit monotonically increases with the the price of the Low plan. This monotonic pattern suggests that the firm can increase profit by making the Low plan less attractive to the consumers. There are two possible explanations to this finding. First, the Low plan is significantly less profitable than the Medium and High plans; thus, we find that the Low plan, if offered in the first place, would cannibalize Medium plan, hurting firm profitability. Second, since consumers’ switching costs are quite large, those who subscribe to the Low plan are also less likely to switch to the Medium and High plans, further contributing to a reduction in profitability.

In the second stage, we use the optimal set of prices obtained from the grid search as the starting point for further search based on the Newton-Raphson algorithm across the price vector. The final


\textsuperscript{11}We ensure that her consumption decision is subject to the inventory level specified in \(x_{it}\), and the purchase decision is only possible in a payment period, i.e., at the beginning of the month. After her choices are drawn, we evolve the state variables for the next period\((w_{i,t+1}, z_{i,t+1}, x_{i,t+1})\). We continue this procedure until the outside option (i.e., dropped out) is drawn, at which point the consumer’s tenure, total revenue, cost and profit are recorded; then, we move on to the next consumer.
results are shown in Table 7. We find that consistent with the grid search results, it is optimal for the firm to effectively eliminate the current Low plan and offer only two plans - we refer to them as the “new Medium” and “new High” plans. The profit-maximizing price of the new Medium (new High) plan is slightly lower (higher) than the current Medium (High) plan. These results suggest that the Low plan is cannibalizing the other two plans, especially because consumers who subscribe to the Low plan when the content set is smaller, might want to upgrade to a higher plan later when the content set increases, but the switching cost might discourage such a change. Further, we find that the firm should increase the difference in the plan prices to better differentiate across consumers with different price sensitivities. Overall, when the marginal cost of operation is $2.0, the firm can use the optimal menu of plans to increase the per-consumer profit across the approximately two-year counterfactual period from $127.1 to $152.3, a 19.8% increase.\footnote{An important caveat is that the counterfactual exercise is based on existing customers. Changing the price may affect the customers’ willingness to join in the place. Thus, the results must be interpreted with caution.}
### Table 7: Counterfactual for the Optimal Plan Design

<table>
<thead>
<tr>
<th>Plan (Mailing quota)</th>
<th>Current prices</th>
<th>Costs</th>
<th>Revenue</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>$11.95</td>
<td>$10.4(17.3%)</td>
<td>$34.6(19.2%)</td>
<td>$24.3(19.4%)</td>
</tr>
<tr>
<td>Medium</td>
<td>$19.95</td>
<td>$39.0(64.9%)</td>
<td>$119.5(65.9%)</td>
<td>$80.4(66.0%)</td>
</tr>
<tr>
<td>High</td>
<td>$29.95</td>
<td>$36.0(19.9%)</td>
<td>$22.4(19.4%)</td>
<td></td>
</tr>
<tr>
<td>All three plans</td>
<td>$63.0(100%)</td>
<td>$190.1(100%)</td>
<td>$127.1(100%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Optimal prices</th>
<th>Costs</th>
<th>Revenue</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Medium</td>
<td>$18.00</td>
<td>$30.0(58.3%)</td>
<td>$159.1(78.5%)</td>
<td>$129.1(85.4%)</td>
</tr>
<tr>
<td>New High</td>
<td>$32.55</td>
<td>$22.8(41.7%)</td>
<td>$46.0(21.5%)</td>
<td>$23.3(15.4%)</td>
</tr>
<tr>
<td>All two plans</td>
<td>$52.8 (100%)</td>
<td>$205.1(100%)</td>
<td>$152.3 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

### 6.2 Counterfactual 2: Effect of the Reduced Turnaround Time

In the second counterfactual, we examine the case where the firm reduces the turnaround time. We aim to quantify the effect of such a change on consumer choices, and hence firm revenues and costs, which can help the firm to evaluate the expected profitability from reducing turnaround time - an endeavor that likely involves costly investments (e.g., building more distribution centers that are located closer to the consumers), but nevertheless a feasible strategic option for the firm. We examine the change in revenue, cost and profit of our focal firm in the case where it reduces the turnaround time by one day, that is, consumers who currently have 5-day (7-day) turnaround time now have a 4-day (6-day) turnaround time.

The profit impact of turnaround time are not intuitively obvious. To see this, consider the following two opposite effects. On the one hand, it will increase the consumers’ willingness to pay for the RBM service, enabling the firm to collect more revenue. On the other hand, it will also increase consumption rates, thereby incurring additional marginal costs for the firm. The key question is whether the incremental revenue can compensate the marginal cost increase. This question can be precisely answered by our model, which enables us to compute the additional willingness to pay for a shorter wait. Note that because of the separation of consumption and payment, the change in consumers’ willingness to pay is independent of the marginal costs. Thus, conditional on a specific reduction in the turnaround time, the tradeoff between revenue and marginal costs depends on the magnitude of the marginal costs. Intuitively, reducing turnaround time is less likely to be profitable for firms with higher marginal costs, and vice versa.

We investigate this intuition in the counterfactual exercise. Specifically, we compute the firm’s total revenue, cost and profit for marginal costs ranging from $0.4 to $6.0. Figure 8 illustrates
the changes in total marginal cost (red dotted line) and revenue (blue dotted line). We make two observations. First, at the current estimate of marginal cost ($2.0), the total revenue and costs of the focal firm increase by 5.7% and 9.5%, respectively. The total profit increased by 3.5%. Thus, the focal firm can gain modestly from increasing its operational efficiency. Second, reducing turnaround time increases profit only if the marginal cost is not too large (specifically, no greater than $3.2). Third, when the marginal cost is small (e.g., $0.8), the increase in profit becomes much more substantial (e.g., 6.9%). This result intuitively suggests that RBM businesses with lower marginal costs (e.g., DVDs rental services) stand to gain more from building an expansive network of distribution centers, compared with RBM businesses with higher marginal costs (e.g., Rent The Runway). Together with the results in the previous two counterfactual analyses, we conclude that it is important to consider both the marketing and operational aspects of an RBM business.

Figure 8: Effects of Reducing Turnaround Time

---

13 Here, we do not consider the fixed costs that needs to be incurred to reduce the turnaround time, which can be a substantial investment. Instead, this result is aimed at helping the firm to evaluate the return of such an investment.
14 We refrain from conducting a counterfactual “throttling” exercise, which can have implications beyond what is captured by our model. In 2004, Netflix faced a class-action lawsuit, which alleged that Netflix may have been practicing “throttling” by increasing the turnaround time for some of its customers. Source: http://www.nbcnews.com/id/11262292/ns/business-us_business/t/frequent-netflix-renters-sent-back-line/, retrieved on October 27th, 2014.
6.3 Counterfactual 3: Digital Delivery (Online Streaming)

A key component of the RBM service is the mail or courier delivery service. While many RBM services will continue to rely on mail or courier services for products that must be delivered in physical forms (e.g., designer dresses, baby toys), recent technological advances have enabled traditional RBM services (e.g., Netflix and GameFly) to stream digitizable products (e.g., movies and video games) to streaming-compatible devices. For example, Netflix has become the largest online movie and TV show streaming provider, with more than 40 million streaming members as of 2013. There are other firms like Amazon and Apple, which deliver movies via digital distribution for a rental price in the range of $3-7. Thus, it becomes important to assess the impact of a migration to the digital distribution channel for both consumers and the firm.

Firm profitability can increase or decrease when using digital distribution. First, observe that digital delivery or streaming increases consumers’ ability to use the service on every single occasion, and there are no inventory stock-outs experienced by the consumer. In contrast to the RBM model, where forward-looking consumers forgo consuming all products in a single period to avoid being left with no products for the duration of the turnaround time, in digital delivery there are no such impacts. Thus, firms are arguably creating more value for the consumer, which can increase their willingness to pay. However, it may reduce the consumers’ incentive to purchase a large service capacity as a method to cope with a long turnaround time. In this counterfactual exercise, the firm uses a la carte price format for the online-streaming service. This price format is popular with companies that include Amazon, Apple. Since there are no intertemporal tradeoffs induced by the closed loop model present with online streaming, we assume the same consumption utility for online streaming as in equation (2):

$$ u_{ict} = \left( \sum_{k \in C_t} \alpha_i^k I[c_{it} = k] \right) + \alpha_i^{cw} c_{it} I[t \in T_w] + \alpha_i^{cs} \log(\omega_t) c_{it} + \alpha_i^p p c_{it} + \epsilon_{ict} $$

(11)

where $u_{ict}$ is the utility from watching $k \in \{1, 2, 3\}$ streamed movies. The first three terms represent the baseline consumption utility, additional utility from weekend consumption and a larger content set, respectively, and are defined similarly as in equation (2). The error term, $\epsilon_{ict}$ is the idiosyncratic shock that is assumed to follow a Type I Extreme Value distribution and is independent across consumers, consumption decisions and time periods. The rest of equation (11) differs from the subscription-based utility (as specified in equation (2)) in three ways. First, the payment disutility is incurred concurrently with the consumption occasion (vs. at the beginning of the payment cycle),
and is proportional to the number of moves consumed (vs. a fixed fee that is independent of the consumption). Second, with *a la carte* pricing, online streaming imposes no quota constraints on consumption. Third, there is also no switching costs because there is no need to change plans. The latter two points imply that the consumer’s decision problem is reduced to one that maximizes the daily-level contemporaneous utilities.

The counterfactual analysis is based on the same set of consumers in the estimation sample, over a two-year horizon, and is conducted as follows. For each price level, we simulate the consumption (i.e., purchase) choices of each consumer over a two-year horizon, using individual-level preference parameters previously obtained. We then aggregate across all consumers to compute the total revenue, and use the assumed marginal cost to obtain the total costs. The profit-maximizing price is then found using a grid search. To examine the sensitivities of optimal price and profit with respect to the marginal cost, we repeat the same exercise for three levels of marginal cost ($0, $1.0 and $2.0).\(^{15}\) Results are summarized in Figure 9. We find that when the marginal cost is $0, the optimal pay-per-view price is $3.05, resulting in a per-consumer profit of $49.1. For higher marginal costs of $1.0 and $2.0, we find that the optimal prices are higher ($3.44 and $4.62), but the per-consumer optimal profits are lower ($34.30 and $23.30 respectively). Notably, the per-consumer profits at all three levels of marginal costs are lower than that contributed by current customers of the firm ($127.1).

Overall, the results suggest that switching to digital delivery may *reduce* the firm’s profit even if it is technologically feasible. Admittedly, caution is warranted in interpreting this finding because of three reasons. First, we make the assumption that the consumers’ preference for online streaming is the same as that from watching physical DVDs. Arguably, there may be differences in the consumption experiences between these two outlets. While the video quality of streamed movie may be lower than that of the DVDs (especially Blue-rays), it may afford the consumer extra convenience in accessing movies (e.g., there is no need to open the mail, or to ship it back). If these conveniences are valued by consumers, then our results should be interpreted as a conservative estimate for the profit that can accrue from online streaming. In fact, the recent success of Netflix streaming is likely attributed to the ease of use of online streaming. Second, we use a two-year time horizon in this exercise. With *a la carte* pricing, consumers are not bound by any contract.

\(^{15}\)The marginal cost of online streaming includes the data transfer fee and the licensing fee for the movie. While the data transfer fee is small - approximately 6 cents for a standard, 2-hour movie and 9 cents for a high-definition movie, the licensing fee typically varies across movies and can be as high as $4 per movie. Source: [http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Stream-This!-Netflixs-Streaming-Costs-65503.aspx](http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Stream-This!-Netflixs-Streaming-Costs-65503.aspx). Retrieved on March 18, 2015.
Thus, they may use the service for a longer time, as there is no notion of leaving as a subscriber. Third, it is possible for the firm to utilize both the RBM model and online streaming. It would be interesting to explore whether these two distribution channels are likely to be complements or substitutes (Liu et al., 2010; Sriram et al., 2010). It should be pointed out that if additional data from digital streaming users becomes available, our modeling framework can incorporate the data to examine the implications of quality differences in the two delivery methods for the consumers, as well as the synergy between RBM and online streaming.

Figure 9: Digital Delivery Counterfactual

7 Discussions and Conclusions

The Rental-by-Mail business model can be used by many rental products for which (1) the firm can efficiently maintain a large pool of rental products to better match consumers’ idiosyncratic preferences and needs (Elberse, 2008); (2) the value of consumption declines significantly, such that the same consumer derives a relatively large value from renting the product for a short period of time, compared to owning the product; and (3) the firm can potentially build a cost-effective infrastructure to deliver products to its consumers via a delivery service. Rental products that
fit these criteria include both information goods, such as movies, books, video games, and non-information goods, such as art works and toys. Thus, an investigation of RBM consumers can be useful for a wide array of rental services.

This paper takes a first step in developing an integrated framework for understanding consumers’ decision-making with RBM services. Specifically, we model RBM consumers’ daily-level consumption and monthly-level purchase decisions, explicitly recognizing the two constraints of consumption: the instantaneous quota and nontrivial turnaround time. Consumers in the model are forward-looking and obtain utility from consuming the product at hand, but can also wait for better consumption occasions in the future. Their value of waiting also depends on the number of rental products in possession, as well as expectations of the future availability of rental products. In addition, consumers choose plans that allow them to access a different number of products. We link consumers’ consumption and purchase decisions in the same modeling framework, built from micro foundations. This framework enables us to rationalize consumers’ purchase decisions, and to obtain estimates of their willingness to pay for each rental plan; this enables us to conduct a number of counterfactual analyses. As the counterfactuals show, the firm can leverage such estimates to improve product designs and reduce cannibalization across plans.

A second contribution of this paper is that we provide a characterization of the closed loop rental process, which has nontrivial implications for the endogenous consumer demand for rental products. We demonstrate the key role played by the turnaround time, an important operational feature of the RBM model that has never been studied. An implication of the turnaround time is that, unlike the classical pay-as-you-go rental model, consumers pay different effective prices per consumption occasion under the RBM model, due to the heterogeneous turnaround time required to serve them and the fixed subscription price.

Consumers’ consumption rate, as well as the value derived from the RBM service, decreases with the turnaround time. The effect of turnaround time on consumption rates is further impacted due to forward-looking consumers, who strategically plan their consumption decisions to maximize total consumption utilities over time. Regarding the firm, whereas reducing the turnaround time requires costly investments for the firm (e.g., building and staffing more distribution centers, increasing the size of the product inventory to reduce stock-out instances), we show that the return on investment from reducing turnaround time can be precisely quantified, thus creating new opportunities for the firm to attain higher profitability by increasing operational efficiency. We caution, however, that while reducing turnaround time represents a gain in operational efficiency, such a gain does not
necessarily translate to higher profit. The fixed subscription price format implies that the firm does not get compensated for each rental; and a reduction in the turnaround time increases the cost of operations, especially for serving heavy users. For example, Netflix claims that it loses money on consumers whose effective price per rental is less than $2.0. This caution is particularly relevant for RBM services with high marginal costs unless the consumers’ valuation for consumption is sufficiently high. In sum, our findings suggest that it is important for the firm to account for both operational and pricing decisions, and our model can potentially help the firm to better evaluate its return on investment by lowering the turnaround time.

As a third contribution, we show that the firm can account for both consumer forward-looking and individual-heterogeneity. We adopt a Bayesian estimation framework, which provides a high degree of flexibility in modeling heterogeneous consumers, but which is computationally intensive in dynamic structural models. However, it is now feasible to implement, due to recent advances in estimation methods and computing power. We show that the heterogeneity in consumer preferences is too substantial to ignore, and such heterogeneity can serve as a meaningful basis for the firm to customize its marketing mix.

**Limitations and Future Research**

Our current research has a number of limitations, some of which open up important avenues for future research. First, our results are based on a specific rental service, and caution is advised in generalizing our findings to other settings. Particularly, while utilities from movie consumption are generally derived within a day, other rental products (e.g., video games and designer dresses) may give consumers a stream of utilities over a longer time interval. Future research can extend our modeling framework to those types of services.

Second, whereas our setting features a highly differentiated firm with unique products that faces little direct competition, we could extend the framework to a competitive market. Third, we use the inferred, rather than observed dates for when a consumer receives or sends out a movie. Specifically, we used the dates when the returned rental product was received by the firm and the firm’s estimated turnaround time. These inferences are based on the best information available to us and are consistent with the fact that there is no evidence that the firm has intentionally slowed down the service during the observation period. However, random delays by the USPS service may potentially lead to some discrepancies between the inferred and actual dates of shipment. Data of better quality (e.g., consumption diaries from a consumer panel) can be used to construct the state
space more precisely and avoid potential bias in the estimates for consumption utilities. We also
do not explicitly consider the heterogeneity in consumers’ preferences across rental products. It is
possible to extend our model by allowing for product differences. Finally, we focus on product rental
decisions, and future research can analyze both the rental and purchase decisions of consumers, if
both are offered by the firm, in order to generate further insights for the firm. In sum, there are
multiple ways to extend the first empirical modeling framework for RBM service proposed in this
paper, which we expect to see developed further in future research.
References


Appendix A: Estimation Details

Estimation of $\Omega$

We estimate separately from the data in a first stage. We first discretize the data into $N_\omega = 5$ bins, and estimate the $(N_\omega \times N_\omega)$ transition matrix, imposing the following restrictions. Since the content set only evolves as an increasing process, we set the non-diagonal elements of the lower triangular matrix to zero. We also allow increasing transitions only to the next higher state for simplicity, and because the data in the content set support this transition.

\[
\Omega = \begin{bmatrix}
\Omega_{11} & \Omega_{12} & 0 & 0 & 0 \\
0 & \Omega_{22} & \Omega_{23} & 0 & 0 \\
0 & 0 & \Omega_{33} & \Omega_{34} & 0 \\
0 & 0 & 0 & \Omega_{44} & \Omega_{45} \\
0 & 0 & 0 & 0 & \Omega_{55}
\end{bmatrix}
\] (12)

We estimate the parameters $\Omega_e = (\Omega_{11}, \Omega_{12}, \Omega_{22}, \Omega_{23}, \Omega_{33}, \Omega_{34}, \Omega_{44}, \Omega_{45})$ non-parametrically using a bin estimator. Note that $\Omega_{55} = 1$ is fixed, since there is no other state to which it can transition.

Estimation for the structural parameters is done using the IJC method of the Bayesian estimation of dynamic discrete choice models developed by Imai et al. (2009). The estimation was implemented in R. To expedite the estimation, the computation-intensive part was written in C++, which significantly reduces the computation time. The vector of heterogeneous parameters $\theta$ follows a continuous distribution. Furthermore, for each $\beta \in \theta$, we use $b$ to denote the mean of $\beta$, and we assume a normal prior for $b$, $N(b_0, s_0)$. 

49
Outline of the MCMC Algorithm

Below, we outline the steps of the MCMC algorithm that generates the posteriors of the parameters $\Theta$ in the utility function: $\Theta = \{\alpha_w, \alpha^{sw}, \alpha^1_i, \alpha^2_i, \alpha^{cs}_i, \alpha^p_i\}$, as well as the hyper-parameters that govern the distribution of the heterogeneous parameters: $\Xi = (\Delta, V)$.

Note that while all seven structural parameters in our proposed model are modeled as heterogeneous, the estimation steps outlined below provides an illustration for the estimation procedure of both homogeneous and heterogeneous parameters.

The MCMC process has three blocks: Block I (Step 1) draws the set of $NPAR_{homo}$ homogeneous parameters $\gamma = (\alpha^w, \alpha^{sw})$. Block II (Step 2) draws the set of $NPAR_{hetero}$ heterogeneous parameters, $\alpha_i = (\alpha^1_i, \alpha^2_i, \alpha^{cs}_i, \alpha^p_i), \forall i$. Block III (Steps 3 and 4) draws the hyper-parameters $\Xi$.

Step 0

At the beginning of each iteration $r$, start with the history:

$$
\mathcal{H}^r = \{(\gamma^*, \alpha^*_i)^l, EV \left(s, \gamma^{sl}_i, \alpha^{sl}_i\right)_{l=r-N}^{r-1}, q(\gamma^{r-1}_i, \alpha^{r-1}_i|DATA_i)_{l=1}^{I_i}\}
$$

where $I$ is the total number of consumers in the estimation sample, and $N$ is the number of past iterations that will be used for the approximation of the $Emax$ function in Steps 1 and 2. The IJC algorithm makes an important distinction between the history of candidate parameters and accepted parameters. We use plain subscript (e.g., $r - 1$) to denote the previously accepted parameters, and use superscript $*$ to denote the candidate parameters.

The first component of the history, $(\gamma^*, \alpha^*_i)^l = (\alpha^w*, \alpha^{sw*}, \alpha^1_i, \alpha^2_i, \alpha^{cs}_i, \alpha^p_i)$, is the vector of candidate parameters from the past iteration $l$. The second component, $EV \left(s, \gamma^{sl}_i, \alpha^{sl}_i\right)$ is the corresponding pseudo-value function. Conceptually, the pseudo-value function is an approximation to the full solution of the true value function: the sequence of pseudo-value function converges to the true value function in probability uniformly (Imai et al. 2009). The third and final component, $q(\gamma^{r-1}_i, \alpha^{r-1}_i|DATA_i)$, is the pseudo-likelihood function, which depends on the previously accepted parameters $(\gamma^{r-1}_i, \alpha^{r-1}_i)$. Apparently, $q(\gamma^{r-1}_i, \alpha^{r-1}_i|DATA_i)$ also depends on $DATA_i$, which is the sequence of the observed consumption and purchase decisions $\{d_i\}_{i=1}^{T_i}$ plus the payoff relevant state variables (e.g., sequence of DVDs, previous plan choice, weekend, day of the month, and...
DVD content size) at time $t$. To reduce clutter in the notation, we suppress the dependence of $q$ on $DATA_i$, and write $q(\gamma^{r-1}, \alpha^{r-1}_i)$ instead.

**Step 1**

In this step, we draw the homogeneous parameters $\gamma^r$ using the random-walk Metropolis-Hastings algorithm.

(1a) Based on the current draw of $\gamma^{r-1}$, use a multivariate normal density (i.e., $N(\gamma^{r-1}, \Sigma)$) as the proposal density. $\Sigma$ is an $NPAR_{homo}$ by $NPAR_{homo}$ diagonal matrix whose elements are proportional to the parameter estimates of homogeneous parameters obtained from the maximum likelihood estimation. The vector of candidate parameters is denoted $\gamma^r$.

(1b) For each consumer, compute the pseudo-likelihood at the candidate vector $\gamma^r$, i.e., $q(\gamma^r, \alpha^{r-1}_i)$, conditional on $DATA_i$.

$$q(\gamma^r, \alpha^{r-1}_i) = \prod_{t=1}^{T_i} \left[ \frac{\exp \left( V^r_{dit} \left( s_{it}, \gamma^r, \alpha^{r-1}_i \right) \right)}{\sum_d^{D_{it}} \exp \left( \bar{V}_d \left( s_{it}, \gamma^r, \alpha^{r-1}_i \right) \right)} \right]$$

(14)

where

$$V^r_{dit} \left( s, \gamma^r, \alpha^{r-1}_i \right) = u_{idt}(\gamma^r, \alpha^{r-1}_i) + E\hat{V} \left( s, \gamma^r, \alpha^{r-1}_i \right)$$

and $E\hat{V} \left( s, \gamma^r, \alpha^{r-1}_i \right)$ is the Emax function approximated by the weighted average of the $N$ past pseudo-value functions:

$$E\hat{V} \left( s, \gamma^r, \alpha^{r-1}_i \right) = \sum_{l=r-N}^{r-1} \frac{K_h \left( \gamma^r - \gamma^l, \alpha^{r-1}_i - \alpha^l \right)}{\sum_{k=r-N}^{r-1} K_h \left( \gamma^r - \gamma^k, \alpha^{r-1}_i - \alpha^k \right)}$$

(15)

Notice that the pseudo-value functions, which are the bases of the non-parametric approximation, correspond to the previous candidate parameters, which have larger variations compared with the accepted parameters. The weight for each of the $N$ pseudo-value functions are based on the distances between the candidate vector and the previously stored vectors $\gamma^l$. We use the standard Gaussian kernel for the non-parametric approximation:
\begin{align}
K_h(\gamma_r^* - \gamma_l^*, \alpha_i^{r-1} - \alpha_i^{l}) &= \frac{2\pi}{NPAR} \exp \left[ \sum_{j=1}^{NPAR_{homo}} \frac{(\gamma_r^* - \gamma_l^*)^2}{2h_j} + \sum_{j=1}^{NPAR_{hetero}} \frac{(\alpha_i^{r-1} - \alpha_i^{l})^2}{2h_j} \right] \\
\text{where } NPAR \text{ is the dimension of } \Theta, \text{ and } NPAR = NPAR_{homo} + NPAR_{hetero}. \text{ We use } h_j \text{ to denote the bandwidth or smoothing parameter for the } j\text{th parameter. The selection of the bandwidth is based on the tradeoff between the bias and variance of the resulting estimator. Using Silverman’s rule of thumb (Silverman 1986), we set } h_j = \hat{\sigma}_j N^{-\frac{1}{5}}, \text{ where } \hat{\sigma}_j \text{ is the sample standard deviation of } N \text{ sample points.}
\end{align}

(1c) Similarly, we compute the pseudo-likelihood at the previously-accepted vector \(\gamma_r^{-1}\):

\begin{align}
q(\gamma_r^{-1}, \alpha_i^{r-1}) &= \prod_{t=1}^{T_i} \frac{\exp \left( \bar{V}_{dt} (s, \gamma_r^{-1}, \alpha_i^{r-1}) \right)}{\sum_{d'} \exp \left( \bar{V}_{d'} (s, \gamma_r^{-1}, \alpha_i^{r-1}) \right)} \\
\text{Assuming a diffuse prior for } \gamma, \text{ we determine whether or not to accept } \gamma^r \text{ based on the following acceptance probability:}
\begin{align}
Prob_{\text{accept}} &= \min \left[ \frac{\prod_{i=1}^{T} q(\gamma_r^*, \alpha_i^{r})}{\prod_{i=1}^{T} q(\gamma_r^{-1}, \alpha_i^{r-1})}, 1 \right] \\
\text{If accept, set } \gamma^r = \gamma_r^*; \text{ if reject, set } \gamma^r = \gamma_r^{-1}.
\end{align}
**Step 2**

In this step, we use the random-walk Metropolis-Hastings algorithm to draw $\alpha_i^r$ for each consumer.

(2a) $\alpha_i$ is distributed as $N(\Delta^r Z_i, V_\alpha)$. We first generate a candidate $\alpha_i^{r*}$ as $\alpha_i^{r*} \sim N(\alpha_i^{r-1}, \Psi)$, where $\Psi$ is an $NPAR_{hetero}$ by $NPAR_{hetero}$ diagonal matrix whose elements are proportional to the parameter estimates of heterogeneous parameters from the maximum likelihood estimation on the pooled data (i.e., similar to Manchanda et al. 2004, we ignore the difference between the consumers in this step).

(2b) Compute the pseudo-likelihood for consumer $i$ at the candidate vector $\alpha_i^{r*}$, i.e., $q(\gamma^r, \alpha_i^{r*})$, conditional on $DATA_i$.

$$q(\gamma^r, \alpha_i^{r*}) = \prod_{t=1}^{T_i} \frac{\exp \left( \tilde{V}_{dit} (s_{it}, \gamma^r, \alpha_i^{r*}) \right)}{\sum_d \exp \left( \tilde{V}_{d} (s_{it}, \gamma_i^r, \alpha_i^{r*}) \right)}$$  \hspace{0.5cm} (19)

where $\tilde{V}_{dit} (s, \gamma^r, \alpha_i^{r*}) = u_{idt}(\gamma^r, \alpha_i^{r*}) + E \hat{V} (s, \gamma^r, \alpha_i^{r*})$

and $E \hat{V} (s, \gamma^r, \alpha_i^{r*})$ is the Emax function approximated by the weighted average of the $N$ past pseudo-value functions: $E \hat{V} (s, \gamma^r, \alpha_i^{r*}) = \sum_{l=r-N}^{r-1} E \hat{V} (s, \gamma^r, \alpha_i^{r*})$

$$E \hat{V} (s, \gamma^r, \alpha_i^{r*}) = \sum_{l=r-N}^{r-1} \frac{K_h (\gamma^r - \gamma^l, \alpha_i^{r*} - \alpha_i^{l*})}{\sum_{k=r-N}^{r-1} K_h (\gamma^r - \gamma^k, \alpha_i^{r*} - \alpha_i^{k*})}$$ \hspace{0.5cm} (20)

The weights for the $N$ pseudo-value functions are based on the distances between the candidate vector and the previously stored vectors $\alpha_i^l$. Similar to Step 1(b) above, we used the Gaussian kernel:

$$K_h (\gamma^r - \gamma^l, \alpha_i^{r*} - \alpha_i^{l*}) = \frac{2\pi}{NPAR_homo} \exp \left[ \sum_{j=1}^{NPAR_{homo}} \frac{(\gamma^r - \gamma^l)^2}{2h_j} + \sum_{j=1}^{NPAR_{hetero}} \frac{(\alpha_i^{r*} - \alpha_i^{l*})^2}{2h_j} \right]$$ \hspace{0.5cm} (21)
and again we use Silverman’s rule of thumb (Silverman 1986) to determine the optimal bandwidth \( h \).

(2c) Similarly, we compute the pseudo-likelihood at the previously-accepted vector \( \alpha_i^{r-1} \):

\[
q(\gamma^r, \alpha_i^{r-1}) = \prod_{t=1}^{T_i} \frac{\exp \left( \tilde{V}_{dt}(s, \gamma^r, \alpha_i^{r-1}) \right)}{\sum_{d' \in D_t} \exp \left( \tilde{V}_d(s, \gamma_i^r, \alpha_i^{r-1}) \right)}
\]  

(22)

Then we determine whether or not to accept \( \alpha_i^r \) based on the following acceptance probability:

\[
Prob_{\text{accept}} = \min \left[ \frac{q(\alpha_i^r \pi(\alpha_i^r))}{q(\alpha_i^{r-1}) \pi(\alpha_i^{r-1})}, 1 \right]
\]  

(23)

where \( \pi(\cdot) \) is the prior density: \( \pi(\alpha_i) \propto (\alpha_i - \Delta_i^{r-1} Z_i)^{\nu} V_{\alpha}^{r-1}(\alpha_i - \Delta_i^{r-1} Z_i)^{\nu'} \)

If accept, set \( \alpha_i^r = \alpha_i^{r*} \); if reject, set \( \alpha_i^r = \alpha_i^{r-1} \).

Note that Step 2 is iterated for all consumers \( i = 1, ..., I \), where \( I \) is the total number of consumers in the estimation sample.

**Step 3**

Based on \( \alpha_i^r = (\alpha_1^i, \alpha_2^i, \alpha_3^i, \alpha_i^{cs}, \alpha_i^p, \alpha_i^w, \alpha_i^{sw})^r \) : draw the posterior mean \( \Delta^r \) from the posterior density:

\[
\Delta^r \sim N \left( \tilde{\Delta}, \tilde{V}\alpha \right)
\]  

(24)

where \( \tilde{\Delta} = \tilde{V}_{\alpha}(A_{\alpha}^{-1} \bar{\delta} + \sum_{i=1}^{I} Z_i^t V_{\alpha}^{r-1} \alpha_i^r), \) and \( \tilde{V}_{\alpha} = (A_{\alpha}^{-1} + \sum_{i=1}^{I} Z_i^t V_{\alpha}^{r-1} Z_i^t) \). We set the priors to be uninformative \( \bar{\delta} = 0 \), and \( A_{\alpha} = diag(100) \).

**Step 4**

Draw \( V_{\alpha}^r \) from the inverse Wishart distribution:

\[
V_{\alpha}^r \sim IW \left( \nu + I_{|\alpha|}, \sum_{i=1}^{I} (\alpha_i^r - \Delta^r Z_i)(\alpha_i^r - \Delta^r Z_i)^{'} \right)
\]  

(25)
We set $I_{[\alpha]}$, the prior mean of $V_\alpha$, to $\text{diag}(0.1)$; and the prior degrees of freedom to $\nu=\text{NHPAR}+3$, where NHPAR is the number of heterogeneous parameters.

**Step 5**

Use the candidate parameters $\alpha^{*r}_r$ to compute the pseudo-value function $E\tilde{V}(s, \gamma^{*r}, \alpha^{*r}_r)$ at the current iteration $r$. This step uses the Emax approximation computed in Step 2.

**Step 6**

For each consumer $i$, compute the pseudo-likelihood at the current iteration $r$: $q(\gamma^r, \alpha^r_i)$.

**Step 7**

Use the pseudo-value function and the pseudo-likelihood function from Steps 5 and 6 to update the history $H^{r+1}$. Go to iteration $r + 1$. 

55
Appendix B: Estimation Results

Nonparametric First Stage Content Set Transition

We first estimate Ω using a bin estimator. Specifically, we find:

\[
Ω = \begin{bmatrix}
0.9974 & 0.0026 & 0 & 0 & 0 \\
0 & 0.9944 & 0.0056 & 0 & 0 \\
0 & 0 & 0.9958 & 0.0042 & 0 \\
0 & 0 & 0 & 0.9916 & 0.0084 \\
0 & 0 & 0 & 0 & 1.000
\end{bmatrix}
\]

Figure 10: Convergence of Pseudo-likelihood

Note that Ω_{55} = 1 is fixed, not estimated.
Consumer Individual Level Utility Parameters

Figure 11: Histogram of Individual Parameters from Dynamic Semiparametric Model

<table>
<thead>
<tr>
<th>Parameter for 1 DVD ($\alpha^1$)</th>
<th>Parameter for 2 DVDs ($\alpha^2$)</th>
<th>Parameter for 3 DVDs ($\alpha^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Histogram for 1 DVD" /></td>
<td><img src="image2" alt="Histogram for 2 DVDs" /></td>
<td><img src="image3" alt="Histogram for 3 DVDs" /></td>
</tr>
<tr>
<td>Weekend ($\alpha^w$)</td>
<td>Content Set ($\alpha^{cs}$)</td>
<td>Price ($\alpha^p$)</td>
</tr>
<tr>
<td><img src="image4" alt="Histogram for Weekend" /></td>
<td><img src="image5" alt="Histogram for Content Set" /></td>
<td><img src="image6" alt="Histogram for Price" /></td>
</tr>
</tbody>
</table>
### Table 8: Estimates of Hierarchical Parameter $V_β$

<table>
<thead>
<tr>
<th></th>
<th>$α^1$</th>
<th>$α^2$</th>
<th>$α^3$</th>
<th>$α^P$</th>
<th>$α^w$</th>
<th>$α^{cs}$</th>
<th>$α^{sw}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$α^1$</td>
<td>6.05 (0.28, 10.52)</td>
<td>11.97 (0.12, 20.98)</td>
<td>-49.94 (-92.07, 0.03)</td>
<td>1.17 (-0.005, 2.03)</td>
<td>25.92 (-0.007, 46.76)</td>
<td>-0.99 (-1.74, -0.01)</td>
<td>-0.18 (-0.47, 0.02)</td>
</tr>
<tr>
<td>$α^2$</td>
<td>11.97 (0.12, 20.98)</td>
<td>26.78 (0.48, 47.98)</td>
<td>-115.58 (-209.46, -0.03)</td>
<td>2.68 (0.003, 4.67)</td>
<td>57.72 (0.06, 105.96)</td>
<td>-2.07 (-3.60, -0.01)</td>
<td>-0.44 (-1.14, 0.007)</td>
</tr>
<tr>
<td>$α^3$</td>
<td>-49.94 (-92.07, 0.03)</td>
<td>-115.58 (-209.46, -0.03)</td>
<td>1052.26 (1.17, 1896.6)</td>
<td>-21.30 (-37.0, -0.01)</td>
<td>-467.49 (-876.07, -0.14)</td>
<td>8.51 (-0.004, 16.11)</td>
<td>3.28 (-0.007, 7.19)</td>
</tr>
<tr>
<td>$α^P$</td>
<td>1.17 (-0.005, 2.03)</td>
<td>2.68 (0.003, 4.67)</td>
<td>-21.30 (-37.0, -0.01)</td>
<td>0.74 (0.055, 1.31)</td>
<td>14.83 (0.01, 29.12)</td>
<td>-0.19 (-0.35, 0.002)</td>
<td>-0.07 (-0.16, 0.006)</td>
</tr>
<tr>
<td>$α^w$</td>
<td>25.92 (-0.007, 46.76)</td>
<td>57.72 (0.06, 105.96)</td>
<td>-467.49 (-876.07, -0.14)</td>
<td>14.83 (0.01, 29.12)</td>
<td>365.57 (0.46, 894.48)</td>
<td>-4.31 (-7.97, 0.005)</td>
<td>-1.41 (-3.42, 0.20)</td>
</tr>
<tr>
<td>$α^{cs}$</td>
<td>-0.99 (-1.74, -0.01)</td>
<td>-2.07 (-3.60, -0.01)</td>
<td>8.51 (-0.004, 16.11)</td>
<td>-0.19 (-0.35, 0.002)</td>
<td>-4.31 (-7.97, 0.005)</td>
<td>0.23 (0.05, 0.36)</td>
<td>0.02 (-0.015, 0.071)</td>
</tr>
<tr>
<td>$α^{sw}$</td>
<td>-0.18 (-0.47, 0.02)</td>
<td>-0.44 (-1.14, 0.007)</td>
<td>3.28 (-0.007, 7.19)</td>
<td>-0.07 (-0.16, 0.006)</td>
<td>-1.41 (-3.42, 0.20)</td>
<td>0.02 (-0.015, 0.071)</td>
<td>0.24 (0.08, 0.35)</td>
</tr>
</tbody>
</table>

**Note:** We use boldface to denote parameters whose 95% highest posterior density intervals do not contain 0.