Essays on the Digital Consumer: Models of Engagement, Upgrade, and Referral Behaviors

A dissertation presented

by

Clarence Lee

to

Sunil Gupta

Vineet Kumar

Elie Ofek

Thomas J. Steenburgh

in partial fulfillment of the requirements

for the degree of

Doctor of Business Administration

in the subject of

Marketing

Harvard University

Cambridge, Massachusetts

May, 2014

Dissertation Advisors: Clarence Lee

Sunil Gupta

Vineet Kumar

Essays on the Digital Consumer: Models of Engagement, Upgrade, and Referral Behaviors

Abstract

In this dissertation, I present three essays on two empirical models of consumers using Web services. Business-to-consumer (B2C) Web services, such as Facebook, Dropbox and Pandora, have become a major part of the economy. Due to the low cost of digital distribution, these firms can provide their services for free, with the goal to attract a large customer base, while earning revenue by relying on advertisements or charging a small subset of customers for premium features. My goal is to characterize the various stages that a consumer faces when using a Web service: from adopting the service, to using it for personal and social needs, and to paying for the service. In addition, I also model the customer referral process, where the customer becomes a marketing instrument to encourage other customers to adopt. In the first essay, I explore the relationship between how customers find out about a service and how active they are when using the service. I estimate a hidden Markov model (HMM) of consumer behavior, and I characterize how the firms' social media efforts may encourage customers to be more active. In the second essay, I examine the relationships among usage, payment (upgrades), and referrals. I estimate a single agent dynamic structural model to capture these consumer decisions. Lastly, I conclude with an essay that presents the computational challenges in estimating the HMM and the dynamic structural model in a Bayesian fashion, and I also discuss how I use various estimation techniques, parallelization, and Amazon Elastic Compute Cloud to address these issues.

Contents

Acknowledgments

I am profoundly thankful for the guidance, care and wisdom that my dissertation committee have shown me over these last few years. Sunil Gupta, Vineet Kumar, Elie Ofek, and Tom Steenburgh – without this group of amazing individuals, I would not have been able to accomplish this today.

First, I would like to thank my co-chairs Sunil and Vineet. Sunil inspired me with his work on customer lifetime value and encouraged me to come to Harvard to study digital marketing with him. I am always amazed at his quick intellect and his remarkable ability to cut straight to the heart of the analysis – to this day I am still hoping I could be more like him in this aspect. His timely introductions to industry contacts has made it possible for me to visit Silicon Valley to acquire data for this dissertation. I also want to thank Vineet for shepherding me through the gory details of building a dynamic structural model. His generosity with his time and willingness to always meet with me, everyday, so we can hash out the nitty gritty parts of my model, have been essential to my development as a doctoral student. It is true that there is so much care that goes behind the scenes in these models! His eye for detail and his amazing ability to simplify my arguments have shed light on an invaluable process of how to chisel and make one's research better – I am hoping to reap from what I have learned from him in the years to come.

I am immensely grateful for what Elie has done for me. His creative and positive energy never fail to encourage me, and I feel so thankful that he took me as a student in the first year of my doctoral program. Since then, he has guided me through the process of class selection, field exam, paper submission, job market preparation, and now my dissertation. Not only he has done so much for me, I am thankful for all that he has done for the entire HBS doctoral program. I greatly thank Tom for all of the empirical tools and Bayesian statistics he has taught me all these years. I miss the fact that I can just walk into his office and learn from the way he thinks about data. His humor, wit, and kindness has been a source of encouragement during the tough times of the doctoral years. Without the guidance of Elie and Tom, I would not have been able to turn the second chapter of my dissertation into what it is today.

I also want to thank Anita Elberse for working with me on my first project at Harvard, for giving me the amazing access to data that resulted in my first paper, and for being the awesome person that she is. Her ability to be hilarious never fails to impress me, and I admire the attention, support, and care that she shows, not only to her doctoral students, but also to her legions of MBA students over the years from teaching SMICI. Special thanks to Rajiv Lal, Doug Chung, Thales Teixeira, John Gourville for their feedback on my job market talk and helpful comments over the years.

I want to thank Glen Urban, John Hauser, Michael Braun and the Marketing department at MIT for inspiring me to go into this field. It was they who had first shown me how exciting doing business research is and how wonderful a career like this could be.

I want to also thank John Korn, Jennifer Mucciarone, and everyone at the doctoral office for the financial support that has made this dissertation possible. It has been always a pleasure to visit Wyss to see everyone's smiling faces. Thank you Jen for meeting with me weekly during the months prior to the job market to help me stick to a schedule. I also want to thank Kathy Randel for always being incredibly helpful to us doctoral students. I want to thank Heather Ramsey for her tremendous help in the job market application submission process. A big thank you to the staff at FRCC for providing the computing power when I needed it the most.

There are so many friends I want to thank for helping me these years, but I want to especially recognize my roommate Matt Lee, Sujin Jang, Anoop Menon, and Sen Chai for their help and camaraderie over the years. I want to thank Tom Wollmann and Haris Tabakovic for making the FRCC lab such a fun place to be, and for helping me refine my research ideas and structural model. A big thank you to Jeff Lee, Taylan Yalcin, Tuan Phan, Joseph Davin, Lingling Zhang, Pavel Kireyev, and the marketing doctoral students for helping me chisel my work to the way it is today. These guys make up an invaluable support group that is crucially needed in a doctoral student's journey, and they have positively affected not only my professional work, but also my personal life.

My family have always believed in me, even during times when I did not in myself. I am truly blessed to have parents like Huei and Kuo Lane. I want to thank my grandmother Der-Fang for being the constant source of inspiration in my life – the kind of love she has shown me since I was a child is something that has kept me going even during the toughest times.

Lastly but not least, I want to thank my lovely wife Denise for sharing this journey with me. She has been an outstanding source of strength, and is the only person I know who is willing to fly multiple trips from Los Angeles to Boston, when we were dating, to take care of me, during the final months before going on the job market. I am still amazed at how she had the stamina to proofread my essays in between shifts in the operating room. Her love and care have made it possible for me to finally see all this work come to fruition.

Chapter 1

Introduction

In this dissertation, I present three essays on models that characterize customer usage of Web services. Business-to-Consumer (B2C) Web services, such as Facebook, Dropbox and Pandora, have become a major part of the economy. Many of the largest firms in this industry exceed a market cap of billions. In Silicon Valley alone, several companies have raised millions of dollars in venture funding, only to be sold for price tags in the billions. The common feature of these services is their usage of the Web to directly interface with customers. Via the Cloud such services are accessed by customers using a variety of devices such as desktops, laptops, and mobile devices. Due to the low cost of distribution, these companies are able to experiment with business models that allow them to provide free service to consumers, while earning revenue from advertisements or from a small subset of customers that pay for premium features. This business model has become known as "freemium" (free + premium).

Because all customers do not pay for the service, firms are required to both incentivize customers to sign up and to continue to engage them by keeping them active. Companies often report their Daily Active Users (DAU) or Monthly Active Users (MAU) as an indicator of the success of their services to the press and potential investors. Once firms gain a cohort of active customers, they then consider how to obtain a profit. To do this many companies adopt the following three strategies: 1) relying on the sale of advertisements to firms who are interested in reaching their customer base, 2) relying on a small portion of consumers to pay for premium versions of their service, or 3) a mixture of both strategies. The goal of this dissertation is to understanding consumer behavior patterns in online services, including engagement, referrals, and purchase. Importantly, I account for the social effects of these Web services, where customers use existing social features to use these services in a collaborative fashion.

In the first essay, I study how digital media services can develop an active customer base, focusing on the following two basic questions. First, how does the way customers use the service post-adoption to meet their own needs (personal usage) and to interact with one another (social usage) vary across methods of customer acquisition? Furthermore, how do firm-to-customer and customer-to-customer communications promote usage? I study these questions by collecting two unique data sets from different Web services and by developing a multivariate hierarchical Poisson hidden Markov model. This model captures the joint dynamics of customer engagement (personal and social usage) at the individual customer level and fits the data significantly better than univariate models. We show that post-adoption behavior varies depending on the method of customer acquisition. In one empirical context, namely an annotation and note-taking web service, we find that customers who hear about the service through Search and Mass-Invite exhibit significantly higher usage behavior as compared to customers who joined through Word-of-Mouth (WOM), whereas in a different context, namely a cloud-based file storage web service, customers that joined the service through WOM referrals exhibit the highest usage behavior. Regardless of how a customer is acquired, and in both empirical settings, we find that customer-to-customer communication post-adoption is more effective than firm-to-customer communication in keeping customers engaged. These findings suggest that firms should pay close attention to how the mode of customer acquisition affects subsequent usage intensity for their particular offering and that they should actively encourage customers to share information with each other post-adoption. Our goal is to provide a methodology that allows marketers to simultaneously capture the personal and social usage behavior of their service in an integrative model, and to provide a way for managers to test the relationship between routes of customer adoption and subsequent customer dynamic usage behavior.

In the second essay I look more closely at one online service that relies on the freemium business model, and I construct a dynamic structural model that captures not only the engagement behavior of customers, but also the upgrade and referral behavior. The goal of the essay is to link engagement to customer payment, and then to calculate the value of customers in this particular context. The freemium model leads to several questions interesting to marketers, which we explore in our framework. How much value should the free product provide to consumers relative to the premium product, given the inherent cannibalization effect? What is the right referral bonus incentive to offer to customers? How does sharing influence customers' likelihood of upgrading to the premium product? My collaborators and I develop an empirical microfoundations-based framework to understand dynamics of consumer behavior of plan choice, usage, and referral in the freemium setting and apply it to a novel panel data set from a leading cloud-based storage service. Using Bayesian methodology, we estimate the structural model and perform counterfactual analysis. We find that the value of free consumers is approximately \$36 per year, and that the existence of the referral program contributes to at least 60% of this value – signifying the importance of the referral program. In addition, we conduct profit maximization simulations, and we observe an asymmetry in the magnitude of the change in upgrade rates as we increase and decrease prices. Lastly, we explore simulations to maximize the average consumer referral rate by changing the referral incentives. Contrary to the belief that more is better, we find the existence of an optimal incentive point for referrals. Thus, we are able to characterize both the individual value of consumers to the firm as well as the network value of customers, providing a mechanism to capture the impact of consumer-to-consumer interactions.

In the third essay, I discuss the computational challenges in estimating the previous two models. I discuss the approach that I have taken, the challenges I have faced, and the limitations that I have yet to solve. I end the essay with a discussion on the potential methods that could be used to solve these limitations, as well as some of the latest techniques involving Cloud computing that may hold promise to estimating these models. I conclude the dissertation by a short discussion of the typical challenges in modeling customers in this setting, and what may be of interest for future research pursuits.

Chapter 2

Where Do the Most Active Customers Originate From and How Can Firms Keep Them Engaged?

2.1 Introduction

Digital media start-ups obsess about customer engagement. Unlike firms that sell physical goods, many of these start-ups provide free services to consumers, generating profits either by selling ad space on their Web sites or by having users who want "more" to upgrade to premium accounts. Therefore, making money depends not on the number of signed up accounts, but primarily on customers actively using the service. Recognizing this fact, investors typically view a start-up's customer engagement level, often measured by the number of users who have been active on the service over some time frame or by the amount of total user activity on the site during a certain period, as being a critical factor in their investment decisions.[1] Additionally, many investors incorporate customer engagement metrics into the funding offer as necessary conditions for receiving the full investment amount. Consequently,

[1] Various metrics are used in industry. For example, Zynga reports the number of unique log-ins everyday as a measure of how active its customer base is (Techcrunch 2012).

digital media firms trumpet the number of *active* users under their fold whenever they raise funds. For example, the first investors of Facebook gave the company a 1.5 million active user requirement for receiving the entire series A investment (Kirkpatrick 2010). Hence, understanding customer engagement is one of the key issues for many firms in this modern digital economy.

Nevertheless, developing an active customer base is a challenging and multifaceted problem. To begin with, customers commonly use Web services to satisfy both personal and social needs. For instance, personal usage occurs when a customer logs onto Pandora to listen to her stored music or when a customer reads an article on CNN.com. Social usage occurs when a Facebook user sends a message to a friend or when a Dropbox user shares a folder or file with a colleague. And, while a service's main value proposition may be individualistic, customers could still use it in a social manner via features included on the service (for example, CNN.com allows sharing an article read online with one's Twitter followers or Facebook friends). Therefore, there exists a need for a general framework that allows capturing both personal and social usage in an integrated manner. This paper develops a new methodology that allows for the joint examination of both consumer behaviors over time.

Furthermore, firms have several tools at their disposal to affect consumers' use of their systems. Some of these tools work by fostering a relationship between the firm and its customers (e.g., Blogs, Twitter feeds, the firm's Facebook page). Other tools enable a dialogue between customers (e.g., comment boards, chat forums and single-click e-mail sharing). For firms, a key question is what tools are more effective at increasing customer engagement, that is, heavier usage of the service, and by how much? Previous work has not compared the effectiveness of firm-to-customer vs. customer-to-customer communications at engaging customers.

Finally, firms acquire customers through various means. Some companies rely on Word-of-Mouth (WOM) campaigns to increase friend-get-friend adoption, while others rely on

Search advertising to boost their customer count. Most treatments of customer relationship management (CRM) efforts deal with the acquisition phase of increasing the customer base, but in post-acquisition analysis do not distinguish based on how they joined. In other words, they assume that "a customer is a customer is a customer" and no distinction is made based on the mode of acquisition. However, if customers acquired from different routes exhibit differential persistent usage behavior post adoption, then they have the potential to generate different long-term value, implying that firms would be well-advised to consider which acquisition method will yield the most active customers for their service. While some scholars have studied how certain marketing and social (WOM) actions affect the likelihood of first-time adoption and repeated purchases (Chevalier and Mayzlin, 2006; Choi et al., 2011; Chan et al., 2011), there are few studies that shed light on how acquisition routes affect post-adoption service usage behavior. Our goal is to see if there are indeed post-adoption usage variations among customers acquired from various routes and offer a methodology for quantifying the size of these variations. For instance, do WOM-acquired customers always use the service more than Search-acquired customers, and if so by how much? Do WOM-acquired customers share more information with other customers than Search-acquired customers post adoption? These are all questions that have not been adequately addressed in previous works and which can bear on the profitability of firms' CRM strategies.

Using consumer usage data from two different digital services, we develop a new multivariate hierarchical Poisson hidden Markov model that captures the joint dynamics of customer engagement (personal and social usage) at the individual level. We show that significant differences exist in customer engagement levels among customers acquired through different adoption routes. Specifically, we segment the customers by three typical acquisition tools for Web start-ups: WOM (e.g., a friend referring a friend), Mass-Invite (e.g., a link by a popular blogger on a site like Techcrunch), and Search (e.g., an organic search result on Google). Given that previous research suggests that WOM is an effective customer acquisition tool

(Trusov et al., 2009; Villanueva et al., 2008), we might expect to find that WOM-acquired customers are the heaviest users. We show that this does not hold in all situations. In the context of a Web annotation service, Search-acquired customers on average make 82% more personal usage, and Mass-Invite users make 37% more personal usage than WOM-acquired users. Meanwhile, in the context of a cloud-based file storage service, we find that WOM-acquired customers end up using the service about 8% more for personal purposes than customers from other routes. Our results therefore suggest that it is important for firms to empirically test where their most active customer originate from.

We further find that customer-initiated communication is more effective than firm-initiated communication at elevating customers' usage behavior. For instance, receiving an inbound message related to the service from another customer significantly increases the probability of nudging a user from a passive to an active personal usage state; and this finding held irrespective of the type of service we examined. This suggests that, post-adoption, it is effective to provide incentives for customers to interact with each other and share content than to try and directly communicate with them. One of our contributions, therefore, is providing a methodology that allows marketers to simultaneously capture the personal and social usage behavior of their service in an integrative model and to provide a way for managers to test the relationship between routes of customer adoption and subsequent customer dynamic usage behavior. Moreover, our findings show that it is indeed important for companies to assess these factors as they attempt to build an active customer base for their service.

In the next two sections, we discuss how our work fits into several streams of literature, and then present theories relevant for predicting how customers' expected usage behavior might depend on the adoption route and the source of customer communication post-adoption (firm-initiated or customer-initiated). Section 4 describes the empirical context and the research methodology and Section 5 discusses the specification of the multivariate Poisson hidden Markov model we employ. Section 6 discusses our findings, and in Section 7 we conclude with a set of managerial implications emanating from our work.

2.2 Literature Review

Although the two proposed research questions we focus on have not been previously studied at our level of detail, some related literature does exist. The closest work is Villanueva et al. (2008). This study compares the aggregate log-in behavior of Word-of-Mouth acquired customers and traditional marketing acquired customers, finding that WOM customer acquisition routes add nearly twice as much long-term value to the firm. Our work differs from this in that we model customer usage behavior at the individual level while accounting for customer usage dynamics and heterogeneity. We also are the first to jointly model personal and social usage behavior. Furthermore, we examine how firm-to-customer and customer-to-customer interaction affect usage behavior, something that Villanueva et al. (2008) do not directly observe.

Also related, Choi et al. (2011) compares the purchase behavior of customers from traditional acquisition methods (offline WOM and magazine advertising) and "IS-enabled" acquisition methods (online WOM and online Search) for an online retailer. They find that the purchase behaviors of customers acquired through IS-enabled techniques are independent of geography, while traditionally-acquired customers' purchase behaviors vary significantly according to geography. Although the main reasons customers shop online in their study are lower-prices and convenience, these reasons primarily apply to the repeated purchase of bulky, perishable goods and are unlikely to characterize the usage of Web services. Moreover, our distinction of acquisition modes is different, in that we separate between WOM and Search and do not lump them together.

Past research has argued that WOM is an effective tool for acquiring valuable customers, especially when compared to what is known about customers acquired from traditional routes like television and radio advertisements. For example, Villanueva et al. (2008) finds in a Web-hosting services context that WOM-acquired customers have higher customer lifetime value.[2] Trusov et al. (2009) and Aral and Walker (2011) have found that WOM is more effective than traditional marketing at acquiring customers. On the other hand, Search-acquired customers have been shown to be not-as-desirable customers since they are price-sensitive (Bakos 1997). This research might lead us to believe that WOM-acquired customers will always exhibit higher usage patterns, especially social usage, and that Search-acquired customers will exhibit the least amount of usage. However, this post adoption usage issue has not been previously studied and, moreover, past research has simply grouped Search customers together with WOM customers without explicitly comparing the two.

Some work in the service literature has examined post-adoption customer behavior. However, this literature mostly studies the customer's decision to keep/drop the service, hence by-passing the need to model usage behavior. In addition, this literature mainly describes services under contractual settings, such as cell-phone services (Lemon et al. 2002; White et al. 2007; Hogan et al. 2003; Iyengar et al. 2007; Verhoef and Donkers 2005). This differs from our research context in two ways. The first is that the contractual setting makes customers commit to pay upon sign-up, and contracts are usually set for a fixed period of time and a pre-specified usage level (with a penalty for ending a contract early). Secondly, these works primarily focus on customer *churn* when contracts end and do not explicitly model week-to-week usage of the service. Within the Web services context, however, adoption is typically free, so signing up does not necessarily imply usage; in fact after sign up an account can be dormant indefinitely without any penalty or cost. Additionally, Web services can be "dropped" at any time, and therefore, are not adequately addressed by the analyses from the telecommunication sector.

[2]We point out that in Villanueva et al. (2008) the term "WOM" is broad and captures tools such as referral from friends/family, magazine and newspaper articles, and referrals from search engines.

Finally, our work is somewhat related to literature that examines how various forms of communication impact customer behavior, particularly in an online setting. There are two dimensions that may be useful to think about when comparing different types of communications. The first dimension is whether the communication is under the control of the customer or of the firm. Although traditionally firms prefer to use communication tools that they can craft and control, these messages may not be very effective since customers realize that firms are trying to persuade them (Friestad and Wright, 1994). The second dimension is whether the communication is direct and personalized, versus messages that are broadcast (i.e., sent from one entity to many people). Past works have compared the effectiveness of directed, personalized communication versus passive, broadcast communication and found that the broadcast communication can be more effective at influencing customers to be active (Aral and Walker, 2011). Following this logic, one might expect firm-to-customer broadcast communication to be more effective at reaching a broad set of customers and influencing them to be active than directed, personalized communication between customers. That said, few have actually simultaneously examined the effect of firm-to-customer and customer-to-customer communications at the individual customer level and the implications of these for usage behavior.

To sum up, the contribution of our work relative to the extant literature is three-fold: 1) We establish that customer usage dynamics can differ according to how customers were acquired, 2) we develop a framework to simultaneously account for the personal and social nature of various services and find that it is important to distinguish between the two, and 3) we show how one can incorporate various forms of communicating with customers, firm-initiated and customer-initiated, and that these can have differential effects on usage.

2.3 Acquisition Modes and Engagement Levels

Our first goal is to examine how post-adoption usage behavior may differ depending on the mode by which a customer is prompted to join a Web service. Specifically, customers may adopt via WOM, hearing about or being referred to the service by a friend, family member or colleague, or they may influenced to sign up by other means common in the online domain, such as Mass-Invites and Search. Previous research suggests that all of these tools are more persuasive than traditional marketing instruments (Brown and Reingen 1987; Herr et al. 1991) and are more effective at attracting customers with greater long-term value (Villanueva et al. 2008). Given our research focus and the data we have, we do not examine the effectiveness of these methods in getting consumers to sign up, for example, the conversion rate from being exposed to a Mass Invite and signing up. Rather, our focus is on how knowing which route led a consumer to sign up is related to their post adoption behavior. Notwithstanding, we provide the following discussion for how most digital media start-ups use each of these acquisition tools as it will help set the context for how mode of acquisition might be associated with subsequent usage patterns.

WOM is a common approach in practice to acquire customers, and it works in a few different ways. One way is to be invited by other customers through a directed referral program, with firms often providing incentives for customers to make these referrals. For instance, Zipcar gives existing customers free driving credit for referring other customers to join. Another way is when customers organically refer other customers to the service, and in the Web services setting, customers recommend to other potential customers that they sign up. We classify customers who state that they heard about the service from family members, friends, and colleagues as being acquired by WOM.

Digital media start-ups also acquire customers through Mass-Invites. Popular technology news sites, such as TechCrunch and Mashable, periodically cover new digital/Web companies and include invite-links for customers to join the featured services. We are interested in this adoption route because it is an effective way for start-ups to acquire a large number of customers in a short amount of time. The "Techcrunch bump" is a well known effect among Silicon Valley entrepreneurs in which thousands of customers often sign up after a Web service is featured on a news site like TechCrunch.[3] Companies can sometimes exert influence on the likelihood of being featured by dedicating employees or hiring PR agencies to contact these news sites. Moreover, there are many popular blogs in this industry that can garner the same level of exposure as the major news sites. Therefore, firms can also directly pay or incentivize those who contribute to these blogs to promote these products.

Finally, customers can be acquired through Search, which is one of the largest interactive marketing channels.[4] Recent work has assessed the effectiveness of Search as a customer adoption route (Ghose and Yang 2008; Rutz and Bucklin 2011; Yao and Mela 2011). There are two main ways in which firms can reach Search customers. The first method is through search engine optimization (SEO). The more effort spent at optimizing search results at the major search engines, the higher the likelihood that customers will find out about the service organically (i.e., as part of the organic search link results that come up). The second way, which is typically more expensive than SEO, is to advertise through sponsored search results (as part of the links that appear on the right or top of the page and are paid for by the company). We classify customers as acquired through Search if they specified that they heard about the service through any of the major search engines (Google, Yahoo, or Microsoft Bing).

Given these three distinct adoption routes, the question arises as to which acquired customers will use the service the most after sign up, and whether or not the usage patterns are the same regardless of the nature of the service? Since customers who come through

[3] Josh Kopelman, 2008. "After the Techcrunch Bump." Redeye VC. http://redeye.firstround.com/2008/01/after-the-techc.html.

[4] "US Interactive Marketing Forecast, 2013 to 2018," Forrester Research 2013.

WOM heard about the service from someone known and likely trusted, we might expect these customers to use the service more than any of the other customer segments. Customers tend to recommend products to others who exhibit strong relationships with them (Aral and Alstyne 2011). Since customers typically recommend products to individuals whom they know relatively well, the customers who adopt as a result of WOM may deem the service's functionality or benefits as more relevant to them. Moreover, if one signs up through WOM, presumably he or she has at least one more person they know who is a user of the service, thereby presenting more opportunities to share information or to benefit from any network effects. Therefore, we might believe that WOM-acquired customers will be more active users both to satisfy their own personal needs and to share content with other people.

However, there is room to consider the other adoption routes as being associated with high levels of subsequent usage. In particular, a customer who purchases a product or signs up to a service as a result of a Search they conducted is likely to have initiated the process because of an expressed desire they had for something that performs the functions or provides the benefits that the product/service in question offers. Hence, they may be more inherently motivated and more in need of using the product than customers who had no ex-ante expressed need for it. Furthermore, research in psychology has shown that people tend to embrace more passionately and exhibit more attachment to items that they discovered actively on their own as opposed to being told about them passively(Aronson and Mills, 1959; D. Kahneman and Thaler, 1991; Plott and Zeiler, 2005) – once again suggesting that Search-acquired customers will exhibit high propensity to use the service heavily post adoption. As for Mass-Invite-acquired customers, perceptions of the recommender, for example, a high-tech guru or influential expert, may prompt readers to not only consider sign-up but also have confidence that using the service will deliver the purported benefits. And this would be consistent with evidence from the innovation adoption literature with respect to key opinion leaders and influential customers (Rogers, 2003; Venkatesh et al., 2000; Van den Bulte and

Joshi, 2007). In addition, the write-up featuring the product or service might include ways to avoid pitfalls or achieve better results/experience, all of which could lead to more robust usage.

2.3.1 Post-Adoption Communications

Once customers sign up, what can the firm do to encourage greater personal and/or social usage? Should it conduct campaigns to communicate directly with its customers using tools like Twitter and blogs, actions we refer to as firm-initiated or firm-to-customer communications? Or would customers interacting and sharing information amongst each other, actions we refer to as customer-initiated or customer-to-customer communications, be more effective at driving future usage? This is interesting because companies can facilitate customer-to-customer communication by providing appropriate means that enable such behavior. Facebook is well known to create and update features that encourage customers to interact with each other and to spend more time on the site. This was the original purpose of its "Newsfeed" and "Messaging" features, and since their introduction, have both been redesigned many times in order to facilitate usage[5].

With firm-initiated communications, companies have a high degree of control over what is conveyed, how the information is presented, and when it is delivered to customers. As such, the firm may be able to more positively tell customers how they can benefit from the service or provide them with useful updates and tips; and attempt to do so at those moments when the information is relevant for customers (Ajzen and Fishbein, 1980; Davis et al., 1989; Davis, 1986). However, customers may realize that the goal of firm-to-customer communications is to influence their beliefs or behaviors in favor of the vendor, and trigger resistance to these persuasion efforts (Friestad and Wright 1994, 1995). Conversely,

[5] "More Facebook Changes, Aimed at Users on the Go." http://www.nytimes.com/2013/01/03/technology/personaltech/facebooks-latest-mobile-interface-expands-features.html.

customer-to-customer interactions as part of a service, even if originating from individuals one does not know personally, are perceived as uncontrolled by the firm and the result of a spontaneous desire to communicate and share relevant information. As such, recipients of customer-initiated communications may be more open to these interactions and engage with the service. Consequently, we might expect these latter type of communications to lead to greater engagement levels and drive robust usage. This is consistent with studies showing that customers' sharing behaviors on social networking sites increase their friends' usage (Trusov et al., 2009; Aral and Walker, 2011); and this may be particularly true with respect to social usage due to reciprocity or because the communication triggers a desire or need to respond.

2.4 Empirical Application

For our empirical application, we identified two companies that have well-defined personal and social usage features as part of the their offering. The first firm offers a note-taking and web annotation service, similar to Evernote. The second firm offers a cloud-based file and document storage service similar to Google Drive.[6]

The first company provides a Web annotation service that lets customers make virtual sticky notes, highlight text on any Web page, and organize annotated web pages in folders for easy retrieval in the future. Sign-up is free, and once a customer joins, she downloads a browser plug-in. For this web service we define highlighting and placing sticky notes as forms of personal usage. For instance, a customer can browse the CNN.com Web site to read an article. Once there, she can use the plug-in to create a sticky note next to an item of interest on the Web page, write comments on the sticky note, and highlight certain parts of the article. The Web service then stores all sticky notes and highlights in the user's account,

[6]We are precluded from explicitly identifying the names of these companies due to their desire to stay anonymous. For information on the similar web services, please see http://www.google.com/drive/ and http://www.evernote.com for product descriptions of Google Drive and Evernote.

and it retrieves and reinserts them at the appropriate locations anytime the user revisits the Web site in the future. It is important to note that the customer is not actually making a direct change to CNN.com or commenting on the site itself (so others visiting CNN.com do not see annotations); rather they are virtual and only appear on the user's downloaded web pages. These activities have individual value, for example, if a person is conducting research on a topic over several weeks she can save time and enjoy convenient access to her notes and comments when revisiting Web pages previously reviewed. That said, the service also has collaborative value. In particular, if a customer would like to share an annotated Web page, such as a CNN.com article, so that others can quickly focus on what the sender highlighted or access their comments within the context of the web page, this can be done through the service's sharing feature. The receiver can view the annotated Web page, add their own annotations, and send a response back if they so choose. We refer to the action of sharing an annotated Web page as social usage.

The data set we analyze for the Web annotation service consists of 986 customers who made at least one annotation over a 56 week period[7]. Table A.1 shows a summary of usage statistics. We define personal usage Y_{it}^{p} as the number of highlights and sticky notes customer i makes in week t. We define social usage Y_{it}^{s} as the number of messages customer i sends to other users in week t using the service's sharing feature. An adoption route is defined as the way in which a customer learns about the firm's service. At sign-up, customers are asked how they heard about the service. While this is not a perfect measure, it is a standard method for inferring the route of customer adoption (Choi et al. 2011). Table A.1 shows the distribution of users by the different adoption paths relevant for this Web service. Since we

[7]We have compared the adoption route breakdown of this sample with the break down of excluded inactive customers (who did not make any annotations). The breakdown of the 986 customers are 227 for Search (23%), 246 for WOM (25%), and 513 for Mass-Invite (52%), while the breakdown for the excluded 887 customers are 172 for Search (19%), 209 for WOM (24%) and 506 for Mass-Invite (57%). Therefore, we do not believe there are no major differences in proportion of adoption routes between the mentioned sample and the excluded sample.

want to compare the usage level among customers from the different adoption routes, we incorporate these variables as dummies in the model.

To measure the amount of customer-to-customer interaction, we code the variable Inbound Sharing as the number of messages that a customer receives from other customers in a given week. Furthermore, we also have data on firm-to-customer communication. The company keeps an active blog and a Twitter account to communicate with its existing customers. Therefore, we also incorporate the number of blog posts and number of Tweets per week into our model[8].

One type of demographic information we have for this Web service is a customer's occupation. The management team indicated their belief that the product is most useful for customers who conduct much research on the Web. Indeed, academics and PR professionals are two occupations that make up a significant portion of the customer base. We account for the heterogeneity of customers from these different professions. Lastly, we also look at actions customers take when they sign up, such as whether they gave feedback in the optional comment section on the sign-up page, or if they invited others to join the service. Table A.5 displays all the variables we include in our model as observed customer heterogeneity.

The second data set we analyze is from a leading cloud file storage service. Similar to a service like Google Drive, this service allows consumers to store files on cloud servers and to share folders with anyone who is also a member of the service. Users typically have an indicator on their desktop toolbar that alerts them, or allows them to see, if a new file has been placed in a folder that they share with others. The service delivers individual value in that it provides storage on the cloud, allowing customers to access their files anywhere, anytime and on any internet enabled device (provided the free software is installed). The

[8] We note that customers-to-customer communication is observed at the individual level, and the firm-to-customer communication is observed at the aggregate level. We do not observe who actually reads the blog posts and Tweets. This is a common problem when empirically modeling advertising variables with a broadcast nature.

service also yields a collaborative value, in that users that are "shared" on a particular folder have access to the most recent material that has been uploaded to that folder; thus all updates are immediately visible to all and confusion over which version is most recent is minimized.

The data set for the file storage service consists of a sample of 1200 customers who made at least one file synchronization over a 206 week period. We define social usage in this context (Y_{it}^s) as the number of files a customer places in all of his or her shared folders in week t and define personal usage (Y_{it}^p) as the number of files placed in all non-shared folders in week t, in the customer's cloud account.

In this data set, we had information on whether a customer signed up for the service as a result of a referral sent from another member of the service. We code those customers who adopted the service via the company's referral invite program as WOM-acquired customers, and all other customers in a "baseline-acquired" category. In this case there was no direct question customers responded to about how they heard about the service, thus we are in effect lumping together customers who came through Search, Mass-invites, PR events, and other categories into this baseline category. While this offers fewer segmentation levels, it will allow us to test if customers originating from WOM referrals have different post-adoption usage behavior than customers from other adoption routes. Customer-to-Customer interaction in this data set is coded as the number of inbound files that other users place in a customer's shared folders. We also obtained the Firm-to-Customer interaction from the company's Twitter account and blog postings. The demographic information that we have is whether a customer has invited other customers during the time frame of the data.

2.5 Model Development

2.5.1 Sources of Dynamics

In modeling a customer's propensity to use a Web service over a period of time, it is important to think about how often a customer will consider accessing it and what may be driving behavior dynamics. Most likely, a customer will periodically contemplate using the service – becoming engaged with it every once in a while and then reverting back to a passive state where she no longer needs or considers it much. Past research has further shown that customers may be prompted to use a product when factors from their environment remind them to do so (Laibson 2001; Wood and Neal 2007, 2009). Examples of these "contextual cues" include firm-to-customer communication such as advertising, social media campaigns, or sales force interactions. In our context, we observe the Blog and Twitter posts of both Web services, where these efforts are enacted with the goal of inducing customer engagement.

In addition to contextual cues, research has also demonstrated how customer sharing and user activities may induce other customers to consider utilizing a service more heavily (Trusov et al. 2010; Aral and Walker 2011). Therefore, another source of dynamics that is important to take into account is the customer-to-customer communication that occurs organically among users. In our setting, we account for the total level of inbound sharing behavior that a customer receives at each time period as a potential source of dynamics. We confirm the existence of dynamics using the Run Test as in Frank (1962) and Netzer et al. (2008). Since Run Tests are conducted for binary random variables, we transform the personal and social usage variables into a 1 if there is any usage activity at all, and a 0 if there is none. The QQNORM plots of the Run Test Normal Deviates are shown in Figure A.1. Since the deviates are clearly below the standard normal distribution line, this evidence points towards the existence of dynamics in both the personal usage and social usage process that cannot be ignored. For comparison on fit performance, we compared three specifications of the HMM against models that do not account for dynamics (latent class models). We find that the out-of-sample fit for all HMM's are better than models with no dynamics. For a discussion of this, please see Table A.3, and the description in Section 2.6.1 (Model Selection) in the Results section.

2.5.2 Hidden Markov Model

One natural way to capture the dynamic behavior just described is to model a customer's latent state as a Markov chain, where his or her state at any given week depends on the previous week's state. Hidden Markov models have been used in marketing to describe many phenomena, including: Web-path analysis (Montgomery et al. 2004), tracking visual attention (Liechty et al. 2003), customer relationship interactions (Netzer et al. 2008) and services under contractual settings (Ascarza and Hardie 2011). Aside from their ability to capture the dynamics of unobserved and observed behaviors, HMM's also have enough flexibility to incorporate covariates to influence the state transition probabilities as well as the moments of the state dependent distributions.

There are reasons to believe that the process governing a customer's personal usage and social usage are different. Psychology and IT literatures suggest that customers continue to exhibit strong personal usage if they find products to be relevant (Venkatesh et al. 2000) or if they experience a high level of satisfaction (Bolton and Lemon 1999). However, customers may decide to interact and share with other customers for entirely different reasons: altruism, self-enhancement (Wojnicki and Godes 2012), and social exchange (Homans 1958). Hence, it is important to model both underlying processes distinctly, as opposed to restricting the observed personal and social usage behavior to one hidden Markov process. We propose a novel way to model both processes jointly through a multivariate hidden Markov specification.

To better understand our approach we begin by describing the hidden Markov model in a univariate fashion, modeling the personal and social usage processes separately, one latent process at a time, without accounting for the correlation among these behaviors. The structure of the HMM allows estimating the personal or the social process by simply changing the dependent variable Y_{it} accordingly. For ease of exposition, from here on Y_{it} can refer to either personal or social usage, interchangeably. At the end of this section, we discuss how to jointly model personal (Y_{it}^p) and social usage (Y_{it}^s) together in a multivariate hidden Markov model.

We seek to model customers' latent states over time, as well as their observed usage behavior. Let $i \in \{1, ...I\}$ be individual customers who joined the service at week T and we observe their usage behavior (Y_{it}) at week $t \in \{T, ..., T^{max}\}$. Furthermore, we assume that each customer's usage behavior is driven by an underlying unobserved (latent) state S_{it} that evolves over weeks t for each individual i. We also assume a set of time-invariant customer individual characteristics W_i and firm/customer-initiated communications (Z_{it-1}) that can influence each individual's transition probabilities from period $t-1$ to period t.

2.5.3 Initial State Probabilities

An important feature of the HMM is the initial state probability distribution. This is the probability that a customer starts out in a particular state after signing-up. Let π_{is} be the probability that customer i begins in state s upon sign-up. We assume that $\sum_{k=1,2,...,NS} \pi_{ik} = 1$, and a total of NS number of states. Therefore, let

$$\pi_i = [\ \pi_{i1},\ \pi_{i2}, \ldots,\ \pi_{iNS}]. \tag{2.1}$$

Many works using HMM assume the individual initial probability distributions to be the stationary distribution (Netzer et al. 2008; Montoya et al. 2010). However, we have enough data to allow us to actually estimate the individual π_i's for all customers, and this removes an unnecessary assumption on each customer's initial state probabilities.

2.5.4 Latent States - Markov Transition Probabilities

We assume S_{it} follows the first-order Markov assumption:

$$P(S_{it}|S_{it-1}) = P(S_{it}|S_{it-1}, S_{it-2}, ..., S_{iT}) \tag{2.2}$$

The transition probabilities for all enumerated states at time t can be described by a NS x NS matrix $Q_{i,t-1\rightarrow t}$ in which each corresponding (j,k) entries denotes the quantity $q_{it,jk}$, where $j,k \in \{1,2,...,NS\}$. For instance:

$$Q_{i,t-1\rightarrow t} = \begin{bmatrix} q_{it,11} & q_{it,12} & \cdots & q_{it,1NS} \\ q_{it,21} & q_{it,22} & \cdots & q_{it,2NS} \\ \vdots & \vdots & \vdots & \vdots \\ q_{it,NS1} & q_{it,NS2} & \cdots & q_{it,NSNS} \end{bmatrix} \tag{2.3}$$

where, $\sum_{k=\{1,2,...,NS\}} q_{it,jk} = 1, \forall j \in \{1,2,...,NS\}$.

2.5.4.1 Covariates on State Transition Probabilities

We investigate what factors may influence the transition from one state to another. We include time-varying covariates and control for individual-level observed and unobserved heterogeneity. Time-varying covariates on the transition probabilities may include promotion or communication. We specifically compare the effects of customer-to-customer communication and firm-to-customer communication as described in Section 4.

One way to incorporate covariates in the state transition probabilities is to use an ordered logit to link the covariates of interest to each of the transition probabilities similar to Netzer et al. (2008).

$$
\begin{aligned}
q_{it,s1} &= P(Transition\ from\ s\ to\ state\ 1) \\
&= \tfrac{exp(\theta_{is1}-\rho_s'\cdot Z_{it-1})}{1+exp(\theta_{is1}-\rho_s'\cdot Z_{it-1})}, \\
q_{it,ss'} &= P(Transition\ from\ s\ to\ s') \\
&= \tfrac{exp(\theta_{iss'}-\rho_s'\cdot Z_{it-1})}{1+exp(\theta_{iss'}-\rho_s'\cdot Z_{it-1})} - \tfrac{exp(\theta_{iss'-1}-\rho_s'\cdot Z_{it-1})}{1+exp(\theta_{iss'-1}-\rho_s'\cdot Z_{it-1})}, \\
q_{it,sNS} &= P(Transition\ from\ s\ to\ state\ NS) \\
&= 1-\tfrac{exp(\theta_{isNS-1}-\rho_s'\cdot Z_{it-1})}{1+exp(\theta_{isNS-1}-\rho_s'\cdot Z_{it-1})},
\end{aligned} \tag{2.4}
$$

for $s \in \{1,\ldots,NS\}$ and $s' \in \{2,\ldots,NS-1\}$. Where ρ_s is the parameter vector that captures the effect of time varying covariates Z_{it-1} on individual transition dynamics from $t-1$ to t. $\theta_{iss'}$ is the threshold parameter for individual i.

Observed and unobserved heterogeneity is incorporated in the following way.

$$
\begin{aligned}
\Theta_i &= [\ \theta_{i11}\ \ \ldots\ \ \theta_{iNS,NS-1}\ ], \\
\Theta_i &= \delta' \cdot W_i + \epsilon_{i\theta},
\end{aligned} \tag{2.5}
$$

where W_i are observed individual-level characteristics, δ is the corresponding parameter to be estimated, and $\epsilon_{i\theta}$ are the unobserved individual characteristics. We assume $\epsilon_{i\theta} \sim N(0,\Sigma_\epsilon)$, and Σ_ϵ is distributed Inverse-Wishart with proper diffuse priors.

2.5.5 State Dependent Personal and Social Usage Behavior

Depending on the latent state s that customer i is in at week t, he or she will exhibit a different level of personal usage Y_{it}^p and social usage Y_{it}^s. Since Y_{it}^p and Y_{it}^s are unbounded counts, a natural way to model these quantities is through a Poisson distribution. We therefore model the state dependent personal usage Y_{it}^p and social usage Y_{it}^s as independent Poisson random variables with a different mean parameter for each possible state s. These dependent variables can be estimated jointly, and we discuss how this is specified in the last section of the model description.

2.5.5.1 Covariates on the State Dependent Means

There are conceivably many factors that can affect a user's personal and social usage for any particular state. A significant covariate for one particular state means that it impacts a customer's average usage in that state at a particular time t. We use the standard exponential function to link the covariates to the Poisson mean parameter $\lambda_{i|s}$.

$$E[\ Y_{it}\ |S_{it}\ =\ s] = \lambda_{i|s} = exp(\tilde{\beta}_{0s} + \beta_s \cdot X_i), \tag{2.6}$$

where $\widetilde{\beta}_{0s}$ is the state dependent intercept and β_s are the estimated effects for the vector of covariates X_i in state s.[9]

Next, we discuss how we account for the correlation between the observed personal and social usage behaviors using a bivariate Poisson distribution, and how we incorporate covariates into the personal and social usage mean parameters.

2.5.5.2 Joint Estimation of the Personal and Social Usage Models

Many services offer personal-use features that encourage usage in an individual manner and also offer social-use features that encourage usage in a collaborative manner. This is certainly true in our data sets as explained in the previous section. Therefore, we need to be able to account for both personal and social usage in customer behavior. We now present a model that links the route of acquisition to both types of usage, and that allows for the joint-estimation of both processes. In addition, a customer's personal and social usage processes could be correlated[10]. Therefore, we propose two model enhancements that simultaneously characterize both processes and allow accounting for the correlation in observed personal and

[9] We place a restriction on $\tilde{\beta}_{0s} = \beta_{01} + \sum_{k=2}^{s} exp(\beta_{0k})$ in a similar fashion to Netzer et al. (2008) so that $\tilde{\beta}_{01} \leq \tilde{\beta}_{02} \leq \cdots \leq \tilde{\beta}_{0NS}$. This prevents the label-switching problem and ensures identification of states.

[10] We verified that such correlation exists. The Pearson Product-Moment Correlation for personal and social usage is $\rho = 0.20$ (significantly different from zero at $p < 0.01$). To ensure that outlier observations and zeros are not driving this correlation, we conducted the same test on a) aggregated usage data, b) usage data without zeros, and c) usage data without outliers (greater than five standard deviations from the mean). We still find equal or greater significant correlation in all cases.

social usage. These two enhancements are: 1) multivariate Markov chains and 2) multivariate state dependent distributions.

To account for both the individual and collaborative aspects in the underlying latent states, we extend the Markov assumption to jointly account for the personal and social usage processes:

$$P(R_{it},\ S_{it}|R_{it-1}, S_{it-1}) = P(R_{it},\ S_{it} \mid R_{it-1}, R_{it-2}, ..., R_{i,T}, S_{it-1}, S_{it-2}, ..., S_{iT}), \qquad (2.7)$$

where R_{it} and S_{it} are the corresponding latent personal and social usage states[11]. In the results section, we show that this joint model with $NS = 2$ outperforms several competing models, both for in-sample and out-of-sample fit tests. We therefore assume two states for both personal and social usage, but stress that the model can be generalized to any number of states. Note that this totals to 16 transition probabilities. For models with $NS \geq 3$, more stringent requirements on the data are necessary to identify all of the possible transition probabilities, for both personal and social usage. For instance, a multivariate Markov chain model with $NS = 3$ for both personal and social state transitions would consist of nine possible state pairs. Therefore, there are a total of 81 transition probabilities. When incorporating covariates on each of these transition probabilities, it becomes very easy to overparameterize the model, and therefore should be approached with care. We discuss the fit tests we conducted to check for model parsimony in our model selection section.

Let $P(R_{i1} = r,\ S_{i1} = s)$ be the joint probability that customer i begins in a personal usage state r and social usage state s at week 1, where $r, s \in \{1, 2\}$. Let the joint initial state probabilities be

$$P(R_{i1} = r, S_{i1} = s) = [\pi_{i11}\ \pi_{i12} \quad \pi_{i21}\ \pi_{i22}] \quad .$$

To jointly estimate the transition probabilities, we combine the 2 x 2 transition matrix of

[11] While another way to incorporate both processes would be to define one latent Markov chain with double the number of states, doing so would be forcing both processes to be correlated. We have conducted fit tests with HMM's with a univariate latent variable in Tables A.2 and A.3, and we show that our approach performs dramatically better.

the univariate models into a 4 x 4 transition matrix. We define $P(R_{it} = r', S_{it} = s' \mid R_{it-1} = r, S_{it-1} = s) = q_{itrsr's'}$, then we can express the joint transition probability matrix Q_{it} as

$$Q_{i,t-1\to t} = \begin{bmatrix} q_{it1111} & q_{it1112} & q_{it1121} & q_{it1122} \\ q_{it1211} & q_{it1212} & q_{it1221} & q_{it1222} \\ q_{it2111} & q_{it2112} & q_{it2121} & q_{it2122} \\ q_{it2211} & q_{it2212} & q_{it2221} & q_{it2222} \end{bmatrix}. \tag{2.8}$$

If we assume independence between R_{it} and S_{it}, we can incorporate heterogeneity and covariates into the transition probabilities by using two binomial distributions with logit links to covariates in the following fashion:

$$\begin{aligned} q_{itrsr's'} &= P(R_{it} = r', S_{it} = s' \mid R_{it-1} = r, S_{it-1} = s) \\ &= \left(1 - \frac{exp(\theta^p_{ir1} - \rho_r \cdot Z_{it-1})}{1+exp(\theta^p_{ir1} - \rho_r \cdot Z_{it-1})}\right)^{r'-1} \left(\frac{exp(\theta^p_{ir1} - \rho_r \cdot Z_{it-1})}{1+exp(\theta^p_{ir1} - \rho_r \cdot Z_{it-1})}\right)^{(2-r')} \\ & \quad \left(1 - \frac{exp(\theta^s_{is1} - \omega_s \cdot Z_{it-1})}{1+exp(\theta^s_{is1} - \omega_s \cdot Z_{it-1})}\right)^{s'-1} \left(\frac{exp(\theta^s_{is1} - \omega_s \cdot Z_{it-1})}{1+exp(\theta^s_{is1} - \omega_s \cdot Z_{it-1})}\right)^{(2-s')}. \end{aligned} \tag{2.9}$$

For instance, $P(R_{it} = 1, S_{it} = 2 \mid R_{it-1} = 1, S_{it-1} = 1)$, the probability that a customer moves from state $\{Personal, Social\} = \{1, 1\}$ to state $\{Personal, Social\} = \{1, 2\}$, can be expressed as:

$$q_{it1112} = \left(\frac{exp(\theta^p_{i11} - \rho_1 \cdot Z_{it-1})}{1 + exp(\theta^p_{i11} - \rho_1 \cdot Z_{it-1})}\right) \cdot \left(1 - \frac{exp(\theta^s_{i11} - \omega_1 \cdot Z_{it-1})}{1 + exp(\theta^s_{i11} - \omega_1 \cdot Z_{it-1})}\right).$$

There are many possible candidates for the multivariate state dependent distribution. One natural way to model the two unbounded counts while accounting for correlation is to use a multivariate Poisson distribution. For instance, we can model the state dependent distribution using a bivariate Poisson distribution:

$$P(Y^p_{it}, Y^s_{it} | R_{it} = r, S_{it} = s, \lambda^p_{ir}, \lambda^s_{is}, \lambda^c_{rs}) =$$

$$e^{-(\lambda^p_{ir} + \lambda^s_{is} + \lambda^c_{irs})} \frac{(\lambda^p_{ir})^{Y^p_{it}}}{Y^p_{it}!} \frac{(\lambda^s_{is})^{Y^s_{it}}}{Y^s_{it}!} \sum_{i=0}^{min(Y^p_{it}, Y^s_{it})} \begin{pmatrix} Y^p_{it} \\ i \end{pmatrix} \begin{pmatrix} Y^s_{it} \\ i \end{pmatrix} i! \left(\frac{\lambda^c_{irs}}{\lambda^p_{ir} \lambda^s_{is}}\right)^i,$$

where λ^p_{ir} and λ^s_{is} are the means of the personal and social usage for customer i, in states r and s. λ^c_{irs} is the parameter that captures the correlation between personal and social usage in states r and s. We can then further incorporate covariates into the first two mean parameters in the following fashion:

$$\lambda^p_{ir} = exp(\beta^p_r \cdot X_i),\ \ \lambda^s_{is} = exp(\beta^s_s \cdot X_i). \tag{2.10}$$

The estimates of s β^p_r, ρ_s, and $\theta^p_{iss'}$ are shown in the middle columns of Tables A.4 and A.9, while β^s_r, ω_s, and $\theta^s_{iss'}$ are shown on the right-most columns of the respective tables.

2.5.6 Likelihood Specification

Under the standard HMM specification[12], an individual's usage probabilities are correlated within each path through the unobserved states s_t, and therefore the joint likelihood must be summed over all the possible paths that an individual could take over the entire time periods:

$$\begin{aligned} L_i = \ & P(Y^p_{iT} = y^p_{iT}, ..., Y^p_{iT^{max}} = y^p_{iT^{max}}, Y^s_{iT} = y^s_{iT}, ..., Y^s_{iT^{max}} = y^s_{iT^{max}}) \\ = & \sum_{r_T=\{1,2\}} ... \sum_{r_{T^{max}}=\{1,2\}} \cdot \sum_{s_T=\{1,2\}} ... \sum_{s_{T^{max}}=\{1,2\}} [P(R_{iT} = r_T, S_{iT} = s_T) \\ \times & \prod_{t=T}^{T^{max}} P(R_{it} = r_t, S_{it} = s_t | R_{i,t-1} = r_{t-1}, S_{i,t-1} = s_{t-1}) \\ \times & \prod_{t=T}^{T^{max}} P(Y^p_{it} = y^p_{it}, Y^s_{it} = y^s_{it} | R_{it} = r_{it}, S_{it} = s_{it}) \Bigg] . \end{aligned}$$

[12]For full specification of the priors and the conditional distributions of the parameters, please refer to the appendix of Netzer et al. (2008).

2.6 Results

We estimated the personal and social usage hidden Markov model jointly in a MCMC hierarchical Bayesian fashion using a Gibbs sampler with a random walk Metropolis algorithm implemented in R/C++. We placed non-informative priors on all parameters, and we ran the MCMC with random initial values, for 80,000 iterations until convergence. Convergence is assessed using the Gelman-Rubin statistic (Gelman and Rubin 1992). We kept the last 20,000 iterations for inference. We first discuss the results from analyzing the Web annotation service data, and subsequently compare them to the results from analyzing the cloud-based storage service data. Full results for both services are provided. Section A.1 Tables A.4 - A.7 include the estimation results of the Web annotation service and Section A.2 Tables A.9 - A.12 include the estimation results of the Web file storage service.

2.6.1 Model Selection

Table A.2 shows the fit comparison between the joint-multivariate model that we present, univariate HMM models, and latent class models for the Web annotation service. We use several measures of fit, including the log marginal density, DIC, and Monte Carlo simulation versions of AIC and BIC (Raftery et al., 2007). We observe that for all measures, our two-state multivariate HMM model greatly outperforms all other models, even when the alternative models allowed for three states in each type of usage.

We note that we were unable to analyze a multivariate Poisson HMM model with three states for each usage type. As explained, such a model would entail 9 possible state pairs and 81 transition probabilities; we were unable to get such a model to converge. In examining Table A.2 it does seem that 3-state univariate and 3-state latent class models provide some improvement over their two-state counterparts in terms of fit. However, the improvement in the fit measures when moving from a univariate to a bivariate Poisson HMM is much more dramatic. As an additional check, we have also performed in-sample and out-of-sample fit tests using the same measures. For the out-of-sample test, we save the last eight weeks of each customer as the hold-out. Then, we calibrate the model using all periods prior to the hold-out sample, excluding the last eight weeks. The results of the hold-out fit measures and the calibration log marginal density, for all five models, are provided in Table A.3. Again, the joint-multivariate model beats all other models in in-sample and out-of-sample fit. In addition, the out-of-sample tests show that all HMM models perform better than models that do not account for dynamics. Consequently, we settle on interpreting results from the two-state multivariate HMM.

2.6.2 Interpreting the State Dependent Usage

The coefficients in the state dependent usage model should be interpreted as a Poisson regression, i.e., taking the exponential of the sum of the relevant estimates gives the expected personal and social usage in the respective states. The interpretation of the states is based on the average amount of usage in a given state. The expected personal and social usage in state 2 are at least an order of magnitude higher than those in state 1. Since customers in state 1 seem to be in an "unengaged" state and make little use of the Web service, we accordingly label state 1 as "passive". Furthermore, we label state 2 as "active" since customers in this state are engaged in the sense that they make substantial use of the Web service.

2.6.3 Results from the Web Annotation Service

2.6.3.1 The Relationship between Adoption Routes and State Dependent Usage

Table A.4 presents the posterior means and the 95% highest posterior density (HPD) intervals for the joint Bayesian estimates of the multivariate HMM for the Web annotation service data. We first focus on the results of the state dependent usage Poisson coefficients.

To begin with, Table A.4 shows that active customers use the system much more intensely than passive customers do. Giving these estimates economic meaning by using equation 2.10, we find that WOM-acquired customers make only 0.11 annotations per week on average when in the passive state, yet they make 18.96 annotations per week when in the active state. Similarly, Search and Mass-Invite acquired customers make only 0.20 and 0.09 annotations when in the passive state, respectively, and they make 40.13 and 26.05 annotations when in the active state per week.

Comparing the customers acquired through the different routes, we find that Search customers are more intense in their personal usage than WOM customers. Table A.4 shows that this difference is quite large. In the active state, Search-acquired customers conduct 112% more weekly personal usage than do WOM-acquired customers ($(\frac{40.13}{18.96} - 1) * 100\% = 112\%$). In the data, we find that an average customer spends 7.5 minutes for each annotation per daily session. Thus, a Search-acquired customer in the active state may on average spend an additional 11 hours per month using the service than a WOM-acquired customer ($\frac{7.5\ min\ *\ (40.13-18.96)\ *\ 4\ weeks}{60min} \sim 11\ hours$). Although neither type of customer uses the system much in the passive state, Search customers conduct 82% more weekly personal usage than WOM customers in this state ($(\frac{0.20}{0.11} - 1) * 100\% = 82\%$).

Furthermore, we find that Mass-Invite-acquired customers are more intense in their personal usage than WOM-acquired customers, but only when they are in the active state. Comparing the active state usage for Mass-Invite customers of 26.05 annotations with a WOM customer's 18.96 annotations per week, the former use the service 37% more. A 37% increase in annotations would translate to a customer spending an additional 4 more hours using the service per month ($\frac{7.5\ min\ *\ (26.05-18.96)\ *\ 4\ weeks}{60min} \sim 4\ hours$). For the firm, this would translate to significantly more customer engagement. Neither Mass-Invite nor WOM acquired customers use the service much in the passive state, respectively making 0.09 and 0.11 annotations per week on average.

Since various literatures tout the power of WOM on the adoption of products in general (Muller et al. 2010; Chevalier and Mayzlin 2006), one would naturally assume this relationship applies to usage behavior post-adoption as well, something that many industry observers do assume[13]. However, our findings reveal that this is not always the case insofar as personal usage is concerned in the Web annotation service context. In particular, those customers acquired through *non* WOM-based methods tended to exhibit more robust personal usage in the Web annotation service.

Unlike the personal usage results, we found no difference in the customers' social usage behavior depending on how they were acquired. All Search and Mass-Invite coefficients are insignificant in the right-most column of Table A.4, meaning that there is no difference among WOM, Search, and Mass-Invite acquired customers in social usage intensity. In other words, acquisition mode was not associated with different persistent social behavior post-adoption. This might be surprising too, as we might have expected customers who were acquired through WOM, presumably at the recommendation of friends, colleagues or family members, to be more active social users than customers acquired through other means or to have a greater network to leverage and share with post-adoption.

2.6.3.2 Dynamics in State Transitions: Impact of Customer-to-Customer and Firm-to-Customer Communications

From week to week, a customer may move from a passive to an active state, from an active to passive state, or stay in the state they were in (passive or active). Customers in an active state use the service much more than in the passive state, regardless of adoption route. There are two types of interactions that may influence a customer to transition from one state to another. The first type is customer-to-customer interactions. For instance, receiving an inbound message from a friend may be attention grabbing and an effective reminder of

[13]Based on our discussions with management at relevant companies.

the service, resulting in the customer moving from a passive to an active state. Furthermore, receiving the inbound message from a friend may trigger a desire to reciprocate, serving as an effective call-to-action that drives the customer to reply in social usage.

The second way to influence a customer's transition probability is through firm-to-customer interactions, where the firm uses digital media to communicate directly with customers. The company blog is the main instrument through which the company we studied engages in a dialogue with its customers. The company also uses Twitter for updates as well as to respond to suggestions and complaints. One of the primary jobs of the company's CMO is to write entries on the company blog and respond to inquiries on the company's Twitter profile. Accordingly, we incorporate the variables Inbound Sharing, Tweets, and Blog Post into the Markov transition matrix.

An important feature of the HMM is its ability to investigate what factors can significantly move a customer from one state to another. We find that Inbound Sharing among customers of the Web annotation service significantly transitions customers from a passive state to an active state, for both personal and social usage, per Table A.4. On average, an inbound share increases the personal usage passive-to-active transition probability from 0.4% to 6%, an increase of 15-fold. Similarly, an inbound share dramatically increases the social usage passive-to-active transition probability, from 0.3% to 80%. These percentages are obtained using the mean of the parameter distributions of the transition probability covariates from Table A.4 with equation 2.9. The first of these results is quite intriguing because it shows that allowing customers to share using the social feature (which of course results in others receiving an inbound share) further increases overall *personal usage*. This result confirms and adds to related findings in the literature where social influence positively affects the choice and usage of new media technologies (Nam et al. 2010). Notably, the social influence in question occurs post-adoption.

Tables A.6 and A.7 illustrate these changes to the joint personal and social transition probabilities from Equation 2.9. Notice that customer-to-customer communication significantly changes the probabilities in Table A.6, while the probabilities in Table A.7 do not change significantly. Receiving an inbound share increases the probability of a customer transitioning from a passive social state to an active social state from 0.74% to 61.03%. This effect also spills over to personal usage: it changes the transition probability of a customer who is in the passive-personal-passive-social state into the active-personal-active-social state from 0.04% to 10.97%. Moreover, a customer who receives an inbound share is more likely to stay in an active social state – an improvement from 12.22% to 75.90%. Again, we see this effect spilling over to impact personal usage: a customer receiving an inbound share is more likely to stay in an active personal usage state, an increase of 3.8% to 17.74% in probability.

Figure A.2 shows the dynamic effects of customer-to-customer communication on personal and social usage. We observe that when a customer receives an inbound message from another customer, it will take almost 3 additional weeks for their personal and social usage levels to return to the normal baseline. In addition, although the effect is significant for both usages, by comparing the normalized expected personal usage levels with the social usage level, we see that the effect of an inbound message is almost five times more effective in influencing social usage than personal usage.

While we included data on the media efforts of the company involved in our study, we did not find any significant relationship between the number of blog posts and/or Twitter entries per week and the transition probabilities from a passive state to an active state. As a result, companies similar to the one studied here may be better off focusing their marketing efforts on promoting features and providing incentives for customers to share among themselves. Putting the finding above in the context of the WOM literature, we gain a more nuanced understanding of how social interaction may impact behavior. Even though customers who adopted through WOM may not exhibit higher usage relative to customers from other adoption routes, WOM in the form of post-adoption communication linked to the service among customers is effective at increasing usage behavior; it works by bumping customers up to active states where these customers are on average more engaged.

These findings seem consistent with some observations in practice. Vish Makhijani, Zynga's Senior Vice President of Business Operations, noted that an important insight from their focus groups is that customers desire more ways to interact with their friends, in addition to the value they derive from just playing games. As a result, Zynga committed to developing more social features for users to interact with each other[14].

2.6.3.3 Observed Heterogeneity and State Transitions

We account for observed heterogeneity in the annotation Web service setting by incorporating individual-level differences in profession, feedback when joining, and the invitation of others. We now examine the hierarchical parameters in Table A.5.

We incorporated the individual-level profession information because we wanted to test if there are certain types of customers that the firm should focus on serving. The management team stated that they designed many features for academic researchers and public relations (PR) professionals. Hence, we wanted to see if these types of customers have a higher propensity to move up to an active state or stay in the active state once there. We find two notable dynamic usage differences among the various customer types. Specifically, we find that PR professionals are more likely to transition out of an active personal usage state. Furthermore, we find that academic researchers are more likely to move to an active personal usage state, but less likely to move to an active social usage state. Thus, academic researchers tend to use the personal features on the service, rather than using it to share with others, and they are "sticky" once they transition into an active usage state.

The coefficient for the covariate Invited Others is significantly positive for the active-to-active state transition for both personal and social usage. This means that customers who

[14]Piskorski, M., D. Chen, 2011. "Zynga." Harvard Business School Case.

have sent invitations to other people are, on average, more likely to stay in an active personal or social state than an average customer. This makes sense because the number of invites a customer sends to his or her network is probably a good indicator of how much he or she likes the service and/or how relevant they may find it. Hence, if they invite others, they are more likely to stay in an active personal or social state. Moreover, we also find that these customers are more likely to transition to an active personal usage state. Our finding would seem consistent with managerial literature that calls on firms to measure to what extent customers recommended their service as an indicator of their value to the firm(Reichheld, 2003).

Lastly, we observe that customers who give feedback when joining have a higher probability of staying in an active personal usage state. Translating this into probabilities, these customers have a 70% greater propensity to stay in an active state than a customer who did not give feedback upon adoption. Furthermore, they are also more likely to transition to an active social usage state. The fact that customers took the effort to make comments when signing-up could signal that they have a higher interest in the service and are more committed to it. Hence, once they are piqued to transition into an active state, they are more likely to stay there and use the service more.

2.6.4 Results from the Cloud-Based File Storage Service

2.6.4.1 The Relationship between Adoption Routes and State Dependent Usage

Next, we discuss the results from the second data set of a cloud-based file storage service. We start with the relationship between adoption routes and state dependent usage. The top of Table A.9 shows the state-dependent covariate estimates for personal and social usage for customers who adopt from WOM referrals versus all other customers (denoted Others).[15]

[15] In the cloud-based file storage service, given the incentives offered by the firm (extra free storage to both the referring and referred parties) management indicated to us that this was the predominant mode of

Using the mean estimates from Tables A.9 with Equation 2.10, we see that WOM customers sync a weekly average of 212 and 173 files into their personal (non-shared) and social folders when they are in an active state. While in a passive week, WOM customers sync an average of 1.04 and 4.06 files into their personal and social folders. For customers from all other adoption routes, we see an average of 196 and 165 in the active state, and an average of 1.45 and 3.35 in the passive state. When comparing across states, it is noted that customers' usage level is orders of magnitude greater in the active state than the passive state.

When comparing across adoption routes, the coefficient for WOM is significant. Thus, for this Web service, we find that customers who adopt through WOM referrals do exhibit higher personal and social usage behavior relative to customers who adopted the service through other routes or sources. Specifically, consumers who adopt as a result of WOM tend to add 8% more files into their non-shared, i.e., personal, folders than customers from other acquisition categories. Furthermore, WOM-acquired customers add 4.7% more files in shared folders than customers that adopted through other routes.

These finding are important for three reasons. First, they re-confirm our premise that acquisition mode can be associated with varying levels of customer usage behavior post-adoption. Second, and combined with the results from the previous data set, which acquisition modes are associated with greater customer usage behaviors can depend on the nature of the service. One way to think about our findings across the two data sets is that when a service has considerable collaborative value in addition to personal value, such as when a group of people want a common repository to store file on a common project, then WOM-acquired customers tend to exhibit heavy usage (social as well as personal). When the perceived value is relatively less collaborative and more individual, such as when Web page annotations are relevant mainly for the person making them, then WOM as a mode of acquisition is associated with lower personal usage and may not be a precursor for greater

WOM to affect adoption.

social usage. Lastly, our method allows for quantifying the magnitude of customer usage, where firms can gain insight to what extent customers from various routes are more active and to what extent various types of communication affects usage state transitions. This helps firms make informed decisions with regards to managing their marketing agenda. Therefore, managers would be wasting an opportunity to better leverage their customer base if they do not track this information.

2.6.4.2 Dynamics in State Transitions: Impact of Customer-to-Customer and Firm-to-Customer Communications

In terms of factors that impact the transition probabilities, we observe from Table A.9 that inbound sharing significantly transitions customers from the passive states to the active usage states. First, we analyze the personal and social state transition probabilities separately. We find that receiving an inbound share increases the transition probability of passive to active personal state by 3% of its original transition probability. Similarly, inbound sharing affects the transition probability of passive to active social state with a 13% increase. While the original transition probabilities are not very large themselves, these findings reveal that customer-initiated communications nudge customers to be more active for both types of usage, and the coefficient for the social usage process is an order of magnitude larger than the coefficient for the personal usage process. This reinforces the idea that WOM in the form of customer-to-customer communication is predominantly effective in *activating* customers into states where they exhibit higher levels of social usage. Furthermore, we observe that the coefficient of customer-to-customer communication (inbound sharing) on the social transition probability is an order-of-magnitude larger than the firm-to-customer communication (blog and tweets). This signifies, once again, that an effective way for firms to encourage customers to engage with their services may be to implement policies or design features that facilitate communication among customers post adoption.

That said, Table A.9 reveals that firm-to-customer communication is more effective at influencing customer transition probabilities in this context than the previous one analyzed (i.e., the Web-annotation service). For instance, we find that the coefficient on the variable Tweets, which relates to the change in the transition probability as a result of a firm's tweets, for customers already in the active state, is significant. In addition, the Blogs coefficient from the right-most column of Table A.9 is significant, suggesting that the company's blog posts are effective in transitioning customers from a passive to an active social usage state. The effect of a blog post in a given week increases the customer probability of moving from a passive to an active social state from 2.26% to 2.38%, a 5% increase. Thus, it seems customers attend more to the firm's communications in this instance than the previous one. This may be the result of the content of the firm's communication being perceived as more relevant to customers who sign up. While the actual magnitude of the transition probability is small, this could be due to the fact that the Tweets and blog posts are, by their broadcast communication nature, not directed at any individual customers. Since our model investigates the average effect of how each blog post influences a customer, it may appear weaker than the case where if we had data on who actually reads the blogs[16].

2.6.4.3 Heterogeneity and State Transitions

Similarly to the Web annotation context, we account for the observed heterogeneity in customers by incorporating individual-level characteristics into the intercept term of each customer's transition probabilities. In this case, we have information on each customer's propensity to invite others. This propensity is approximated by a binary variable that is 1 if the customer has sent out an invitation to other customers to join the service in the course of their tenure with the firm[17]. From examining the hierarchical parameters in Table A.10, we

[16]We duly note that this is a common problem that all data on advertising and broadcast communication would face, and it is a limitation of our data.

[17]We chose a time-invariant binary measure since the vast majority of customers (95%) send their first invite in the first three weeks, if they send it at all.

can see that on average customers who have invited at least one customer to join the service not only have a higher probability of transitioning from a passive to an active state, but also have a higher probability of staying in the active state once they reach it. More specifically, customers with an invitational propensity have probabilities of 3.50% and 6.56% of transitioning from passive to active personal and social states, respectively. This is compared with baseline customers transition probabilities of 0.34% and 1.14%, respectively. Similarly, invitation-inclined customers have 19.0% and 23.8% probabilities of staying in the active personal and social states, as compared to other customers' probabilities at 3.7% and 9.8%, respectively. This is consistent with the findings from the annotation service data, whereby the Invited Others variable has a similar significant impact on transition probabilities, likely for the same reasons we outlined in that context.

2.6.4.4 Summary from Both Contexts

What is important to take away from the comparison of the two applications is that, at a general level, the issue of route to adoption matters for being able to predict the intensity and pattern of customer usage post adoption. Furthermore, because increasingly Web-based companies offer services that allow customers to perform personal as well as social actions – it is important to account for these two usages yet at the same time jointly model them because the processes governing them may be correlated. In our data sets this is indeed the case. And lastly, various communications may impact usage. The fact that we find differences between these two services on many of the above characteristics should serve as a motivation for companies to indeed attempt to understand how they pertain to their specific context. As such, our model provides an effective way to achieve this goal.

2.7 Discussion

In this paper, we examine differences in customers' engagement level post-adoption using data from two Web service companies. We develop a novel multivariate Poisson HMM model to jointly estimate personal and social usage behavior and find differences in customers' usage behavior depending on how they joined the service. Our work is important because managers and entrepreneurs in the digital media industry have to care about customer usage, not just initial sign up– their advertising revenue, potential for future monetization, and ability to raise venture funding typically depend on the level of customer engagement.

Our work has implications for how managers attempt to acquire new customers, as not all acquisition routes are associated with equally engaged customers. In particular, blindly following advice to focus on finding ways to facilitate adoption through viral means may lower upfront marketing costs but potentially result in poor post adoption customer engagement levels. Building on this idea, because not all services offer the same type of value proposition to customers, it may be important to consider how the link between acquisition route and post adoption usage might be impacted by the nature of the service. In particular, some firms offer products that are primarily intended for personal uses, while others, for collaborative use (and some for a a combination of the two). For instance, Evernote is an on-line service that allows customers to take and archive notes on the cloud for retrieval on any Internet connected device. While it has social features that allow customers to share notes with each other, the primary value proposition of the software is "to make it easy to remember things big and small from your everyday life," in other-words, personal note-taking[18]. On the other hand, a product like Google Docs allows cooperation among users by creating a platform for the joint sharing and editing of documents, and therefore its value proposition has a strong collaborative benefit in addition to its personal-use advantages (such as access to one's files

[18]Flagship product description, Evernote Website: https://evernote.com/evernote/

on any internet-connected device). The question arises as to whether the way customers are expected to interact with the product or service and derive value from it matters in terms of which acquisition route will be associated with greater post adoption usage. Our findings suggest that this may indeed be the case. Consequently, companies would be well-advised to conduct an analysis similar to the one we performed here to understand the implications of different adoption routes on the engagement level of customers post adoption for their specific product or service.

For example, for a Web-annotation Web service, where it seems customers primarily plan to use the service in ways relevant to satisfy their personal needs, it may be highly relevant to consider acquisition through discovery-based routes such as Search. We speculate that customers who found out about the offering through Search had explicitly recognized a need for the service and actively sought it out; thereby being more highly motivated to use it. This may explain why customers acquired through Search are likely to exhibit the highest level of personal usage when the service provides strong individual-use benefits, such as a Web annotation service. Also, in this context, firms should be content with their customers from Mass-Invite. Being featured on a major "expert" site is still one of the cheapest and most effective ways to gain a large number of engaged customers. By comparison, for a cloud-based file storage service, where it seems customers also use a service to satisfy collaborative needs, methods that encourage WOM to acquire new customers, such as referral programs, can be relevant to implement. The problem that firms like this may face, especially in the early stages of the company, is that certain adoption routes like Search require customers to realize they "need the service." For instance, in 2009, the cloud storage company Dropbox abandoned paid search advertising entirely in favor of organic customer acquisition routes such as WOM referrals because the CEO realized "Search is great for harvesting demand, not for creating it.[19]" Our finding on WOM customers seems to support the notion that

[19]Eisenmann, T., M. Pao, L. Barley, 2012. "Dropbox: It Just Works." Harvard Business School Case.

they are "good" customers, in that even though they are only a little bit more active than customers from other routes, they cost less than paid Search customers to acquire. Thus, WOM as an acquisition mode may be more effective when the value proposition is aligned with collaborative and network benefits.

Beyond showing that adoption routes matter for subsequent behavior, we find that sharing behavior post adoption begets more usage – of both a social *and* personal nature. Specifically, WOM in the form of content sharing among customers post-adoption is more likely to result in customers transitioning into higher usage states. Given that we found this effect in both services suggests this is a rather robust conclusion. Moreover, firms should think carefully about designing their services so they can provide opportunities for sharing that may in turn be effective at increasing overall customer usage.

To summarize, we believe that investigating the relationship between adoption routes and subsequent usage is central to the value of a firm, especially for young, growth-phase start-ups. Moreover, we feel that it is crucial to distinguish between different types of usage, personal vs. social, yet simultaneously measure them. Indeed, we found significant differences among customers who joined through various adoption routes and also uncovered evidence that social features successfully promote customer usage by moving them to active, engaged states. Our work thus calls for a deeper understanding of the mechanisms that drive usage behavior, and for further exploring the possibility of segmenting customers along dimensions such as adoption route. In addition, future research can work with firms to specifically design interventions to encourage customers to generate more social behavior. Our modeling approach can be used to effectively study these issues.

Chapter 3

Designing Freemium: a Model of Consumer Usage, Upgrade, and Referral Dynamics

3.1 Introduction

Over the past decade, several software companies have increasingly turned to the subscription model for revenue generation. Firms often offer a limited but *perpetually free* version of their software in order to rapidly develop a large consumer install-base, with the expectation that users will upgrade to the paid premium version. This business model – referred to by industry as "freemium," a hybrid of free and premium – has been successfully adopted in Silicon Valley, and largely popularized among the newer generation of start-ups. According to the New York Times, freemium is one of the most prevalent business models among Web start-ups because relying on advertising as the sole stream of revenue might not be sufficient or sustainable.[1] To date, over 80% of the top grossing iOS apps have adopted the freemium model, with the largest freemium start-ups having acquired over hundreds of millions in

[1] "Ad Revenue on the Web? No Sure Bet," The New York Times, May 25, 2009. http://www.nytimes.com/2009/05/25/technology/start-ups/25startup.html.

venture funding.[2,3] The success of this business model has been further validated by many of the largest companies across multiple digital sectors, from online social networking sites such as LinkedIn, to music services such as Pandora and Spotify. Even media companies such as the New York Times (and its online pay wall) and mobile payment companies like Paypal utilize freemium. In the offline context, some consumer banks can also be characterized as using this model, with free checking accounts along with premium relationship accounts comprising the differentiated product offerings.

Freemium is often adopted because of its ability to attract a large number of users to its free version, i.e. as a customer acquisition strategy. Start-up technology companies facing capital constraints often choose this strategy over investing in advertising or using a sales force to obtain new customers. When coupled with a powerful consumer-to-consumer referral invite program, its effectiveness in acquisition is often magnified since a free product is easier to recommend. As a result, companies using this strategy see that a large percentage, often as high as 90% or more, of their consumer base are free users who do not contribute directly to the firm's revenues. While attracting a large user base is vital for establishing company value, firms must generate revenue for sustainability. Hence, the challenge requires balancing dual tasks: growing the consumer base by offering a free service and maintaining premium services to incentivize upgrades in order to stay profitable (Needleman and Loten, 2012).

The freemium business model raises several important questions in marketing that we investigate in this study, which uses a data set from a leading online file synchronization, backup and sharing service as its focus. Our research questions in this context include the following:

1. **Product Design:** How much value should the free product provide to consumers

[2] "Freemium apps continue to flourish in 2012." IntoMobile, December 22, 2011. http://www.intomobile.com/2011/12/22/freemium-apps-continue-flourishing-2012/.

[3] TechCrunch Crunchbase for Evernote, Pandora, 37 Signals, Spiceworks, and Dropbox, accessed June 5, 2013. http://www.crunchbase.com/company/.

relative to the premium product? A better free product would encourage more to join its service, but also would cannibalize sales of the premium version, by reduce the likelihood of upgrading to the premium version.

2. **Referrals:** What is the right referral bonus incentive to offer to consumers?

3. **Shared Product Use:** How does sharing influence customers' likelihood of upgrading to the premium product?

4. **Customer Value:** While free customers do not provide any direct revenue to the company, they have potential to generate revenue by either upgrading in the future, influencing others to upgrade via social features, or by referring new customers who may upgrade. How can we then value these "free" customers?

We use a unique panel data set of consumer activities to examine these questions relating to the freemium setting. There are multiple sources of value consumers obtain from the service. First, their files in their accounts are synchronized immediately across all connected devices, including computers, mobile phones and tablets. Second, the files are backed up in the firm's online storage repositories, and accessible from any Internet-connected computer using a Web interface. In the course of using the service, consumers add, delete and share files and also refer other consumers to the service; however, the primary revenue generating activity is when consumers upgrade from a free to a premium account, allowing for more storage capacity.

We develop a framework to characterize the dynamic behaviors of consumers in this setting, accounting for their motivations to undertake these various activities in their accounts. Consumers using the free version of the product in each period choose whether to upgrade to an annual or monthly plan that provides them an increased storage quota. In addition to this decision, they also have the choice of referring a friend, and obtain a referral bonus quota if the friend becomes a customer of the firm by subscribing to either the free or premium service. Consumers can also choose to delete files to maintain the limited space in their account, freeing up storage for future use. Thus, in the model consumers obtain a *flow utility* from the amount of storage current used, as well as decision or *action utilities* corresponding to the addition, deletion or sharing of files, and face a potential disutility related to the decision to upgrade their service by paying a price. Note that the benefits of upgrading accrue over time, since the action increases the constraint on storage from the free quota (2 GB) to the premium quota (50 GB). In this setting, inter-temporal dynamics and trade-offs play a very significant role, since the consumer has to predict future usage (and available storage) in determining the tradeoff of current decisions on upgrading, deleting or referring friends weighted against the costs of those decisions. Thus, we model consumers to be forward-looking, in order to trade off the cost of upgrading to a premium plan with the cost of finding and determining older files to delete, when newer content needs to be synchronized over time, as well as the likelihood that they would hit the limit of the free product. Consumers refer friends to the service, and while they receive a referral bonus when the friend joins, the referring consumer is not able to control the timing of joining, and thus forms an expectation over how many of her referred friends might join during each usage period. Our microfoundations-based model thus incorporates discrete and continuous choices for consumer upgrade, usage, and referral behaviors.

The estimation of our dynamic structural model involves several computational challenges, given that the state space has both continuous (amount of usage) and discrete (type of plan, referrals accepted by friends) dimensions. In addition, whereas upgrading and referral behavior are discrete choices, the amount of files to delete to create free storage space is a continuous decision, complicating the modeling and estimation process. We find that our likelihood function is highly irregular and jagged, making it important to use a robust method to obtain the global maximum. We use a conjunction of different approaches to overcome these computational and estimation-related challenges. First, we use a Bayesian methodology, using a modified version of the Imai-Jain-Ching (IJC) algorithm (Imai et al., 2009) that helps deal with the complex, highly irregular likelihood function. Second, we make extensive use of quadrature approaches to computing integrals for the likelihood to improve accuracy and computational time. Finally, to deal with the constrained continuous decision, we use analytical inversion to obtain the exact value of the unobservable shock corresponding to the decision using a stochastic Euler-equation based approach. In contrast, the grid inversion technique used by Timmins (2002) is not only more computationally demanding, but depends significantly on the accuracy of the discretization, and its use in a setting with highly non-monotonic likelihood like ours could be problematic.

We find that consumers on average obtain significant flow utility from having an amount of storage to synchronize and back up their files, which is expected since it is the primary value of the service. They also have a high and convex cost of deleting files, likely from being able to pick appropriate files that are no longer needed in order to maintain sufficient free storage capacity for future usage. Consumers also have a negative utility, or a cost of referring friends to the service, and weigh that against the probability that the referral will be accepted as well as the firm's offered amount of increase in baseline quota from the referral bonus incentive.

With these estimates, we then simulate counterfactuals that help deconstruct the consumer value to consumer personal usage and referral behavior, and in turn, we examine the impact of changing various design policies on consumer value. We find that the estimate of the value of a free consumer is approximately \$3, and at least 64% of this value comes from having a referral program, even in the absence of social features. A more unexpected finding lies in early evidence of a referral-personal usage synergistic effect, in which the existence of a referral program actually encourages consumers to delete less on average at any period, and thereby increases the probability that a consumer will upgrade at any given period.

Examining the impact of changing referral incentives is crucial because it can change the speed of product adoption, and therefore help the firm rapidly reach a critical mass of install-base. Even without considering the cost of supporting free consumers, we find that giving away too much referral incentive may actually decrease the overall output of referrals. If a consumer can receive the same amount of bonus space for one referral, which is sufficient for use, then what motivation is there to send out another two or three? We find that the shape of the consumer response of referral incentives is an inverted-U, implying that the firm should neither offer too small of an incentive (MB) because consumers may not find it worthwhile to refer anyone, nor too large of an incentive, because it may limit consumers' motivation to send out higher numbers of referral invites. We find the optimal static incentive amount to be approximately 450MB, which is double the amount observed in our data.

From a managerial perspective our findings have several implications. The existence of a large proportion of free consumers makes it difficult to assess firm value and future potential for a start-up entrepreneurial firm: the firm observes zero cash flow from free consumers, making it impossible to accurately project the future stream of cash flow for a majority of the consumer base. Without an accurate understanding of consumer value, it is difficult to price the product. Additionally, firms are often reluctant to drastically change the price of premium plans in fear of initiating backlash from existing consumers. Therefore, at best, firms can run small-scale pricing experiments on a limited subset of its consumers in a static setting in order to inform price change. However, because the profitability of these services depend heavily on repeated consumer visits and usage of features, without a model of consumer behavior, it is difficult to predict how consumers will respond in usage behaviors in order to compensate for the change in price. To further complicate the question of pricing, the existence of referrals and social features actually links the behavior of consumers together, therefore making it even more difficult to account for these factors in a pure experimental setting.

Taking a narrow view of the consumer can be highly inaccurate. Past research has shown that it is more expensive to acquire a new consumer than to maintain a current one. Firms that assign negligible worth to free consumers risk losing an important opportunity to maximize benefit from the self-perpetuating consumer base inherent to the freemium model, especially because it takes a long time for a free consumer to upgrade to the premium product. The value of the free consumers can be derived from four possible areas. The first area is their personal usage level. The more that a consumer uses the service, the more likely they will upgrade over time. This is the central assumption that many firms make when they utilize the freemium model, hoping that free consumers will eventually convert into premium consumers. Secondly, for services where social features exist, the social usage of these features can also contribute to consumer value. If two free consumers are sharing with each other over time, the social act of sharing actually contributes to the probability that both consumers will eventually convert to premium consumers. Third, free consumers add value via Word-of-Mouth (WOM) referrals. More specifically, in the context where a referral program exists, free consumers bring in other consumers, and over time, other consumers may eventually convert to premium consumers. Lastly, there could synergistic effects in the interaction among the three components. We provide a method to calculate the customer value of these free consumers, and the firm can account for and use this information in designing products more appropriately for their customers.

Our work has several limitations that can be explored for future research. First, we currently model consumer behavior conditional on adoption of the service. This can be allayed in the future by modeling the consumer's choice of adopting the service. In addition, the data set is from a period where this firm had minimal competition, so we do not model an outside option – allowing consumers to only choose between free and premium plans. It will be interesting to examine the competitive effects of this in the future. On the social usage front, sharing often begins when consumers form links with each other by sending share folder invites. We neither model that process nor distinguish with whom the consumers are sharing. Our specification examines only the magnitude of total inbound social usage and is agnostic to the type of individuals with whom consumers share files.

Related Literature

Our work intersects multiple domains of literature from a substantive viewpoint: *consumer-to-consumer referrals*, *product sampling*, *product line design and customer lifetime value*. In terms of referral incentives, past works have recognized the importance of managing referral programs (Buttle, 1998; Silverman, 1997). Biyalogorsky et al. (2001) explored the design of optimal referrals, and Ryu and Feick (2007), through experiments, showed that for strong brands, it is good to reward both the sender and the receiver of the referral in order to maximize referral rates, which is true of the referral program of our context.

Another related literature is product sampling. Prior studies on digital goods focused on the fact that they are experience goods, and contended that consumers require time to learn the value of these goods and services (Jain et al., 1995; Heiman and Muller, 1996; Lehmann and Esteban-Bravo, 2006; Heiman et al., 2001; Chellappa and Shivendu, 2005). Therefore, firms can influence the propensity of a consumer to adopt a service by providing free trials. Our context, however, differs temporally. In lieu of offering a limited-time free trial, the freemium model offers a perpetually free product, which can end up serving as a close substitute. Therefore the issue of cannibalization of the premium product is of significant concern. A growing body of literature is emerging that tackles these issues in the form of theoretical models that explore the economics of freemium (Niculescu and Wu, 2013), but given that the dynamic long-term effects are of first-order importance, the paucity of empirical (and even theoretical, with a few exceptions) research is striking.

In the CLV literature (Berger and Nasr, 1998; Gupta et al., 2006a; Fader et al., 2005; Schweidel et al., 2011), firms consider consumers based on a recurring stream of revenues, leading to a “lifetime” value associated with each consumer. Firms can then determine their acquisition and retention strategy based on these value estimates. The value of free customers has also been examined by Gupta et al. (2006b), who evaluate the average of aggregate value of buyers (to sellers and the two-sided platform) in an eBay-like online market place context, where the primary driver of the value of the free consumer comes from the nature of the two-sided platform, and the marketplace obtains fees from sales of goods that buyers bid and purchase from sellers.

From a methodological perspective, our work follows the stream of dynamic discrete-choice structural models in the tradition of Rust (1987). To our knowledge, while there are other models of discrete and continuous choice models (Hanemann, 1984; Song and Chintagunta, 2007), we are one of the first studies in Marketing to incorporate multiple discrete and continuous actions in a dynamic structural model and to estimate it using a technique that recovers the value function. While Bajari et al. (2007) and Ryan (2012) use BBL to estimate a dynamic structural model with both discrete and continuous actions, their estimation procedure fails to recover the value function of consumers. In addition, we present an analytic solution to the continuous action using the Euler Equation and Envelope Condition in order to ease the computational burden by avoiding having to numerically maximize over all possible continuous actions per discrete-choice action. Several authors have examined constrained continuous choices, most notably Timmins (2002), who uses discretization in conjunction with grid inversion to obtain the unobservable shock corresponding to the continuous choice. The closest work to ours in terms of discrete-continuous actions and analytic solution is Michelangeli (2008). However, the author’s work differs in two aspects: 1) the model is a finite-horizon dynamic programming model, and the approach is not readily applied to our infinite-horizon context, and 2) the author uses measurement errors associated with the continuous action, and therefore also limits the ability to conduct a wide range of counterfactuals. Because a major focus of our approach lies in the ability to conduct many counterfactual simulations, our model includes the ability to incorporate discrete and continuous actions with structural errors and to recover the value function.

Next, we detail the institutional context and the relevant features of the service. In section three, we describe the details of the data set that we use, as well as model-free evidence that supports our initial conjectures of the value of free consumers. In section four and five, we describe the model and the estimation procedure. In section six, we present the estimation results and the findings of various counterfactual analyses. Lastly, we conclude with a discussion of the managerial implication and limitations of this research.

3.2 Institutional Setting

The freemium company that provided the consumer data is a leading online storage company that stores consumer files in the cloud that synchronize across multiple devices (e.g. laptop, desktop, and mobile phone). The company was founded in 2008 and currently has hundreds of millions of users world-wide. Later in 2012, major competitors (such as Apple) entered the space by introducing similar versions of the service.[4] During the time period of our data, the cloud storage industry was fragmented amongst smaller providers. However, our firm quickly emerged as the dominant player in the consumer storage and syncing business, while many potential competitors acted primarily as security and backup services. Therefore, for our purpose, we regard the firm as a monopoly growing a captive user base.

The value proposition of the service is for consumers to store and sync files in the cloud and to share files with other users. Consumers do this by installing an application on their desktop. This application appears as a special folder on the consumer's desktop. Consumers can then add files to their account by simply dragging files into the folder, as one would

[4] All of these other competing services also use the freemium model.

do with any normal folder. Once the files are added to this folder, a copy of the files are then transferred onto the firm's servers and can be accessed through all of the devices that a consumer has the service software installed, or via an online interface (similar to the workings of a web-mail interface). While a consumer can access their files through different means (e.g. desktop, mobile devices, or web interface), the primary method that consumers use to access the service at the time of our observation is via their desktop, and therefore we focus on this point of usage in our analysis.

Once the files are stored in the account, they take up space that counts towards an account quota. When signing up, all consumers are presented with the choice of three different plans: Free, Premium-Tier-1 and Premium-Tier-2. Free is the basic plan where the consumer receives 2GB of quota; Premium-Tier-1 and Premium-Tier-2 are the premium plans where consumers would receive 50 and 100GB of storage, respectively. These premium plans work on a subscription basis, with the options of payment plans of monthly or yearly. The pricing and payment plan scheme is listed in Table 3.1. A majority of the upgraded consumers choose the Free and Premium-Tier-1 plans, and therefore we focus our analysis on these two plans, referring to the Premium-Tier-1 plan as the premium plan hereafter.

Plan	Monthly	Yearly
Free	-	-
Premium-Tier-1	$9.99/Month	$99/Year
Premium-Tier-2	$19.99/Month	$199/Year

Table 3.1: Payment Plans

When a consumer runs out of space, any files that a consumer adds into the service will no longer be uploaded to the server. Most importantly, all file synchronization stops, and since this is the primary value proposition of the service, the software is rendered virtually useless to the consumer. This is costly to the consumer, because then a consumer must take the time out to decide which files are not important, and move files out of their full accounts. And even after doing so, it will take time for the account to return to its functionality, as additional time is required for the local folders to resynchronize with the cloud account, proportional to the amount of files removed. Therefore, from conversations with consumers, most users tend to leave "cushion space" in their account. We observe this behavior in a majority of the consumers in the data.

To make space available in the data, a consumer has several choices to make. She can either a) delete files from her account, b) send referral invites to other consumers to join the service, or 3) choose a plan with higher quota. To delete files from her account, a consumer moves or deletes files from their account folder. This change is then synchronized onto the firm's servers.

A consumer can also earn additional storage space through the referral invite program. In order to use the referral program, a consumer can send out a unique link to other consumers who have yet to join; this unique link includes an identification number that links the invites back to the original sender. There is no limit on the number of referral invites that one can send out, but for the referral to count towards their quota bonus, the friends must join using the attached, unique link. In addition, the referral invite works "both ways", in that a consumer joining through the original invite also receives an additional 250MB of space. Hence, senders have the incentive to always include a link with their word-of-mouth, and receivers have the incentive to join via the links.[5] Therefore, while there may be some cases that consumers will not be identified as "referred" consumers in our data and bias our results of the effect of WOM and the usage behavior of non-referred consumers, this problem may be at a minimum. A consumer accepts the referral invite by signing up for the service. Once this is done, the original sender receives credit for the invite acceptance and earns an

[5]One might be concerned that this incentive system may encourage consumers who are already planning to join to actively seek out invites from other consumers. If this were the case, then these "willing" consumers may already have a favorable disposition towards the service, and are more likely to behave favorable towards the service (i.e. use the service heavily, send out more invites, more likely to upgrade to a plan later on). We acknowledge that this will bias our results upward and one way to possibly check for the existence of this behavior is to conduct surveys on existing consumer population.

additional 250 MB space. While this is a very effective way for consumers to gain space, a consumer can only receive credit for a maximum of 32 acceptances.[6]

Another key feature is for consumers to share files with existing consumers. For instance, if Alice wishes to share files with Bob, Alice would 1) create a sub-folder within their account, 2) send a share folder invite to Bob, and 3) place files into the shared folder. Once Bob accepts the share folder invite, whatever files exists in the share folder is then automatically synchronized across both Alice and Bob's accounts and will also count towards the quota on both consumer's account. While it is possible for consumers to share files with consumers who have not adopted using a special "Public" folder, a majority of the sharing is through this aforementioned private share folder feature, and therefore we define social usage as activity where a current customer places files into a share folder she has with another current customer.

Up to this point, we have focused primarily on the service from the consumer's perspective. It is now helpful to describe the firm's strategy decisions relating to the product: price of subscription plans, size of plan quotas, and size of referral incentives. Before the company's public launch, the firm ran pricing experiments on small subsets of its consumers in order to set the current pricing and timing plans. These experiments assumed that the consumer's behavior is static over time, and does not change during the observed period of data. Running counterfactuals off existing data is especially helpful in a setting like this to help inform the product design process. With the incorporation of customer heterogeneity, our methodology allows for them to rank order consumers according to the CLV, and then retroactively study the usage data for the most heavily used features of the most valuable

[6]As with any reward system, we have to be aware of of consumers trying to gain more space by "gaming" the system, mainly by creating clone accounts using additional email addresses. The firm is aware of this, and spends significant resources exactly to correct these gaming behavior in consumers by not rewarding false referrals. For instance, they can verify the source of two very different emails from the same consumer by verifying if these clone accounts install the software on the same device using machine footprints such as MAC addresses. Because of the efforts from the company in correcting gaming behavior, we assume that the integrity of our data is not comprised by this behavior.

consumers. Then, they can allocate company resources to make certain features easier to use (i.e. easier to share, easier to delete). The data-centric design philosophy is fairly popular among Silicon Valley start-ups, especially with online-gaming firms. Companies like Zynga have teams of data analysts to guide their game design decisions.

3.3 Data

The goals of this section are to show the characteristics of our data set and to describe the model free evidence that will help us identify the structural model that will be described in the next section. We obtained a random sample of 1,363 anonymous consumers who joined during in the first two years of the firm's history using a second-degree snowball sampling methodology. We underwent the following procedures to acquire the total sample of consumers:

1. Randomly sample a seed set of 50 people who have joined in the random sample seed window (Seed Group).

2. Add consumers who are connected (shared files or invited by) to Seed Group (1st Degree Group).

3. Add consumers who are connected to 1st Degree Group (2nd Degree Group).

The random sample seed window includes the first two years after the launch of the service. We then obtain all of these consumers' user activities from their join date until December 31, 2011. Our panel data includes the detailed click-stream data of these consumers over the four year period, which we aggregated into a weekly level. We observe a suite of consumer behaviors that are relevant to our problem.[7] These activities include:

[7] In order to protect the confidential nature of the data, usage statistics such as addition, deletion and storage have been normalized to the maximum observed number in the series in Table 3.2.

- Total number of files stored and the amount of storage.
- Amount of files deleted.
- Total amount of storage added.
- Number of referral invites that are sent to consumers who have not already joined the service.
- Number of referrals accepted each week.
- Plan choice and payment plan when upgrading.

While we cannot disclose the exact percentage of premium to total consumers, many freemium companies observe premium-to-total consumer ratios ranging from the single digits to over ten percent.

Statistic	**Mean Across Sample**	**SD**	**MIN**	**MAX**
Number of Consumers	1,349	-	-	-
Total # of Observations	155,279	-	-	-
Time Periods	115.107	18.527	93	206
Average Storage	0.0121	0.0542	0	1
Average Referral Sent	8.503	48.76	0	966
Average Referral Accepted	1.831	7.626	0	155
Average Referral Rate (Sent/Accepted)	0.169	0.321	0	1

Table 3.2: Summary Statistics of Consumers

3.3.1 Model-Free Data Patterns

In this section we examine the data patterns of consumer behavior with the goal of clarifying the key data features that our model needs to characterize and inspire the major design choices of our model. Ultimately, we observe two patterns in the data that justify the value of a free consumer. First, consumers upgrade themselves over time as they become closer to reaching quota. This is the first value of the free customer. Second, even if free consumers never upgrade themselves, they may recruit an additional consumer whom may eventually upgrade.

3.3.1.1 Free Customers Upgrade Over Time From Personal Usage

First, we first examine the ratio of free to premium consumers over time. The left panel of Figure 3.1 shows that the growth of free consumers greatly outpaces the growth of the premium consumers. This indicates that free consumers cannot be overlooked. There are more consumers who begin as free consumers and convert to premium than there are consumers who join as premium consumers from the beginning.

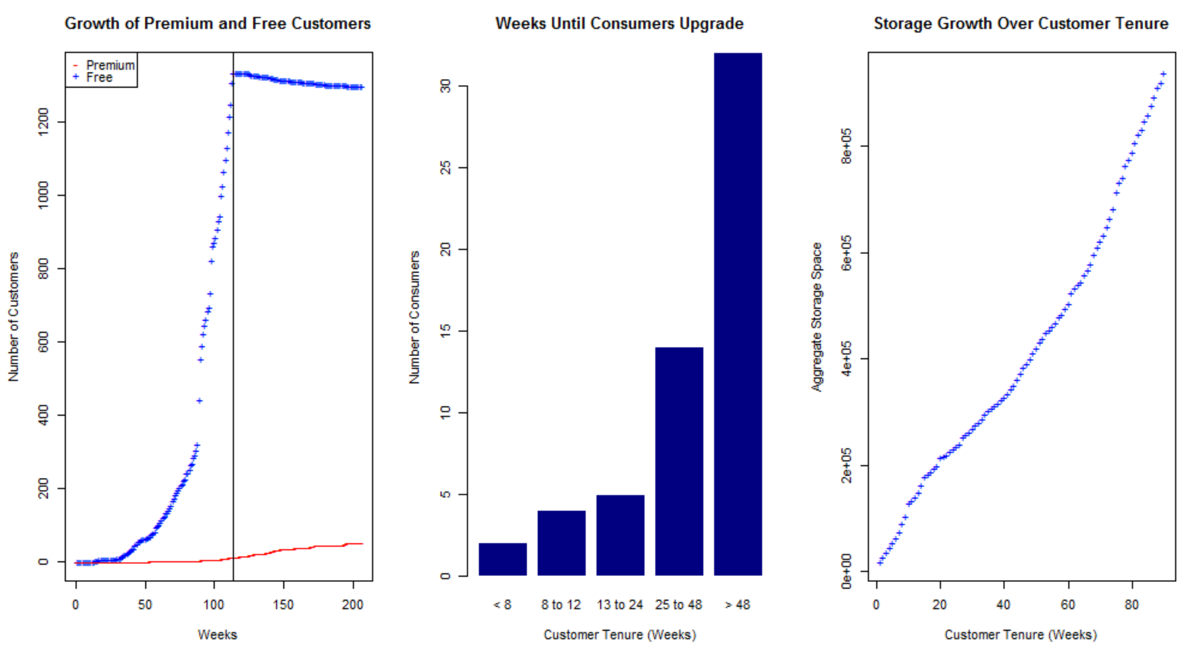

Figure 3.1: Upgrade Patterns

The middle panel confirms that a majority of consumers upgrade after using the service for several weeks. There are only two consumers out of the entire group of premium consumers in our sample who upgrade within the first six weeks of joining the service. The graph on the right shows the growth of the aggregate consumers storage within their first

90 weeks. This graph is from the perspective of the consumers, in that we see the average storage used per consumer grow over time. This suggests that consumers begin using the service as free consumers and later upgrade to a premium plan once they store enough files.

The customer mix, i.e. the fraction of consumers who choose the premium product is a critically important variable in the freemium business model. We detail the dynamic variation of the consumer mix in Figure 3.2. We observe a non-monotonic inverted-U-shaped pattern for the fraction of premium users, suggesting that the firm would have to be patient for customer acquisition to be converted to revenues. If the company incurs costs for each "free" customer, then it might well see diminishing profitability when the fraction of premium customers drops over time.

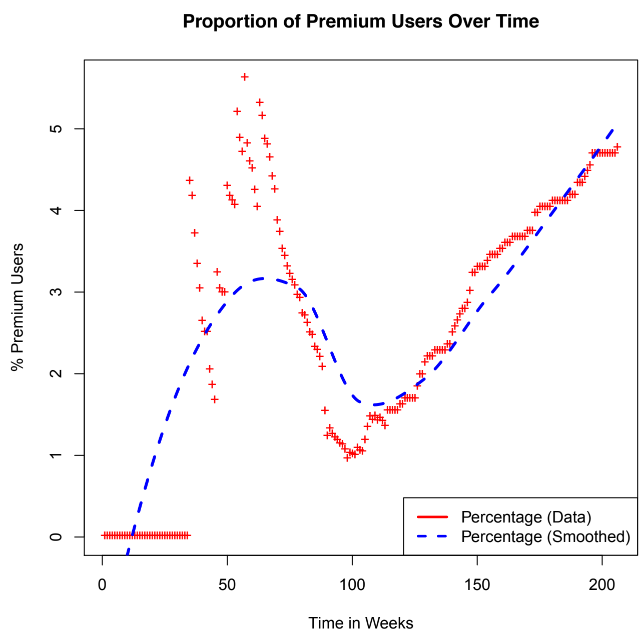

Figure 3.2: Dynamics of Customer Mix over Time

3.3.1.2 Free Customers Bring in Premium Customers Through Referral Invites

We now examine the consumer referral behavior and whether referral invites recruit consumers who later convert to premium consumers. The goal of referrals is to bring in new consumers, with the hope that these consumers will upgrade to premium. Informal interviews with management indicate that 1) the referral program is effective at acquiring new consumers and 2) the growth of referral parallels with the number of new consumers. The latter reflects the viral nature of referral invites, as the number of consumers has the potential to grow exponentially with a viral coefficient greater than 1.

First, the top left panel in Figure 3.3 shows the ratio of consumers who joined directly versus those who joined through referrals. From this graph we observe that referral consumers reach almost 40% of the number of consumers who joined directly, and the proportion is increasing over time. This confirms the company's belief that the referral program is quite effective. Furthermore, in the top right graph we juxtapose the total number of referrals sent alongside the number of referred consumers. We see that the number of referred consumers begins small, and that the slope of growth increases dramatically from weeks 60 to 114.[8] This further endorses the significant impact of the referrals. Interestingly, it takes only a few referrals to obtain a large influx of new consumers, suggesting that even small increases in referral rates can have cascading effects.

[8]The number of new consumers stops at week 114 because that is the cut-off point in any new consumers our sampling scheme.

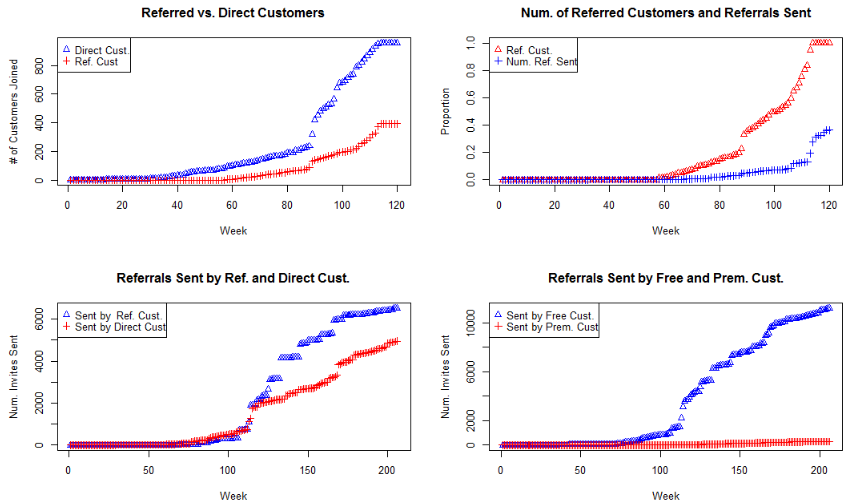

Figure 3.3: Referral Data Patterns

Along the theme of viral growth, in the bottom left panel we see that the average number of referrals sent by referred consumers is greater than the number of consumers who joined directly. This is added confirmation of the potency of the referral program, and evidence of the rapid snowball effect of the original referral sent by the consumer who joined directly. The bottom right panel, perhaps the most interesting of the four, shows that the total number of referral invites sent by consumers before they upgrade greatly outnumbers the total invites sent by consumers who have upgraded to a premium plan. This confirms our intuition that the success of a referral program will depend on a large number of free consumers willing to send out invites.

Lastly, in Figure 3.4, we plot the aggregate growth of premium consumers alongside the growth of the premium consumers who are referred by other consumers. Overall, the referred consumers account for over twenty percent of all premium consumers, suggesting that there may be a significant value to the referral program.

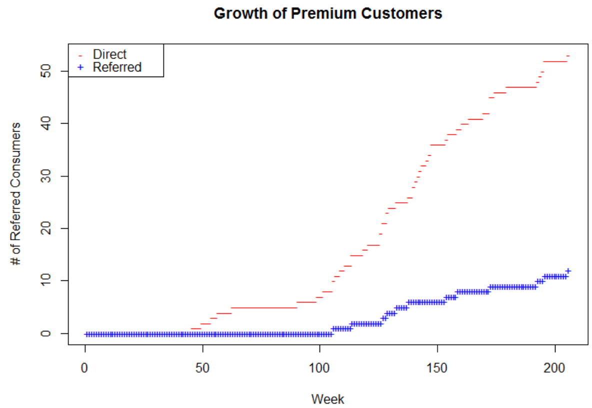

Figure 3.4: Breakdown of Premium Customers Who Joined Directly vs. Referred.

3.4 Model

The freemium business model is defined as a firm offering at least two differentiated variations of their service, one version with limited features but perpetually free, and the other versions with enhanced features at the cost of a subscription fee. The free and premium plans are identical in terms of quality, and the products differ only in terms of the additional premium features offered (e.g. increased storage capacity).

3.4.1 Motivation for Structural Modeling and Sources of Dynamics

A majority of users of freemium services are non-paying consumers who are initially enticed by the free plan. The most valuable asset of free consumers remains in their potential—the potential to upgrade, the potential to refer friends to join the service. Understanding the factors that influence such consumer behavior over time is key to the success of freemium. A structural model that characterizes the dynamic response of consumer behavior is therefore critical for the following reasons. First, we need to account for four different consumer choices – referral, plan-choice, personal deletion, and social deletion – in an integrated model of consumer behavior. Customers send out referral invites to share their enthusiasm for the product with friends and to earn additional free space. However, their motivation to upgrade or delete can be diminished by the extra space earned from referrals. In addition, consumer deletion behavior inherently differs according to their chosen plans. Those who have chosen the free plan may have to delete more in order to maintain enough space to store files, and those who have chosen the upgraded plan do not have to delete much due to the wealth of new space. The four decisions are endogenous, and we need a methodology that can account for this.

Second, we need a structural model because we wish to conduct counterfactual experiments to simulate the value of the free consumer and to observe the effects of changing firm policies on consumer behavior. With atheoretical models, the outcomes of changes in certain product design variables, such as price, free quota size, and referral incentives, cannot be readily characterized as there are often no variations in these variables during the observed data period. A model based on microfoundations of consumer behavior uses theory about consumer behavior to recover primitives of consumer preferences, which are likely to be invariant to changes in these product design and other policy variables. These preference parameters can then be used to evaluate how consumers would make choices in a counterfactual scenario, enabling us to provide recommendations that are under managerial interest.

A fundamental process we need to account for in our model is the inter-temporal tradeoff in upgrade, referral, and deletion behavior. The source of dynamic behavior comes from a combination of three factors: a) uncertainty in file addition, b) uncertainty in ease of upgrade, deletion and referral, and c) substantial penalty of a full account.

3.4.1.1 Uncertainty in File Addition

First, consumers face uncertainty when anticipating the number of additional files needed for storage each period, as some weeks require less than others. In order to store files, a consumer requires space, and therefore must select a plan that fits the amount of data she will receive in the current period. When faced with the space constraint, a consumer can either upgrade to a higher-space plan, gain additional space from referring others to join, or delete files to make space.

3.4.1.2 Uncertainty in Ease of Upgrade, Deletion and Referral Decisions

The upgrade decision is complicated by dynamic factors that facilitate or complicate the decision to upgrade to a premium plan from week to week. We see these examples of weekly unobserved factors from our discussion with current consumers. One such example is a consumer waiting for budget approval so he can pay for the premium plan. For all of the weeks prior to the budget approval, it is "harder" for the consumer to upgrade. However, once the budget is approved, even if a customer's usage is not close to quota, he upgrades. Another example is a consumer who anticipates leaving for an upcoming trip. It is easier for consumers to upgrade while in front of a computer, compared to when they are away during vacation.

If a consumer chooses not to upgrade, she has the choice to gain more space from deleting files. However, in the same manner as upgrades, consumers also faces an uncertainty in the ease of deletion. Users of the service have expressed that certain weeks are easier to delete while other weeks are harder (i.e. deadlines at work or exams at school), and we observe this lumped deletion pattern in the data. This causes the user to continually make the trade-off of whether to upgrade today in order to save the streams of deletions that she has to make in the future. Therefore at a certain point it may be optimal for a consumer to upgrade in order to outweigh the cost of continually deleting in the future.

A consumer faces two forms of uncertainty with regard to referrals. The first is the ease of referral from week to week. Second is the uncertainty of when the invitation will be accepted. Therefore, a consumer cannot simply send out invites the week that she runs out of space, and must consider in advance how likely her invites will be accepted and if her usage will be sufficiently below quota by the time the additional free space is acquired.

3.4.1.3 Substantial Penalty of a Full Account

Lastly, a final source of dynamics is the need to anticipate periods of increased use. The consequences of a full folder in a customer's account include termination of file syncing and subsequent freezing of the account, which renders the service essentially useless. Therefore, our model should account for the fact that the consumer incurs a substantial cost due to the sudden drop in the entire value proposition of the service.

3.4.2 Consumer Decisions

The main value proposition of the service is to help consumers store and sync files in cloud storage (for now we have yet to incorporate social usage). To do this, consumers need storage space in their accounts. The size of this account depends on the different plans that consumers choose. At the time of the data set, the company offered two different plan sizes relevant to our research: 1) 2 GB for free and 2) 50 GB for $9.99/month

In each time period t (week), a consumer $i \in \{1, ..., N\}$ chooses four decisions to maximize her utility: 1) whether to upgrade to a premium plan or remain a free consumer, 2) how many consumers to send referral invites, 3) how many MB's of files to delete from her personal folder, and 4) the how many MB's of files to delete from her shared folders. The time line of events can be summarized in Figure 3.5 and is elaborated below.

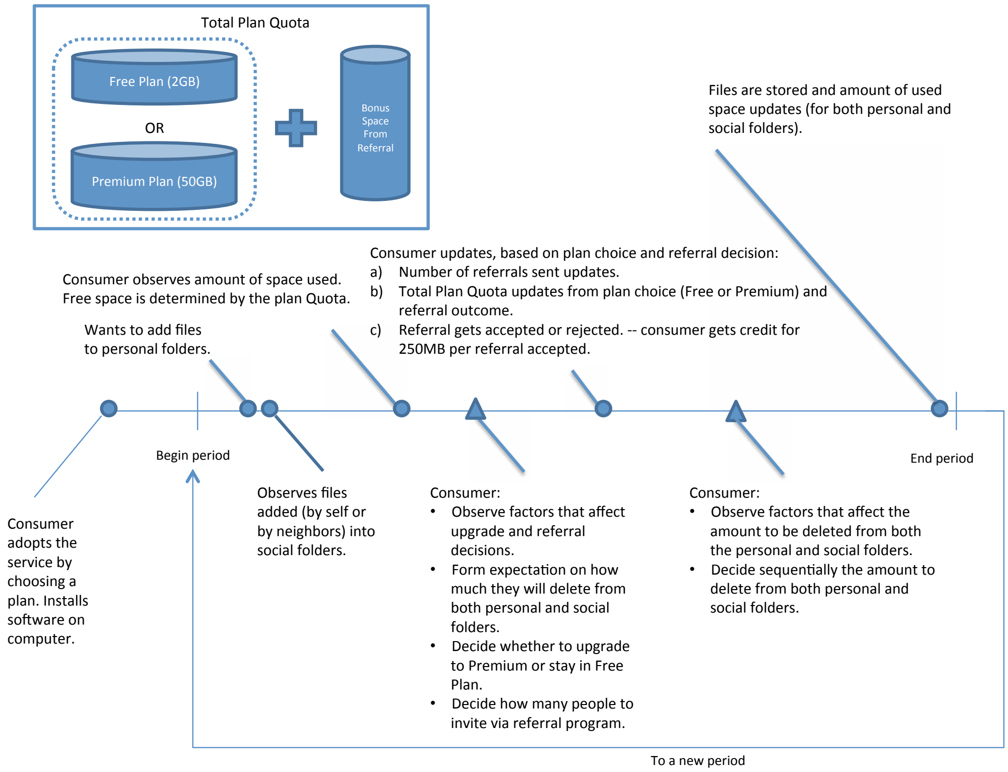

Figure 3.5: Time Line of Events

At each week t, a consumer i:

1. Customer observes all state variables: x_{it}, z_{it}, R_{it}, R^a_{it}.

2. Observes exogenous shocks a^p_{it}, the amount of files that need to be added to a consumer's personal (any non-shared) folders, and a^s_{it}, the amount of files added to their shared folder by the shared neighbors.

3. Observes discrete-choice upgrade/referral shock $\epsilon_{it}(y_{it}, r_{it})$ and chooses decisions y_{it} and r_{it} simultaneously.

4. State variable z_{it} updates base on the upgrade decisions y_{it}. State variables R_{it} updates base on referral decision r_{it}.

5. r^a realizes based on the total number of outstanding invitations $(R_{it} - R^a_{it})$.

6. State variable R^a_{it} updates based on realized r^a.

7. Observes continuous-choice deletion shocks ν^p_{it} and ν^s_{it}, and chooses deletion decisions p_{it}, and s_{it} sequentially. ν^p_{it} and ν^s_{it} are factors, unobserved to the researcher, that affect a consumer's personal (p_{it}) and social deletion decisions (s_{it}).

8. Update state variable x_{it}, where x_{it} is the total amount of space stored in the personal and social folders combined at the end of the period.

At the beginning of the period t, a consumer observes all state variables, and then she observes a^p_{it}, the amount of files (MB) that needs to be added to her personal folder, and a^s_{it}, the amount for her social folder . We start the period with the exogenous addition because it is a natural departure point for the consumer process. Addition of files is the fundamental value proposition of the service and all other consumer decisions depend on the amount of files that need to be added. Then, the consumer makes the joint decision of plan choice (y_{it}) and the number of referral invitations to send (r_{it}). For the upgrade option, we specifically look at two options:

$$y_{it} = \begin{cases} 1, & \text{if customer upgrades with the monthly payment option,} \\ 0, & \text{if customer does not upgrade.} \end{cases}$$

For options $y = 1$, a consumer upgrades from a free to a premium plan and pays a price P^m. We model both the plan price and the plan length because they are both policies that can be changed by the company and therefore are potential subjects of interest in the counterfactual simulations.

Simultaneously, a consumer determines the number of referral invitations to send to other consumers who have yet to sign up. This is modeled as a discrete count variable bounded by R^{max}:

$$r_{it} \in \{0, 1, ..., R^{max}\}.$$

Then, a consumer observes the continuous-choice specific deletion shock ν^p. This is interpreted as the weekly unobserved factors that make deletion easier or harder. For instance, one week a consumer may find it harder to delete from her folders because she is traveling for work, so ν^p would be a low value. Another week could be spring cleaning, which makes it easier for a consumer to delete unnecessary files, rendering ν^p high. ν^s_{it} can be interpreted in a similar fashion.

After these decisions and shocks are realized, the consumer then realizes the value of r^a, the number of accepted referrals in the current period. r^a follows a binomial distribution, with the parameters: 1) p^a, the empirical acceptance probability in the population, and 2) $R_{it} - R^a_{it}$ the number of outstanding referral invites, where R_{it} is the cumulative number of referral invites sent, and R^a_{it} is the cumulative accepted referral invites. Lastly, the four state variables x_{it}, z_{it}, R_{it}, and R^a_{it} are updated at the end of the period. x is the cumulative amount of space used in a consumer's account, and z is the number of premium plan weeks left in their account. By default, z is 0 for all consumers who are in the free plan.

3.4.3 Period Utility Function

Now we describe how each part of the time line component contributes to a consumer's period utility. Note that the consumer's decision making process is not solely based on this utility, but is based on inter-temporal trade-offs as described in §3.4.6. We suppress the i subscript for expositional clarity, even though the model is at an individual level. At each time period t , a consumer gains utility from having files stored in their account and from having folders that are free of files they no longer need. In addition, they incur a cost for the effort to delete files, to pay to upgrade to a premium account, and to send out referral invites to their friends.

Let $\psi(x_t;\theta)$ be the utility contribution of storage, specified as

$$\psi(x_t;\theta_1) = \theta_1 x_t,$$

where θ_1 is the file storage benefit coefficient. Following our reasoning in the model section, we assume linear, first-order effects of storage of files in personal and social folders.

The modification to the utility contribution from personal and social deletion is straightforward, as we employ a flexible quadratic utility for both deletion actions. Let $\chi^d(p_t, s_t, \nu_t^p, \nu_t^s; \alpha)$ be the deletion utility, specified as

$$\chi^d(p_t, s_t, \nu_t^p, \nu_t^s; \alpha) = \alpha_1^p p_t^2 + \alpha_2^p p_t \nu_t^p + \alpha_1^s s_t^2 + \alpha_2^s s_t \nu_t^s,$$

where $\alpha = (\alpha_1^p, \alpha_2^p, \alpha_1^s, \alpha_2^s)$, is the vector of coefficients of the personal and social deletion cost. We then express the utility specification from Equation 3.1:

$$\begin{aligned} u(\mathbf{D_t}, \mathbf{S_t}, \epsilon_\mathbf{t}, \nu_\mathbf{t}^\mathbf{p}, \nu_\mathbf{t}^\mathbf{s}; \Theta) = \psi(x_t;\theta_1) + \chi^d(p_t, s_t, \nu_t^p, \nu_t^s; \alpha) + \rho r_t^2 + \\ \gamma P^m 1[y_t = 1] + \epsilon_t(y_t, r_t). \end{aligned} \tag{3.1}$$

where $\mathbf{D}_t$ is the vector of decision variables such that $\mathbf{D_t} = (y_t, r_t, p_t, s_t)$, $\mathbf{S_t}$ is the vector of observable state variables such that $\mathbf{S_t} = (x_t, z_t, R_t, R_t^a)$. ϵ_t is the vector of discrete-choice private shocks related to the joint upgrade and referral decision, such that $\epsilon_t = (\epsilon_t(y=0, r=0), ..., \epsilon_t(y=2, r=r^{max}))$. Θ is the vector of structural parameters to be estimated such that $\Theta = (\theta, \alpha_1^p, \alpha_2^p, \alpha_1^s, \alpha_2^s, \gamma, \rho)$. The storage utility term, a *flow utility*, contributes to a consumer's utility each period that files are stored in the account.[9] The deletion, referral and upgrade utility terms, all *action utilities*, only contribute to a consumer's utility when the actions are taken each period. Below we examine each of these components of the consumer's utility.

Utility from Using the Service (Storage Utility)

At each period, a consumer receives utility from using the service. x_t denotes the cumulative MB used for personal folder storage. While one can assume various functional forms on this benefit, we assume a linear benefit specification on x_t for parsimony, since we wish to simply capture the relationship that consumers gain more utility from having more files in their folder.

Utility from Deletion

The utility specification must satisfy three aspects:

1. Consumer incurs a cost of deleting files from her personal or social folders.

2. A consumer can only delete as much as there are files in the folder.

3. A consumer is forced to delete the amount of files that causes the account to exceed plan quota.

[9] The flow utility here is similar to a consumer receiving flow utility in each period after purchasing a durable good (e.g. a car or television).

We employ a flexible quadratic utility form for deletion to capture the potential convex cost to deletion. The maximum amount that a consumer can delete at any period is capped by $x_t + a_t^p + a_t^s$, the amount of files in a consumer's account at time t. The constraints on how much a consumer can delete is enforced implicitly in the constraint correspondence in the value function specified in the dynamics section below.

Utility from Referring Other Customers

The utility from referring other consumers must reflect three aspects:

1. For each accepted referral, the consumer gains the referral bonus quota m MB's of space.
2. A consumer faces an uncertainty about the acceptance of each referral.
3. The consumer incurs a transactional and reputational cost for inviting another consumer.

r_t is the number of referral invites that a consumer sends at time t. The benefit of an accepted referral is reflected in the state variable Q_{it}. The uncertainty that customer faces with regard to how many referrals will be accepted in each period is captured via the binomial distributed shock r^a, with parameters $n = R_t - R_t^a$ and $p = p_r$, that the customer realizes at the end of each period. Lastly, ρ is the coefficient of the convex transactional cost that a customer incurs for sending out an additional invite. The convex cost reflects the fact that it becomes increasingly difficult for customers to think of an additional friend to invite in any given week.

3.4.4 Alternate Configurations of Decision Timing

While decision timing is unlikely to significantly alter the results of a dynamic infinite horizon model, we still discuss the implications of other possible orders of the consumer decisions.

First we consider the placement of the deletion decisions. There are two possibilities: 1) place the deletion decisions before the upgrade/referral decisions or 2) model all four decisions simultaneously. First, if we were to place deletions first, we would be assuming that customers have the same deletion behaviors regardless of whether they opt for status quo or they gain more space from upgrading/referring. Given the institutional context, this is inconsistent with their understanding of actual consumer deletion behavior – a key reason why customers refer or upgrade is because they want more storage space. Therefore, customers do not need to delete as much in a premium plan than a free plan.

In addition, we consider the ordering of placing the r^a variable after all of the major decisions. The reason for this is that a majority of the realized acceptances, as observed in the data, do not come immediately after the invitations are sent out. Therefore, it is important to capture this uncertainty in referral acceptance that customers face when they make the upgrade, referral and deletion decisions. There are two other places we can place this variable: 1) at the beginning of the period, or 2) between the upgrade/referral decision and deletion decision. In the first case, since we have a dynamic model, we end up with the same Bellman equation as if we had placed the realization in the end. As a result, the implications of the model stay the same. If we were to place r^a between the upgrade/referral and deletion decision, then we would specifically assume that the referrals per period must be realized before a customer makes the deletion decision. This is a strong assumption, since it assumes that the uncertainty in referral acceptance has no effect on a customer's deletion decision. Aside from being a stronger assumption than what is currently assumed, this also makes the solution of the value function more complex. As a result, we settled on our current specification of the time line as a fairly reasonable assumption.

3.4.5 State Evolution

The state variable x_t keeps track of the total amount of files in MB's that is stored in a consumer's account. This is updated via the linear law of motion $x_{t+1} = x_t + a_t^p + a_t^s - p_t - s_t$, which is simply the sum of the amount of files observed at the beginning of the period and the observed addition amount, subtracted by the amount deleted in the particular period.

The state variable z_t keeps track of the number of periods until the consumer's next payment. z_t is set to 4 if he chooses plan 1. z_t decreases by 1 each period. The state evolution for z_t is specified as:

$$z_t = \begin{cases} 0, & z_{t-1} = 0 \ \wedge \ y_t = 0 \\ 4, & z_{t-1} = 0 \ \wedge \ y_t = 1 \\ z_{t-1} - 1, & z_{t-1} > 0 \end{cases}$$

Only a portion of the referral invites are actually accepted, and we keep track of the total number of successful invites as R_t^a. R^{max} is the empirical number of maximum per-period invitations in the data. Whenever an invite is accepted, a consumer gains an additional m MB's of space to their quota. More specifically, we define the variable Q_{it} as a function of the total number of successful invites (R_{it}^a), the baseline amount of space (Q^{free}), referral bonus capacity (m), and incremental amount of space provided by the premium plan ($Q^{premium}$):

$$Q_{it} = Q^{free} + mR_{it-1}^a + \mathbf{1}[z_{it} \geq 1 \ \vee \ y_{it} = 1 \ \vee y_{it} = 2]Q^{premium}$$

The following Table summarizes the state variables and the corresponding laws of motion.

State Variable and Shocks	Description	Type	Law of Motion / Distribution
x_{it}	Cumulative MB used in Personal Folders	Observed	$x_{it+1} = x_{it} + a^p_{it} + a^s_{it} - p_{it} - s_{it}$
R_{it}	Cumulative number of referrals sent	Observed	$R_{it+1} = R_{it} + r_{it}$
R^a_{it}	Cumulative number of accepted referrals	Observed	$R^a_{it+1} = R^a_{it} + r^a_{it}$
z_{it}	Number of premium weeks left	Observed	$z_{it+1} = z_{it} - 1$
Q_{it}	Total Quota	Observed	$Q_{it+1} = Q^{free} + mR^a_{it} + \mathbf{1}[z_{it} \geq 1]Q^{premium}$
$\epsilon_{it}(y,r)$	Upgrade and referral decision shock	Unobserved	Type I Extreme-Value(0,1)
ν^p_{it},ν^s_{it}	Deletion decision shock	Unobserved	Log-Normal(0,1)
a^p_{it}, a^s_{it}	Personal and Social Addition shocks	Observed	Non-Parametric Distribution

Next, we describe the full dynamic model.

3.4.6 Dynamics in Consumer Decisions

The dynamics in the consumer's decisions stem from the inter-temporal tradeoff of the consumer's current benefit versus the future benefits of upgrading to a premium account, deleting files to gain free space, and referring other consumers due to social and practical benefits. We therefore model this tradeoff as the sum of discounted future period utilities:

$$\mathbf{E}_{a^p,a^s,r^a,\epsilon}\left[\max_{(y,r)}\sum_{t=0}^{\infty}\beta^t\mathbf{E}_{\nu^p}\left[\max_{p}\mathbf{E}_{\nu^s}\left[\max_{s}u(\mathbf{D}_{it},\mathbf{S}_{it},\epsilon_{it},\nu^p_{it},\nu^s_{it};\Theta)\right]\right]|\mathbf{S}_{it}\right] \tag{3.2}$$

where β is the assumed discount factor for all consumers, and the utility function is specified in Equation 3.1. The solution to the above dynamic programming problem is the same as the solution to the Bellman equation, which is hereby referred to as the value function:

$$V(\mathbf{S},\epsilon,a^p,a^s;\Theta) = \max_{y,r\in\Gamma(z)}\mathbf{E}_{\nu\mathbf{p}}\left[\max_{p\in H_p(x,z,a^p,a^s,R^a)}\mathbf{E}_{\nu^s}\left[\max_{s\in H_s(x,z,a^p,a^s,R^a)}u(\mathbf{D},\mathbf{S},\epsilon,\nu^p,\nu^s;\Theta)+\beta\mathbf{E}\left[V(\mathbf{S}',\epsilon',a'^p,a'^s;\Theta)\right]\right]\right] \tag{3.3}$$

The integrated Bellman equation, EV, expresses the fixed-point that we solve to derive the solution of the expected value function with the discrete-choice shocks ϵ as well as a^p and a^s integrated out:

$$EV(\mathbf{S};\Theta) = \mathbf{E}_{a^p,a^s,r^a,\epsilon}\left[\mathbf{E}_{\nu^{\mathbf{p}}}\left[\max_{p\in H_p(x,z,a^p,a^s,R^a)} \mathbf{E}_{\nu^s}\left[\max_{s\in H_s(x,z,a^p,a^s,R^a)} u(\mathbf{D},\mathbf{S},\epsilon,\nu^p,\nu^s;\Theta) + \beta EV(\mathbf{S}';\Theta)\right]\right]\right] \tag{3.4}$$

The difference in interpretation for the expected value function is that it is the value function prior to consumers observing all shocks, and therefore is expressed only as a function of the state variables $\mathbf{S}$. $\Gamma(z)$ is the the choice set for the discrete decisions y and r, specified as $(y, r) \in \Gamma(z)$ where:

$$\Gamma(z) = \begin{cases} \{0,1\} \times \{0,...,R^{max}\}, & \text{z=0} \\ \{0\} \times \{0,...R^{max}\}, & \text{otherwise} \end{cases},$$

again reflecting the fact that when a consumer is already on the premium plan $(z > 0)$, she has no choice to make regarding the product and so the plan is trivially set to 0.

The choice set for the continuous deletion decisions p and s is specified by the correspondence constraints $H_p(\cdot)$ and $H_s(\cdot)$:

$$H_p(x, z, a^p, a^s, R^a) = [\max(0, x + a^p + a^s - Q(z, R^a)),\ x + a^p + a^s]\,,$$

and

$$H_s(x, z, a^p, a^s, R^a) = [\max(0, x + a^p + a^s - p - Q(z, R^a)),\ x + a^p + a^s - p]\,,$$

reflecting the fact that consumers cannot delete more than the total amount stored, and must delete a sufficient amount so that they do not exceed their quota, e.g. when the consumer has stored $x = 1.5$ GB and wants to add $a^p = 1$ GB, with zero social addition, and has a baseline quota of $Q(0,0) = 2$ GB, then, she must delete at least 0.5 GB ($x + a^p - Q(0,0)$) of data in order to stay within the limit. Observe that if the consumer had chosen to upgrade earlier in the period to a monthly premium plan, she could have chosen $y = 1$ resulting in $z = 4$ and a corresponding quota of $Q(4,0) = 50$ GB, allowing for a much higher degree of flexibility.

3.4.7 Identification

γ is the price coefficient and is identified primarily by y_t, the upgrade decision. It would be highly negative if the average number of upgrade from week to week is low. Note that given the fact that the price does not change in the entire observation period, we identify this parameter from the population of consumers who eventually upgrade. In addition, the dynamics in x_t in conjunction with y_t also help identify this parameter. If the magnitude of γ were high, then we would see more occurrences of x_t that are close to the free quota, meaning the consumers will only upgrade when they are close to quota. However, γ would be low if, on average, upgrades occur when the average level of x_t are low.

θ is the marginal utility of storage. It would be high in magnitude if the average level of x_t is high in conjunction with low levels of p_t and s_t. The parameter would be low in magnitude if the average level of x_t is low, and we see high levels of p_t or s_t. This parameter is identified via the dynamics in x's and (p,s)'s. We would not be able to separately identify this parameter from (α_1^p, α^s), the costs of deletion, if the problem were static.

α_1^p and α_1^s are the costs to personal and social deletion. This parameter is identified from the variation in a consumer's weekly deletion behavior. The parameter would be highly negative if the average amount of deletion is low, meaning it is very costly for consumers to delete. On the other hand, the parameter would be low if the average p and s were high from week to week, meaning that it is not very costly for consumers to delete.

Lastly, ρ is the quadratic cost to referral. The parameter is identified from the weekly variation in r, the consumers' referral behavior. The parameter would be highly negative if the average amount of referrals is low, meaning that it is costly for consumers to invite other consumers. The parameter would be low in magnitude if the average amount of referrals is high, meaning that it is not very costly for consumers to send out invites.

3.4.8 Heterogeneity

In this section, we detail our unobserved heterogeneity specification, following on the utility description from 3.1. We define the vector of parameters Θ_i as the vector of individual-level parameters such that $\Theta_i = \{\theta_i, \alpha^p_{i1}, \alpha^p_{i2}, \alpha^s_{i1}, \alpha^s_{i2}, \gamma_i, \rho_i\}$. We further define a distribution over Θ_i such that:

$$\Theta_i = \begin{pmatrix} \theta_i \\ \alpha^p_{i1} \\ \alpha^p_{i2} \\ \alpha^s_{i1} \\ \alpha^s_{i2} \\ \gamma_i \\ \rho_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \theta \\ \alpha^p_1 \\ \alpha^p_2 \\ \alpha^s_1 \\ \alpha^s_2 \\ \gamma \\ \rho \end{pmatrix}, \Sigma\right),$$

where θ, α^p_1, α^p_2, α^s_1, α^s_2, γ and ρ are parameters that represent the population-level mean for θ_i, α^p_{i1}, α^p_{i2}, α^s_{i1}, α^s_{i2}, γ_i and ρ_i, and Σ is the population-level covariance matrix. We assume independent, diffuse normal priors on the population parameters: $P(\theta, \alpha^p_1, \alpha^p_2, \alpha^s_1, \alpha^s_2, \gamma, \rho_i) = P(\theta)P(\alpha^p_1)P(\alpha^p_2)P(\alpha^s_1)P(\alpha^s_2)P(\gamma)P(\rho)$ and a diffuse Inverse-Wishart prior on Σ.

θ_i, α^p_{i1}, α^p_{i2},α^s_{i1}, α^s_{i2}, γ_i and ρ_i are identified from the individual-level variations across time in storage (x_{it}), distance to quota ($Q_{it} - a^p_{it} - a^s_{it} - x_{it}$), person and social deletion activity

(p_{it}/s_{it}), and referral activity (r_{it}) in our panel data. Our panel data has a considerable length in consumer activity, as we observe individual-level consumer behavior over a four-year period. Even aggregated at the weekly level, all consumers have at least 93 weeks of data.

3.5 Estimation

The structural parameters $\Theta = (\theta, \alpha_1^p, \alpha_2^p, \alpha_1^s, \alpha_2^s, \gamma, \rho)$ represent the benefit to storage, deletion cost, price coefficient, and the referral cost. Our model and setting present several challenges in estimation: a) large state space, b) discrete-continuous decisions, and c) jagged likelihood. We considered several possible estimation approaches; we explain why we decided to use the Bayesian Imai-Jain-Ching (Imai et al., 2009) method (IJC), and we discuss how IJC alleviates the aforementioned challenges.

We begin with a choice from two common classes of estimation approaches: iteration-based methods in the tradition of Rust (1987) or simulation based two-step methods that follow the tradition of Hotz and Miller (1993); Hotz et al. (1994). The advantage of the first method is that we obtain an estimate of the value function at the end of the estimation process, but this comes at a higher computational cost than the simulation-based methods. While the simulation-based methods are computationally light, such as BBL (Bajari et al., 2007) and POB (Pakes et al., 2007), their accuracy heavily depends on being able to correctly recover the primitives of the agent's policy function in the first step, as any errors in the first step will propagate into the second step and potentially become magnified through the simulation process.

The fundamental idea of the iteration-based estimators is to nest a fixed-point iteration step within the maximization step of MLE. First, one solves the value function of the consumer dynamic programming problem via a fixed-point iteration of the Bellman equation for a given parameter guess. The solution to the fixed-point is a contraction-mapping and therefore, under regularity conditions, we are guaranteed to find a unique solution to the value function. In the second step of the procedure, conditional on solving the value function, the problem is a traditional maximum-likelihood estimation problem, and one can proceed using tradition optimization routines to obtain a consistent estimate of the structural parameters. The algorithm iterates through these two steps for every guess of the parameter value. This procedure, referred to as the Nested Fixed Point estimator (NFXP), is the work horse in estimating many dynamic discrete-choice structural models. Rust (1987) provides proofs of convergence and the properties of the estimator.

Our setting presents a few immediate estimation challenges. First, the NFXP estimator is computationally demanding because it fully solves the Bellman equation at every guess of the parameter value. In addition, the fixed-point iteration must be solved across all states, and therefore the computational time for each iteration of the NFXP increases as the size of the state space grows. The IJC algorithm alleviates this computational challenge. It does so by 1) evaluating the fixed-point iteration once per guess of the parameters and stores a collection of these values and 2) approximating the value function by using a history of past stored value functions weighted by kernels. Another state-of-the-art estimation algorithm one can use is the Mathematical Program with Equilibrium Constraints (MPEC) algorithm (Su and Judd, 2012). While MPEC also has a lighter computational burden of estimating dynamic discrete choice model over the traditional NFXP, it requires user-supplied specifications of the constraint sparsity patterns in order to fully take advantage of the speed gains. Since we wanted an approach that offers the computational advantages as well as the benefits of Bayesian estimation, we opted for IJC. One of these benefits is the ability to account for individual-level unobserved heterogeneity. This is another advantage over BBL, where we would have been restricted to a limited heterogeneity specification.

Furthermore, at each evaluation of the fixed-point iteration, we reduce the number of states, that we have to evaluate the value function over, by using shape-preserving splines. The spline approximates the value function by only having to evaluate the value function over a small subset of states (knots), and then "fills-in" the rest of the value function over all states. Lastly, we gain an additional computational saving by avoiding the calculation of the numerical integral over the entire support of ν^p and ν^s. To do this, we use two independent Gaussian quadratures to approximate the integrals over a subset of the support over ν^p and ν^s. This idea is similar to that of using splines, where the Gaussian quadrature makes a polynomial approximation of the function over a small number of nodes, a subset of the entire support of ν^p and ν^s.

Another challenge to tackle is the discrete-continuous choice aspect of the model. IJC, like NFXP, is designed to estimate dynamic models with discrete-choice controls. In order to handle the continuous-choice control in our problem, we combine the IJC algorithm with a likelihood modification derived from the Euler equation in the spirit of the continuous-choice dynamic models described in the macroeconomics literature. We obtain log-likelihood values of the continuous-choice shocks ν^p and ν^s in a grid-inversion fashion as Timmins (2002). We discuss the details of the grid-inversion in the likelihood specification section below.

Lastly, a consequence of using the grid inversion technique is that the likelihood can be jagged and multi-modal. Therefore, not only does the likelihood not have a analytic derivative, but the jagged likelihood can cause traditional gradient-based optimization methods to become fixed at local maxima. Therefore, in practice, the model from Timmins (2002) is best estimated using comparison methods with multiple starting values. Once the optimizer is in the locality of a globally optimal region, only then can one trust gradient-based methods to find the global optimum. This can be computationally demanding and can take a bit of a coordination effort in order to ensure one finds the global optimum. In addition, if the number of parameters of one's discrete-continuous model is high, then estimating this model using traditional gradient methods would be practically infeasible. The IJC method can handle this challenge since Markov Chain Monte Carlo's (MCMC) stochastic optimization nature is robust to complex likelihood shapes that are highly non-monotonic and can handle parameter space with high dimensionality Imai et al. (2009). We have verified this in our own context with extensive Monte Carlo simulations, and find that global optimum is achieved regardless how jagged the likelihood is.

To tackle all of the estimation challenges, we use a modified version of the Bayesian Imai-Jain-Ching (IJC) algorithm, in conjunction with several state-of-the-art numerical computation techniques, such as Gaussian quadratures and splines, to estimate a discrete-continuous choices dynamic structural model in a Bayesian fashion. The IJC algorithm is a variant of MCMC. It builds upon MCMC methods, based on full likelihood estimation, in that it uses Gaussian kernels and a stored histories of stored pseudo-value functions to approximate the true value function. It provides the benefits of MCMC while alleviating the heavy computational burden of a estimating full-solution Bayesian dynamic discrete choice model with forward looking agents that requires solving the Bellman equation at each MCMC iteration.

The IJC algorithm is a modified version of MCMC, and it follows these four steps at every MCMC iteration k:

1. Draw proposed parameter values, Θ^{*k}.

2. Evaluates pseudo-Expected Value Functions (pseudo-EVF) at currently proposed parameters and the last accepted parameters, $E\tilde{W}(D,\cdot,\Theta^{*k})$,$E\tilde{W}(D,\cdot,\Theta^{*k-1})$. These pseudo-EVF's are approximations to the Expected Value Function in Equation 3.4, and they are constructed using previously stored pseudo-Value Functions, from $H^k = \{\Theta^{*l}, \tilde{W}(\cdot,\cdot,\Theta^{*l})\}_{l=1}^{l=k-1}$, via kernel methods. This is the key innovation of IJC.

3. Calculates pseudo-likelihood values at the currently proposed parameters and the last accepted parameters, $\tilde{L}(\Theta^{*k}, E\tilde{W}(\cdot,\cdot,\Theta^{*k}))$ and $\tilde{L}(\Theta^{*k-1}, E\tilde{W}(\cdot,\cdot,\Theta^{*k-1}))$, using the pseudo-EVF's calculated in the previous step. These likelihood values are used in a

traditional Metropolis-Hastings step, to decide whether to accept or reject Θ^{*k}. Since the prior and the pseudo-likelihood are not conjugate, we cannot obtain a closed-form distribution on the posterior, and therefore we cannot use a Gibbs sampler.

4. Create a new pseudo-Value Function $\tilde{W}(\cdot, \cdot, \Theta^{*k})$ by evaluating the Bellman operator on Equation 3.3. This is then added to the history of past proposal parameters and pseudo-Value Functions, H^k. In the specific context of a dynamic discrete-choice problem from Ching et al. (2012), the pseudo-Value Function is referred to as the pseudo-Emax function.

3.5.0.1 Likelihood Specification

Now we explain the formation of the likelihood specification. With the standard conditional independence assumption, the individual likelihood for the model can be specified as:

$$
\begin{aligned}
L_i(\Theta) &= \prod_{t=1}^{T} P(y_t, r_t, p_t, s_t | x_t, z_t, R_t, R_t^a; \Theta) \\
&= \prod_{t=1}^{T} P(y_t, r_t | x_t, z_t, R_t, R_t^a; \Theta) P(p_t | x_t, z_t, R_t, R_t^a; \Theta) P(s_t | p_t, x_t, z_t, R_t, R_t^a; \Theta),
\end{aligned}
$$

where $P(y_t, r_t | x_t, z_t, R_t, R_t^a, a_t; \Theta)$ is the likelihood contribution from the consumer's discrete choices: plan-choice (y) and number of referral invites (r). We are able to factor the joint likelihood $P(y_t, r_t, p_t, s_t | x_t, z_t, R_t, R_t^a; \Theta)$ into the products of $P(y_t, r_t | x_t, z_t, R_t, R_t^a; \Theta)$ and $P(p_t, s_t | x_t, z_t, R_t, R_t^a; \Theta)$ due to the timing assumption: consumers make the y and r decisions simultaneously, before the decisions p and s. Assuming $\epsilon(y, r)$ to be distributed type I extreme value, the functional form of the likelihood can be expressed as:

$$P(y_t, r_t | x_t, z_t, R_t, R^a_t, a_t; \Theta) = \left[\frac{exp(V_{jk}(x_t, z_{t,}, R_t, R^a_t; \Theta))}{\sum_m \sum_n exp(V_{mn}(x_t, z_t, R_t, R^a_t; \Theta)} \right]^{1[y_t=j, r_t=k]},$$

where the discrete-choice specific Value Function $V_{jk}(x, z, R, R^a; \Theta)$ is:

$$V_{jk}(x, z, R, R^a; \Theta) = \mathbf{E}_{\nu^p} \left[\max_{p \in H^p(x,z,a^p,R^a)} \mathbf{E}_{\nu^s} \left[\max_{s \in H^s(x,z,a^s,R^a)} u(\cdot; \Theta) + \beta EV(x', z', R' R^{a'}; \Theta) \right] \right].$$

Next, the $P(p_t | y_t, r_t, x_t, z_t, R_t, R^a_t, a_t; \Theta)$ is the likelihood contribution from the continuous choice: personal deletion (p). Since we have a monotonic relationship between p and ν^p, it is possible to invert values of ν^p from observed values of p through $g(\cdot)$, the first order condition from the personal deletion sub-problem. The likelihood can then be formed as:

$$P(p_t | x_t, z_t, R_t, R^a_t; \Theta) = P(\nu^p_t = g^{-1}(x_t, z_t, R_t, R^a_t) | x_t, z_t, R_t, R^a_t; \Theta) \left| \frac{\partial g^{-1}(\cdot)}{\partial p} \right|.$$

From Timmins (2002), the continuous-choice shock ν^p can be inverted from the policy function $p^* = g(x, z, R, R^a, \nu^p)$ for given values of all the other actions and state variables. Once the ν^p is recovered, it can then be evaluated at the density function of the specified distribution of $P(\nu^p | \cdot)$ in order to get the likelihood contribution. Given our sequential nature of the decisions for p and s, we can derive their policy functions separately.

For our particular specification of the utility function and the transitions, the model is part of the class of dynamic programming problems called Linear-Quadratic problems. For this class of models, one can derive an analytic solution to the continuous choice portion of the value function by using a combination of the first order conditions and the Envelope Theorem. The resulting Euler equation gives the analytic solution for p^*, the optimal amount of deletion, and it can be specified as follows:

$$p^* = \max\left(\min\left(\frac{\beta\theta - \alpha_2^p\nu^p(1-\beta)}{2(1-\beta)\alpha_1^p},\ x + a^p + a^s\right),\ x + a^p + a^s - Q(z,r)\right).$$

The min and max statements simply provide the boundary constraints, derived from the correspondence constraint, on the optimal p^* amount. The min on the $x + a^p + a^s$ simply ensures that a consumer cannot delete more than the current amount in the consumer's account, and the max on $x + a^p + a^s - Q(z,r)$ ensures that consumers are forced to delete excess files that puts the account usage over current quota. A similar process is used to derive the analytic expression for s^*, conditional on the customer deciding p^* first.

3.6 Results

3.6.1 Parameter Estimates

In this section, we present the results of our estimation. We obtained these values through 25,000 iterations of the IJC algorithm using random initial values for each individual customers. Convergence is assessed visually, and we use the last 5,000 iterations for inference. To summarize the results of the individual-level posterior distribution, we present in Table 3.3 the median and standard deviation of the population distribution of individual-level posterior means for each parameter. This median can be interpreted as the behavior of the "typical" customer in the population.

Variable	Population Median	Std. Dev.
θ_i: Benefit to Storage	0.010	0.047
α_{i1}^p: Personal Deletion Cost	-0.251	3.55
α_{i2}^p: Personal Deletion Benefit	0.270	0.502
α_{i1}^s: Social Deletion Cost	-0.832	1.022
α_{i2}^s: Social Deletion Benefit	0.413	0.533
γ_i: Price Coefficient	-3.912	16.559
ρ_i: Referral Cost	-10.543	18.518

Table 3.3: Summary of Individual-Level Bayesian IJC Estimates

All of the signs of the parameters are as expected. We now explain the intuitive implications of each parameter, with the first parameter contributing as a flow utility and the last three contributing as action utilities. We suppress the subscript i for simplicity.

First we examine the parameter θ, which is the benefit to storage. This parameter is the linear benefit to a consumer having files stored in her account folder. The positive coefficient indicates that the typical consumer, on average, receive positive flow utility for having megabytes of files stored in her folders overtime. This positive coefficient indicates that consumers get value from having files stored over time as opposed to simply adding files into the folder *temporarily* and then using the service to purely transfer files between different computers and mobile devices.

α_1^p and α_1^s denote the cost of deletion, while α_2^p and α_2^s can be interpreted as the temporary deletion benefits. The negative coefficients on the first two indicate that the typical consumers have convex cost to personal and social deletion. This means it becomes incrementally more costly for consumers to delete files as the amount of files needed to be deleted increases. In other words, consumers prefer many weeks where they delete a modest amount as opposed to a few weeks of a large amount of deletion. This type of "smoothing" behavior indicates that the firm may wish to think about ways to profit from a different storage accounting scheme where, in lieu of establishing a quota for the total amount of storage per month, the firm can adjust an upload/download bandwidth scheme. Other existing freemium companies such as Evernote use such an approach. We do note that in our estimation, we allow all four parameters to freely vary between positive and negative support, allowing us to capture the flexible deletion behavior for each customer, while the typical customer have a convex cost for deletion, there are customers who have positive means estimates for both parameters, indicating that deletion is not a costly activity for them, relative to the other customers in the population.

Next, γ is the price coefficient. This parameter denotes how price sensitive the typical consumers may be. The negative value of this estimate is as expected and indicates the magnitude of the costs that consumers must bear when upgrading to a premium plan. Lastly, ρ is the cost of referral for consumers. This cost could be attributed to the cognitive, social, and other costs of actually sending out invitations to friends. The quadratic nature of this term reflects the fact that, as each consumer sends more and more referrals per week, it becomes harder to think of more friends who do not already have invitations. The benefit for each referral is accounted by an expectation of referral bonus quota included in the correspondence constraints ($H(\cdot)$) in the dynamic problem of Equation 3.4. The quota increases according to a specified referral incentive amount (250 MB).

3.6.2 Counterfactuals

In this section, we present the results of the counterfactual simulations generated from the estimated parameters. The goal of these "what-if" analyses is to see the effects of changing key design variables on consumer upgrade and usage behavior. These design parameters include changing the price charged for the premium plans, the size of free quota, and the magnitude of the referral incentives.

3.6.2.1 Price Changes

We now conduct various what-if simulations to entertain the changes in upgrade subscriptions when we change various design parameters. First, we begin with changes in price of premium plans. We explore what would happen to the percent change in upgrade rate if we were to increase and decrease the price by 50%. The results of the simulations are presented in Table 3.4.

Price Per Month	**Change in Upgrade Rate**
\$14.99	Drops to 0.
\$9.99	(Observed Price)
\$4.99	+2.13X

Table 3.4: Change of Price on Total Upgrade Rate

As expected, increase in price decreases the average weekly upgrade rates for the entire consumer base. On the other hand, a decrease in price yields an increase in the average upgrade rates. The default average weekly upgrade rate is 6%. We find that if the price were to be increased to \$14.99, we see a dramatic drop in upgrade rate to 0%, suggesting, not surprisingly, customers' reluctance to price increases. Furthermore, halfing the monthly price to \$4.99 also more than doubles the upgrade rate. However, running the counterfactual over a broader range of prices leads to another observation. One can see from Figure 3.6 that the change in upgrade rate is nonlinear. This nonlinearity indicates that consumers are resistant to changing their upgrade behavior, even if price gets reduce to be very small. For instance, while halving the monthly subscription rate to \$4.99 yields double the upgrade rate, halving it further to \$2.49 does not yield another 2X increase in upgrade rates. This suggests that the firm cannot simply rely on changing the price to yield higher revenue through increase upgrades, but rather they should also consider the effects of changing other design parameters, such as quota size and referral incentives.

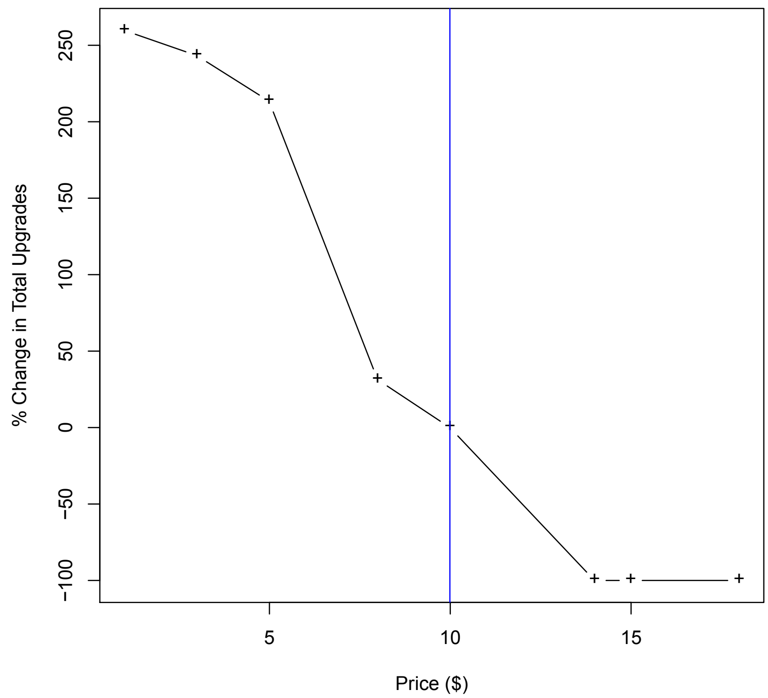

Figure 3.6: Change in Price on Upgrade Rate

3.6.2.2 Quota Changes

Another design variable of interest is the size of the free quota to be offered. While an increase in quota size will appear more attractive to customers, encouraging them to adopt the service and use the service, giving too much free quota can potentially decrease the portion of customers who upgrade to the premium plan. We explore the percentage change in total upgrade rates by increase the free quota by 1GB increments. We also run the counterfactual at 3.5 GB as a halfway point to 5GB.

We observe that as we increase the quota to 3GB and 3.5, the upgrade rate decreases by 4% and 8% as relavtive to the original upgrade rate. However, once we double the free quota, we see the a more dramatic drop of 42%. Furthermore, once we increase the quota to 5GB, we lose almost all of the upgrades. This is shown in Figure 3.7. We speculate that the reason for this dramatic drop is that customers in general have a distribution of file additions that they need to accommodate for (they only need so much space), and if the quota is way above that threshold, then the free version is "good enough" for almost everyone.

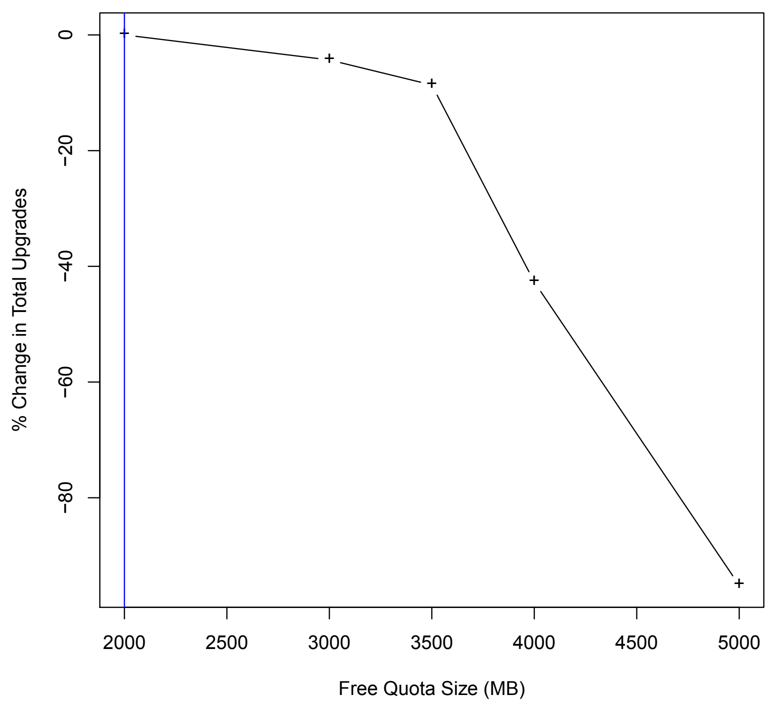

Figure 3.7: Change in Referral Incentives on Percent Change in Average Referrals Sent (MB/Referral Accepted)

3.6.2.3 Referral Changes

Most firms are interested in the freemium business model since, when paired with referral incentives, it has the potential to help the firm grow its consumer base rapidly. Since the most important stated goal for any early stage company is to gain traction in obtaining a large user base, it is of interest for firms to understand how to make this process more effective. First, we explore what would happen if we were to generate counterfactuals for the median customer in the population. We notice that since the personal and social deletion costs for

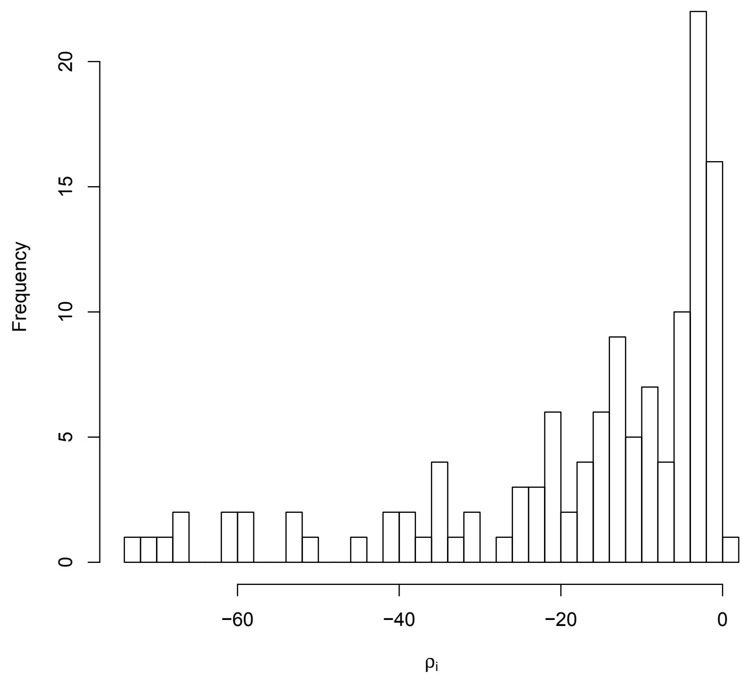

Figure 3.8: Distribution of Referral Cost Posterior Means

these customers are not as high as the referral and upgrade costs, the median customer would resort to deletion more than referrals, therefore we find that it is very difficult to incentivize the median customer to send out any referrals, needing to increase the referral incentive to as high as 1GB per referral accepted.

However, once we segment the customers' individual level coefficients according to how likely they are to send out referrals, we can then identify the top referrers and then counterfactuals where we change the referral incentives to see how these top referrer's referral behaviors would change. Figure 3.8 shows the distribution of the means of the posterior distribution estimates of the individual-level referral costs (ρ_i). From this, we can pick out the top ten customers who are to the right, and we can "create" a representative customer of this selected group by averaging the means of their individual-level posterior means for each parameter. When we identify the top ten referrers, we can then change the referral incentives to see how this group would respond. We explore the results in the next section.

Targeting the High Referrers In this section, we explore what would happen if we were to change the referral incentives offered for each accepted referral invite. Specifically, we explore the effect on the high-referrer's referral activity over a range of referral incentives. The default amount that is given to each referral during our observation is 250 MB per invite accepted. We vary the incentive across a wide range, from as little as 50 MB per referral to as large as 1 GB per referral. We create a representative "high referrer" by taking the average of the posterior means for the ten customers having the lowest mean estimates for referral cost parameter. The result of this is shown in Figure 3.9.

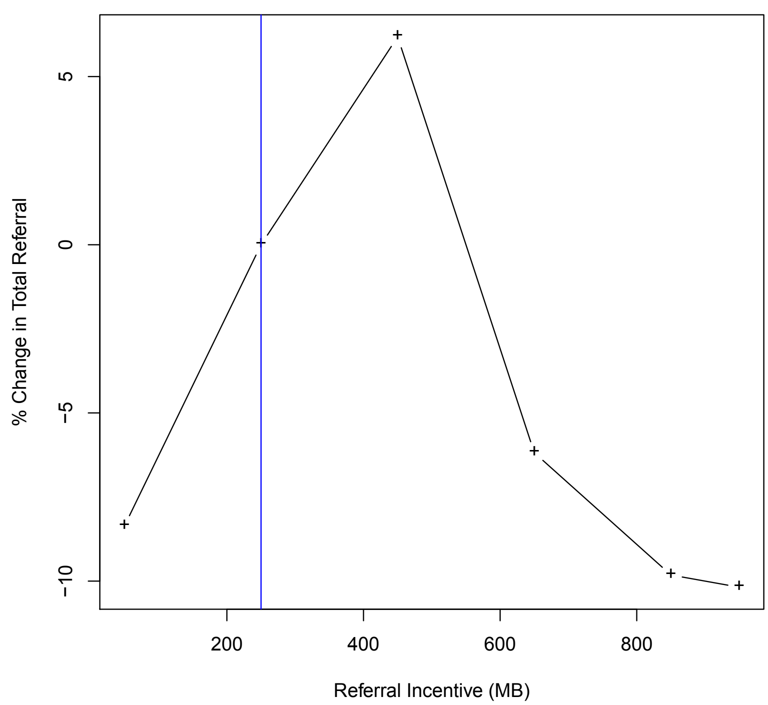

Figure 3.9: Change in Referral Incentives on Percent Change in Total Referrals

First, as we increase referral incentives, we see an increase in the number of referrals sent, and after 450MB of incentive, the average number of referrals sent actually decreases. This indicates that the maximal amount of referral incentive lies around 500 MB if our goal is to increase the average number of referrals sent[10]. If the firm gives too much space for the referral incentive, they are not encouraging more, but rather fewer referrals. We speculate this is because if a consumer can gain enough free space with only one referral, why go through all the trouble of inviting more? In addition, the increase in referrals also comes at the cost of decreasing upgrades. As consumers get more free space from referrals, they are less likely to upgrade, as indicated by Figure 3.10.

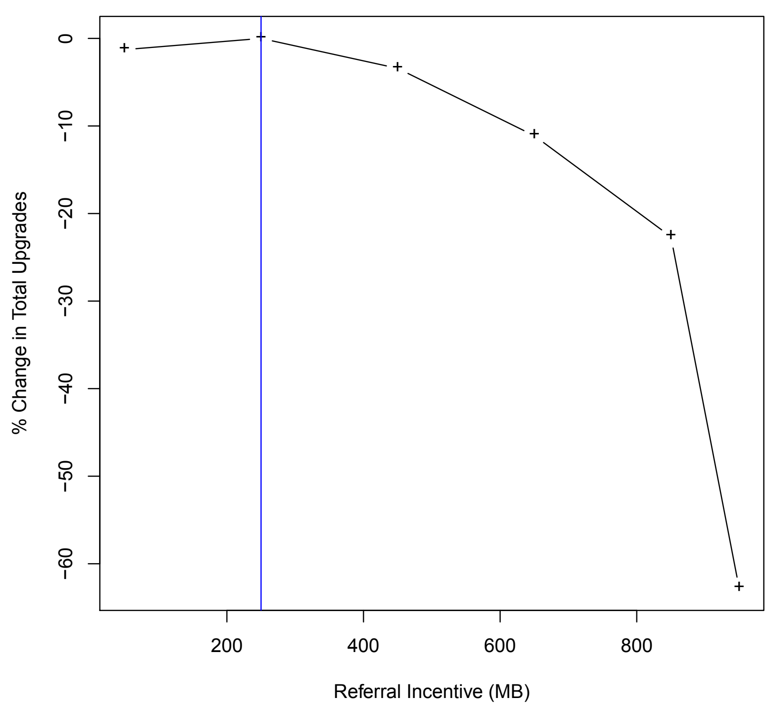

Figure 3.10: Change in Referral Incentives on Percent Change in Upgrade Rate

[10]We observe a similar result in a homogenous model where only the personal usage, upgrade, and referral behaviors are characterized

3.6.2.4 Impact of Referrals on Value of a Customer

In this section, we run a counterfactual to gain a better understanding of the value of a free consumer. More specifically, we find the value of the free consumer from their referral effects. In order to calculate this, we generate 1,000 identical high-referral consumers, each starting out by choosing the free plan, and then we simulate their behavior for six months. Out of these 1000 customers, we let each refer customers, and also make social/personal addition and deletion, and then at the end of the six months, we count up the total number of upgraded customers and assume they pay for just one month of subscription of $9.99. The total revenue divide by the original 1000 customer should give us a conservative estimate of the the value of each of these original 1000 customers.

In addition, we conduct a counterfactual where the referral program or the social features do not exist.[11] The goal of this is to get a sense of the value of each feature. By comparing the number of consumers who upgrade amongst all three settings, we get a sense of the proportion of the value of having a referral feature and the no social feature. [12]

[11]We operationalize the no-referral scenario by: 1) making referral cost parameter ρ take on a large negative magnitude, 2) making the referral bonus quota per referral to be 0, and 3) setting the $R^{max} = 0$. For the no-social scenario, we: 1) turn off all social addition shocks, 2) make social deletion cost parameter α_1^s take on a large negative value, and 3) set social deletion benefit parameter $\alpha_2^s = 0$.

[12]We simulate the baseline setting by using the estimated primitives and setting all incentives to mirror the conditions in the actual data. We do this instead of subtracting the counterfactual from the actual data in order to minimize errors from simulation and model fit. In an additional effort to minimize simulation error, we use the the same sequence of draws for a^p,a^s, ν^p,ν^s, ϵ for the baseline, No-referral, and the No-Social counterfactual. To calculate the baseline counterfactual, we do the following three steps:

1. Start 1,000 baseline consumers on the free plan, then simulate for 115 periods for each consumer.
2. For each period, we sum up the total number of referrals accepted.
3. For each incremental referral accepted each week, we simulate an additional consumer for the remainder of the periods.

	Baseline Scenario	No-Referral Scenario	No-Social Scenario
Total Number of Premium Consumers	279	52	71
Organic Premium Consumers	99	52	20
Referred Premium Consumers	180	0	51
Avg. Pers. Deletion Per Organic Consumer	216 MB	360MB	183 MB
Avg. Soc. Deletion Per Organic Consumer	19 MB	31 MB	0 MB
Avg. Storage Organic Consumer (Free Region)	508 MB	812.8 MB	126 MB

Table 3.5: Comparison of Referral and No-Referral Program Counterfactuals

We now examine the results of this counterfactual in Table 3.5. There are a few interesting observations. First, the amount of referred premium customers in the baseline scenario alone is almost four times the number of premium consumers in the No-Referral scenario. This is interesting because it indicates that the value of the consumer from referrals program is quite substantial, as the referral program brings in consumers that accounts for 64% of the total number of premium consumers in the baseline scenario. We see a similar result in the No-Social scenario, having the referral program bringing almost 70% of the total premium customers.

Second, we observe that in order to compensate for the amount of space that would have been gained from the referrals, consumers in the No-Referral scenario delete more, and therefore deletion is a closer substitute to referral than upgrades are. This is as expected, since consumers need to make space, and without the means to gain space from referrals their only other option is to delete more. We see this by comparing the average amount of personal and social deletion in the original 1,000 consumers who joined by themselves (Organic), during the periods when they are in the free plan, between the Baseline scenario and No-Referral scenario. We see an increase of more than 60% in personal and social deletion in the No-Referral scenario to compensate for the lack for the referral program.

Last and perhaps most importantly, there are twice as many consumers from the Organic consumers who upgrade in the baseline condition over the No-Referral scenario, meaning that the extra space they get from the referrals actually makes them more likely to upgrade in the long run, even if they do not end up using the extra space from referrals. What is counter-intuitive about this is that typically referral programs in freemium firms are seen as an impediment to consumer upgrading to a premium version. It is viewed as a pure customer acquisition tactic via Word-of-Mouth. However, we see evidence that the presence of referral programs, if geared to drive engagement and usage, may work to increase the number of consumers choosing a premium plan. We explore the implications of changing referral incentives in the optimal referral counterfactuals found in section 3.6.2.2. In addition, we also see an even greater increase (almost five times) of Organic upgrades from No-Social scenario to the Baseline scenario. The large increase from both scenarios hint at a potential synergistic effect of having a referral program and social feature on personal usage. More counterfactuals can be conducted to decompose the value of each of the components and its associated spill-over effects. Synthesizing these observations, we estimate the value of a free consumer *per month* to be an average of almost \$3 (279*\$9.99/1000 Consumers).

3.7 Discussion and Limitations

In this study, we present a dynamic structural model of consumer upgrade, usage, and referral behaviors as a framework to develop a deeper understanding of the freemium business model. We hypothesize that the value of consumers come from four sources: 1) consumers eventually convert to premium over time, 2) consumers bring in others via the referral program who then convert to premium consumers, 3) consumers convert to premium product because of increasing usage of social features, and 4) synergistic effects among the three components. We find that the value of a free consumer to be \$3 per month. In addition, we find that having a referral program is quite significant, as it can account for at least 64% of the total premium customers, even without a social feature existing. We look further into the impact of changing design variables, such as price, quota and referral incentives with additional counterfactual simulations. We discover that while changing referral incentives for the median customer may not be effective in increasing referral behavior, changing referral incentive is effective if targeting the high referrers. Furthermore, we found the existence of a static optimal incentive for referral is around double the 250 MB amount. In addition to the substantive finding, we provide a way to account for the social value of consumers via counterfactual simulations that account for both the use of social features and referral programs. Lastly, we also incorporate the dynamic structural model literature by demonstrating a way to incorporate both discrete and continuous actions into an integrated model that allows for the recovery of the value function.

Our findings can inform managers in several ways. First, we confirm the critical importance of the referral program in its contribution not only to the growth of user base (via more referrals), but also to the dynamic lifetime value of consumers, which help to accurately assess the value of the firm. Secondly, the referral incentive is a viable managerial control to experiment with and increase according to our static optimal counterfactual.

A limitation of this work is that we currently do not model consumers switching to the outside option, so therefore we can only assume the results hold for a firm acting as a monopoly with a captive user base. Since the acquisition of this data set, the industry has become an oligopoly, so we need to treat the pricing results with some caution. Future work may incorporate the pricing choices of competitors in order to arrive at a better pricing strategy.

Chapter 4

Computational Challenges and High Performance Computing Solutions

In this section, I discuss the computational challenges with estimating the models that are found in the previous essays and some of the approaches that I have tried to remedy the problems.

4.1 Estimation Challenges of the HMM

4.1.1 Computational Speed

Similar to the log-likelihood of a mixture model, the log-likelihood of the HMM consists of latent part-worths that must be inferred at each guess of the parameter. These part-worths are the transition probabilities of the latent states. Unlike a mixture model, the HMM's part-worths are allow to evolve over time, which makes the HMM estimation problem a rather computationally demanding one. Since the log-likelihood evaluation is unavoidable under a likelihood-based approach, whether *one* employs a frequentist (e.g. EM algorithm) or a Bayesian approach (e.g. MCMC), this is often the first place to start to identify performance bottlenecks.

Originally, I had implemented my entire Bayesian estimation routine in R. However, R is notoriously slow for running loops. Hence, aside from vectorization of one's code, the two standard ways to speed up R are either to a) rewrite R code in a compiled language (e.g. FORTRAN, Java, or C++) or to b) use parallelization. I opted for rewriting the log-likelihood function in C++ since the conversion can realize performance gains of two or three orders of magnitude (Eddelbuettel and François, 2011). For thoughts on parallelization, see section 4.2, where C++ conversion was not sufficient for enough performance gains. Next, I describe the approach I have taken to speed up my HMM estimation algorithm as well as some of the tools I needed for this process.

4.1.2 Using Compiled Code

At a high level, whenever one wishes to rewrite functions in compile code, there are two immediate choices one needs to make. The first is to decide at what level of abstraction will one stop the code conversion. Should one convert all of the R code into compiled code, and call the entire estimation routine using a compiled executable file? Or should one just convert a portion of the R code into compile code, and call the compiled object using a R wrapper function? Generally, I find that it takes less development time to do the latter, since one can use other packages to facilitate the interface between the two languages. The cost however, is that it is very easy to misuse the combination of packages that serve as the interface between the two languages, and one may end up with memory leaks that can crash the entire estimation routine. Ideally, one could prototype everything in R code, and then transfer everything into compiled code, and run the entire estimation routine using the compiled binary. The advantage to this route is that one gains access to the full suite of debugging tools that are designed for development within the compiled code environment.

4.1.2.1 Rcpp: an Interface Between R and C++

Since I decided to only convert the log-likelihood function, I first need to decide what interface I wish to want to use. R provides a standard interface to call C and FORTRAN, and at the time of writing, a popular way to call C++ is using the *Rcpp* package(Eddelbuettel and François, 2011; Eddelbuettel, 2013). The advantages to using *Rcpp* are a) the wide support that it has in the R community and b) the ease of programming transition from R syntax to C++/Rcpp syntax. The contributors to the project also provide other related packages, such as RcppSugar, which provides syntactic sugar that allows users to write C++ functions in a similar fashion as vectorized R code, and *RcppArmadillo,* which provides an interface between Rcpp objects in C++ and the C++ Armadillo Matrix library.

Before one can proceed, one will need the proper C++ compiler installed on the local machine. For Windows machines, one needs to obtain RTools[1], which is a collection of resources that gives the Window user the ability to build R packages. For Linux user, there should be a C++ compiler installed already, and for Mac OS X users, the app Xcode provides a suitable compiler.

Once the compiler is obtained, there are two ways to use *Rcpp*: 1) embedding C++ code directly in R code by using *cppFunction()* function in R, or 2) building an R package with the C++ files included. The *Rcpp* provides a *cppFunction()* that allows users to directly write C++ code within the body of the R code. This function automatically takes care of the compiling, linking and execution, all within the R interpreter during run-time. The purpose of this function is to make prototyping and error-checking individual functions easier. The drawback to this approach is that it cannot call other user defined C++ functions within R. Since I needed to define sub-functions within the log-likelihood function calls, I found that it was necessarily to build a R package with my C++ code.

[1]http://cran.r-project.org/bin/windows/Rtools/

To build the R package, one simply do the following steps:

1. From R, run the following code. This creates a directory named *my_package* in the current directory. This is where one's C++ and R files should reside. All of the other necessary configuration files should be automatically be created. The *Rcpp* or *RcppArmadillo* code choice below depends on whether one wishes to use the Armadillo matrix library. I discuss Armadillo in the next section.

   ```
   R> Rcpp.package.skeleton("my_package")
   ```

   or

   ```
   R> RcppArmadillo.package.skeleton("my_package")
   ```

2. In the "src" folder, create a C++ file. This is where the converted C++ functions reside.

3. In the "R" folder, create a R file. This is where all of the R functions that calls the C++ functions reside. Once the package is loaded in R, these are wrapper functions that a user can call from R to access the C++ functions. An example of a wrapper function is listed:

   ```
   R> rcpp_hello_world <- function() {
           .Call( "rcpp_hello_world", PACKAGE = "my_package" )
   }
   ```

4. To compile, type the following in the command line (DOS or Linux/OS X terminal). Make sure one is in the same directory as the *my_package* directory.

   ```
   terminal$ R CMD CHECK my_package
   ```

   This compiles the package and returns an output of errors if the C++ code does not compile. To install the package into the local system, type from the command line:

```
terminal$ R CMD INSTALL my_package
```

5. To load the package in R, use the library() function call as with loading any packages.

6. When the package is ready for distribution, use the following command to pack the directory into a tarball. This will create the file *my_package_1.0.tar.gz*.

```
terminal$ R CMD BUILD my_package
```

7. To install on other machines, use the command:

```
R> library("my_package_1.0.tar.gz", type="source", repos=NULL)
```

4.1.2.2 Choice of Matrix Algebra Library

Whether one chooses to call C++ from R, or to implement all functions in C++, it is recommended to use a matrix algebra library. This alleviates the user from having to manually create and assign matrix values via many repeated loops. The three most popular matrix libraries are Armadillo, Eigen, and the GNU Scientific Library (GSL). While GSL is primarily for use with C, Armadillo and Eigen are popular for use with C++ and libraries such as *RcppEigen* and *RcppArmadillo* are useful for converting objects between *Rcpp* and the corresponding matrix library. I chose Armadillo due to the ease of use and similar syntax to MATLAB/R. The performance for both libraries are similar, although there are cases in which one is faster than the other, all depending on one's specific problem. There is not a clear winner on performance comparisons between Armadillo and Eigen [2].

4.1.2.3 Memory Leaks and Debugging Tools

The most common problem I ran into while using *Rcpp* is memory leaks that would cause my program to crash. This is where the choice of using R/C++ or pure C++ matters. If one

[2] Rcpp Developers List: http://thread.gmane.org/gmane.comp.lang.r.rcpp/3522.

were to convert all code to C++, debugging is much easier than the R/C++ combination choice. There are standard tools such as Valgrind[3] to detect memory leaks down to the line number of the code. On top of that, tools such as gPerfTools[4] can help profile the code, detecting performance bottlenecks on one's various functions. I chose the route of using the R/C++ combination, and I found that setting up Valgrind and gPerfTools was fairly smooth on OS X, with the help of package managers such as HomeBrew[5] or Rudix[6]. Note that all of these tools are design natively for use on Linux, so to decrease troubleshooting time I encourage users to develop the C++ code on a Linux environment. The support for these debugging tools are minimal on Windows, so users will have to resort to simple print statements and line-by-line checking. Using the R/C++ combination, Valgrind was not able to pinpoint my memory leaks down to the line number, therefore I still had to resort to print statements and careful line-by-line checking.

4.1.3 Other Computational Alternatives

4.1.3.1 Staying within R

BLAS Optimization R, like many other high-level computational languages, relies on Basic Linear Algebra Subroutine (BLAS) to do many of its vector and matrix operations. One way to improve the numerical performance of R is to manually compile a version of R on one's machine. When doing so, one has the option to use a customized version of BLAS tuned specifically to one's machine's CPU. Since BLAS is a standard for scientific computing, there are many variants of BLAS available, such as OpenBLAS, GotoBLAS, Intel's MKL and Apple's Accelerate. There is also the Automatically Tuned Linear Algebra Software (ATLAS), which is an implementation that automatically tunes to one's machine

[3] http://valgrind.org/
[4] https://code.google.com/p/gperftools/
[5] http://brew.sh/
[6] http://rudix.org/

specification. When choosing a particular BLAS implementation, the key fact to look for is one's CPU microarchitecture. There are extensive instructions on how to compile R with tuned BLAS libraries in the Linear Algebra section of the R Installation and Administration Guide.

The R Compiler Package Recent development in Just-in-Time (JIT) byte compiler also gives users a painless way to achieve two to three times of performance gain. Starting with R 2.13, the compiler package will be distributed as part of the standard R package. To use the package, simply do:

```
R> library("compiler")
R> enableJIT(3)
```

Any functions executed after these statements will be byte-compiled after the first execution. One should see speed improvements the second time these functions are executed. Except during times when I'm debugging, I almost always turn this feature on.

Other Versions of R There are versions of R that are tuned specifically for performance gains or stability. There are commercial alternatives such as Revolution R, or other open source alternatives such as pqR (a pretty quick version of R) and Renjin (uses the Java Virtual Machine). More options are described by Wickham (2014). The typical tradeoff regarding these versions is in terms of write code for older versions of R in exchange for speed. In addition, unless one is using a commercial alternative such as Revolution, then the community of users in which one can get help from is much smaller than the global R user community. If one's code does not have too many package dependencies, then it may be fairly painless to use one of these versions of R to obtain the performance gains.

4.1.3.2 MATLAB

A common choice of computational language of economists and engineers, MATLAB's computational speed is often on par with R's. Common advantages of using MATLAB would be the supported API's to commercial optimization software such as KNITRO and AMPL. For a certain type of estimation algorithms, such as MPEC (Su and Judd, 2012), this is the recommended approach for implementation. The drawback to MATLAB is that it does not have as large of available packages for statistical sampling as R, since R is a more popular choice of languages for statisticians.

4.1.3.3 Python

Python is another common language of choice in the engineering community, and is usually faster than MATLAB or R. It is a mature language that has much support in the machine learning and computer science community. If one needs to do much machine learning work as well as text processing, along with moderate estimation tasks, Python may be a reasonable choice. For scientific computation, popular packages that are often use are NumPy and SciPy.

4.1.3.4 Julia

The desire for a high-level language is that accessible like R and MATLAB but can compare to the speed of compiled code spurred the invention of new programming languages. One such development project is called Julia. The advantage of Julia is that it entails the ease and flexibility of a high-level language syntax, while touting computational performance on the level of C(Bezanson et al., 2012). It is natively designed for parallelism and Cloud computing, so in the future this may get some traction. However, the language is very new and currently does not enjoy a widespread user base such as R.

4.2 Estimation of the Single Agent Dynamic Structural Model

4.2.1 Estimation Challenges

4.2.1.1 Jagged, Multi-Modal Log-Likelihood

Building upon the ideas described in section3.5, we now revisit the problem with dynamic models with both discrete and continuous choices. Using the grid inversion technique from Timmins (2002), the resulting log-likelihood will be jagged and multi-modal. Therefore, using standard maximum likelihood techniques can be unreliable, since these routines rely on non-stochastic optimizers that may get stuck in local maxima. In addition, this technique may require multiple starting values and numerous grid comparison evaluations, and this would render any problems with high dimensional parameter space to be unfeasible to estimate.

While techniques such as MPEC may hold promise, especially since one could specify the continuous control constraints naturally into the constraints of the optimization problem, in practice this route still require the multiple starting values and coordination of grid comparisons due to the jagged log-likelihood. This route would be recommended if one were able to derive the analytic gradients, hessians, and Jacobian of the constrained optimization problem for one's specific setting.

4.2.1.2 Continuous Shocks and State Variable

Another challenge that is the existence of multiple observed shocks that the consumer observes prior to making decisions. This involves solving multidimensional integrals in order to obtain the Expected Value Function, as in equation 3.4. Not only is calculating these integrals computationally intensive, but also calculating them non-parametrically would preclude the use of quadrature techniques. In addition, the state variable storage x is a continuous variable. Under the estimation technique of NFXP, one would need to discretize this state variable into a discrete variable, as is done in Rust (1987). The problem with this is that since the transition of x depends on the observed addition shocks and the continuous control variables personal and social deletion. Therefore, one would need to discretize all of these variables. The problem with discretizing both continuous state variables is that, while reducing the mixed discrete-continuous choice problem into a discrete choice problem, the control space is greatly expanded.

4.2.2 Solutions

Given the challenges of a model with multiple shocks, a continuous state variable, and continuous unobserved shocks, I combine several techniques to make the estimation of the structural model feasible.

4.2.2.1 The Imai-Jain-Ching (IJC) Estimator

The first technique is the use of the IJC estimator. IJC is typically used to estimate Bayesian dynamic discrete choice models. The basic idea of the algorithm is to use a Gaussian Kernel to approximate the Expected Value Function, using series of pseudo-Value Functions evaluated at various draws of the parameter value. In a traditional Bayesian dynamic discrete choice model estimator, one would have to iterate the Bellman equation numerous times until convergence, at each draw of the parameter vector. IJC recognizes that this traditional estimator calculates the Bellman equation many times per MCMC iteration, but throws all these "old" calculated values away, with every new draw of the parameter. Hence, IJC proposes to evaluate the Bellman equation once, at each MCMC iteration, and stores away a history of past evaluated Bellman equations. These paste evaluated Bellman equations are what the authors call the pseudo-Value Functions, and the authors show that using the Gaussian Kernels to generate the weights in a weighted sum of the history of pseudo-Value

Functions can converge to the true Expected Value Function as the MCMC converges to the true parameter. Aside from this computational advantage, IJC is subjected to the same costs of running MCMC estimation, as well as the benefits to Bayesian estimation. Some of these benefits include the ability to incorporate observed covariates into the heterogeneity specification, having parameter posterior estimates for individual customers, and the ability to avoid local maxima. The stochastic optimization nature of MCMC is very handy in my context, since the log-likelihood could be jagged and multi-modal. The MCMC procedure would not get stuck in local modes, and over time it should converge to the global maximum.

In addition to these benefits, the IJC estimator also have additional properties that makes it attractive for my context. As mentioned in the previous section, I have to deal with multiple continuous observed shocks (from personal and social addition). Instead of evaluating the multiple integrals at fixed n grids, IJC has a natural way of using Monte Carlo integration to evaluate the multiple integrations. This is very simple to implement, as all one needs is to take k draws of the personal and social additions from their respective distributions, at each MCMC iteration. Imai et al. (2009) show that k can be far less than n and the algorithm will still converge, hence creating another source of computational saving. In addition, one can use the observed distributions of these shocks and sample directly from those distributions, and this will take a nonparametric integral over both shocks. Another benefit to this is that it is not difficult to account for serial correlation in the shocks, by conditioning the distribution of addition shocks on the previous period shocks. Given these benefits, I also used two Gaussian Quadratures to approximate the double integrals over the unobserved personal and social deletion shocks ν^p and ν^s to speed up the numerical integration. On top of this, IJC can also handle continuous state variables, as well as state variables with deterministic transition, both which the NFXP estimator cannot do. Considering that the x state variable is both continuous and transitions deterministically, IJC is a natural choice for estimation.

4.2.2.2 Compiled Code and Parallelization

To speed up the calculation of the log-likelihood, I recoded the functions in C++ and *Armadillo*, and I called the compiled code from R using *Rcpp.* While I gain a speed-up of a factor of 8x, I noticed that I could use parallelization to speed up the process even more. This is crucial especially as I need to do the estimation with heterogeneity incorporated, which will need to run the MCMC over each individual's data.

In order to parallelize the code, the first decision is to choose at what level should one do the parallelization. By parallelization, I refer to the commonly used type of parallelization known as the "embarrassing parallelism," which are the easiest type of programs to parallelize. The basic idea is to look for loops where each individual iteration does not depend on values at other iterations.

There are a few possible candidates for the parallelization entry points, from the most basic level to the highest level:

- Basic matrix algebra operations.
- Evaluation of the Bellman operator over all states.
- Individual Log-likelihood.
- MCMC level.

When choosing which level to parallelize the code, the rule-of-thumb is to make sure that each individual computing blocks that is sent out to each of the other computing nodes will take longer to compute than the coordination time between the processors. Otherwise, the program can potentially run slower than its nonparallel counterpart. This is especially important if one were to utilize CPU's across networks of computers, where the latency on the network can be non-negligible.

The first candidate is to use BLAS libraries that can make use of multiple threads of the computer at every basic matrix algebra operation. The advantage to this is that there is no need to change any of the code, but the drawback is that unless one only plans to conduct parallelization on a single machine with many cores, the communication costs across machines can often be higher than the actual time it takes to calculate each matrix algebra operation.

The second candidate is the evaluation of the Bellman operator over all discrete state variables. At each iteration of the MCMC, IJC still requires the Bellman operator to be calculated once, over all the possible discrete states. If the bottleneck of one's problem is in the calculation of the Bellman equation due to a large number of discrete states, it may be worth to parallelize this part of the computation.

The next candidate would be to move another level up, where one can parallelize the calculation of the log-likelihood by sending out the calculation of each individual customer's log-likelihood to a CPU core. If the number of individuals is high, then it may be possible to estimate a model with a large number of customers by adding a core per customer.

Lastly, one can also parallelize at the MCMC level, where the entire MCMC chain is sent off to a particular core, and this will run multiple MCMC chains. This method takes quite a bit of care, since each MCMC chain will depend on the draw of its previous iteration. There is an algorithm called Parallel Tempering, which allows users to run N MCMC chains and uses the Metropolis sampler to decide whether to exchange values from two random chains after K iterations. This method is shown to improve convergence and may be useful if the sub-operations within the MCMC is difficult to parallelize.

My choice of parallelization depended on my choice of how much code to convert turn into C++. Since I converted the individual log-likelihood function and all of its subroutines in C++, a natural choice for me is to parallelize at the level of the individual log-likelihood. There are many packages in R that facilitate "embarrassing parallelization." I found the package *doSNOW* to be easy to use and sufficient. This package invokes the R package called Simple Network Of Workstations (*SNOW*), which uses an industry standard framework OpenMPI that connects not only various nodes on one machine, but allows multiple workstations and all associated cores to be connected together. The only thing that the user needs to do is to register the list of machines that one would like to use, and replace any *for* loop statements with corresponding *foreach()* calls. The nice thing about *SNOW* is that it allows one to specify connections using socket connections, and one could call slave machines directly from the R interpreter on the master machine using the slave machine's DNS address. Using this method, I was able to run multiple 100-core cluster jobs on the Harvard Odyssey Computing Cluster.

4.2.2.3 Using Amazon Elastic Compute Cloud (EC2)

As a comparison, I also tried the same method on the Amazon Elastic Compute Cloud (EC2). The nice thing about EC2 is that there is no monthly subscription, and one only pays for computing time that one actually uses. There are a few terminologies that are helpful when setting up one's own EC2 clusters:

1. **Compute Instances:** think of these as a single machine with a particular type of CPU in the physical world. There are numerous types of instances that one can rent, each with different pricing plans. I found that for estimation work, the best to use are the Compute Optimized Instances. Each one of these instances have the latest generation Intel Xeon processors, and vary from four to 32 cores per instance. Therefore it is fairly easy to setup a 128-core cluster with just four top-end instances, connected by SSH socket connections. I have found that the performance of a 100-core cluster on Amazon is 1.5 to twice the speed of a 100-core cluster on Harvard Odyssey Cluster.

2. **Amazon Machine Image (AMI)**: these are the snapshot of the system image of one's instance on Amazon. Once an user has set up an instance with a particular

operating system and installed all the required software (R, C++ compiler) and associated packages, one can save an AMI to be deployed to multiple instances, creating many machines with identical setup and user files. This facilitates the parallelization process since all the required packages will be available on the slave computers when the master instance calls them.

3. **Regions and Placement Groups:** when launching multiple instances make sure the instances are all launched within the same geographical region and the same placement group. This ensures that all of the launched instances are actually physically close to each other, often on the same network in Amazon's data centers. This matters because if one wishes to network many instances together, the geographical distances between instances will increase the network latency, leading to performance degradation.

One word of caution when setting up the R cluster call on the master instance is to call the slave instances by their internal Amazon DNS addresses, as opposed to the public DNS addresses. This could have pricing implications since Amazon charges by amount for external data transfers, so one should use internal addresses whenever possible when communicating from an Amazon instance to another Amazon instance.

4.3 Limitation and Future Work

4.3.1 Parallelization

One limitation to parallelizing within R is that the operating system only allows for 128 socket connections, and in practice, R is only able to communicate to 125 other cores due to this limitation on connections. Two possible routes of remedying this are: 1) to use the *doParallel* package, which uses a different connection mechanism, or 2) to conduct the parallelization in C++ using OpenMP. Embarrassing Parallelization seems to be fairly easy to implement on OpenMP, simply with the *#pragma* directive statements in front of for loops. However, only certain compilers will support OpenMP, so the best way to do this would be to compile one's C++ from a Linux environment.

4.3.2 MCMC Adaptive Samplers

One problem I have found with both the IJC algorithm and also the HMM estimation is that the Metropolis draws tend to be very correlated amongst the different parameters. I have had to block the draws of correlated parameters and have to hand-tune the proposal covariance step sizes in order to achieve good mixing and reduce auto-correlation on each series of the parameter draws. Netzer et al. (2008) recommends an adaptive Langevin Monte Carlo sampling technique from Atchade (2006), and the Riemann manifold Hamiltonian Monte Carlo sampler (Girolami and Calderhead, 2011) is another promising adaptive sampler used in conjunction with IJC by Roos (2012). The advantage to these samplers is that the steps of the MCMC will be directed at a "better direction" in the parameter space, therefore the entire MCMC will need less draws before reaching convergence. The tradeoff in these samplers is that each steps usually take longer to compute, and they require the specification of log-likelihood gradient and hessian. While for complex models, analytical forms of the gradients and hessian's may not be feasible to obtain, one possible technique to alleviate this problem may be to use an automatic differentiation library such as Griewank et al. (1996).

Appendix A

Appendix to Chapter 1

A.1 Tables and Figures

A.1.1 Web Annotation Service Dataset

Table A.1: Descriptive Statistics

Key characteristics	Average/Percentage
Overall Observations (Customer-Weeks)	25,823
Overall Maximum Number of Weeks	56
Overall Number of Customers	986
Search Customers	227
Word-of-Mouth Customers	246
Mass-Invite Customers	513
Overall Average Number of Weeks	24
Search Customers	18.4
Word-of-Mouth Customers	19.2
Mass-Invite Customers	28.8
Overall Average Personal Usage (# of Annotations)	24.61
Search Customers	27.01
Word-of-Mouth Customers	21.26
Mass-Invite Customers	25.81
Overall Average Social Usage (# of Msg's Shared)	0.79
Search Customers	0.56
Word-of-Mouth Customers	0.79
Mass-Invite Customers	0.89
Feedback When Joining	16 %
Perc. Customers Who Received Inbound Sharing	5.9 %
Professional Information - PR Professionals	173
Professional Information - Academics & Researchers	395
Professional Information - Other Professionals	418

Table A.2: Fit Measures

	Log Marginal Density	AICM	BICM	-2*DIC
2 State Bivariate Markov Chain, Bivariate Poisson HMM	**-28,420.44**	**-57,532.83**	**-58,652.1**	**-111,879.50**
2 State Univariate Markov Chain Bivariate Poisson HMM	-44,819.45	-90,561.8	-91,976.69	-179,510.59
3 State Univariate Markov Chain Bivariate Poisson HMM	-43,307.22	-86,668.73	-86,898.2	-173,250.26
2 State Latent Class Bivariate Poisson Model	-43,163.73	-86,328.97	-86,440.16	-172,653.50
3 State Latent Class Bivariate Poisson Model	-38,643.27	-77,292.56	-77,458.99	-154,491.01

*The more positive the number the better the fit.

Best model in bold.

Table A.3: In-Sample and Out-of-Sample Fit Measures

	Log Marginal Density (Hold-Out)	AICM (Hold-Out)	BICM (Hold-Out)	-2*DIC (Hold-Out)	Log Marginal Density (Calibration)
2 State Bivariate Markov Chain, Bivariate Poisson HMM	**-2,309.40**	**-5,551.99**	**-5,910.37**	**-9,061.54**	**-19,685.37**
2 State Univariate Markov Chain Bivariate Poisson HMM	-3,430.42	-7,568.95	-7,838.77	-13,547.82	-28,800.73
3 State Univariate Markov Chain Bivariate Poisson HMM	-3,159.04	-6,807.65	-7,080.49	-12,419.49	-27,743.82
2 State Latent Class Bivariate Poisson Model	-4,149.82	-9,017.55	-9,284.66	-16,370.99	-28,969.22
3 State Latent Class Bivariate Poisson Model	-4,131.85	-8,731.04	-8,997.36	-16,319.43	-26,136.47

*The more positive the number the better the fit.

Best model in bold.

Table A.4: Bivariate HMM Estimates

Parameter	Personal Usage Estimates	Social Usage Estimates
	State Dependent Usage Covariates	
β_{01} state dependent intercept	-2.213	-7.212
- WOM (Passive State)	(-2.322, -2.118)	(-8.860, -5.972)
β_{02} state dependent intercept	1.640	1.455
- WOM (Active State)	(1.621, 1.658)	(0.929, 1.917)
Search (Passive State)	0.600	-1.866
	(0.475, 0.729)	(-5.729, 1.088)
Search (Active State)	0.750	-0.116
	(0.697, 0.803)	(-1.579, 1.213)
Mass-Invite (Passive State)	-0.218	-2.870
	(-0.337, -0.095)	(-6.049, 0.334)
Mass-Invite (Active State)	0.318	0.157
	(0.268, 0.364)	(-0.932, 1.230)
	Transition Probability Covariates	
θ_{11} threshold (Passive State)	5.451	5.843
	(3.205, 6.591)	(3.823, 7.214)
θ_{21} threshold (Active State)	2.565	2.662
	(0.935, 3.961)	(-0.367, 4.141)
$Var(\theta_{11})$	1.222	1.137
	(0.696, 1.568)	(0.728, 1.437)
$Var(\theta_{21})$	0.964	1.566
	(0.627, 1.202)	(0.999, 1.977)
Inbound Sharing (Passive State)	2.700	7.229
	(1.759, 3.672)	(5.329, 8.644)
Inbound Sharing (Active State)	0.595	1.149
	(-0.557, 1.598)	(-1.051, 3.551)
Blog Posts (Passive State)	0.092	-0.011
	(-0.083, 0.260)	(-0.813, 0.538)
Blog Posts (Active State)	0.036	0.131
	(-0.253, 0.297)	(-0.733, 0.743)
Tweets (Passive State)	0.004	-0.085
	(-0.016, 0.028)	(-0.155, -0.024)
Tweets (Active State)	0.023	0.005
	(-0.013, 0.053)	(-0.069, 0.065)

Transition Probability Covariates should be interpreted as the effect of moving a customer from the specified state to a passive state.

Table A.5: Mean Estimates for Observed Heterogeneity Parameters

Parameter	Personal Usage Model θ^p_{11} (Passive to Passive State)	 θ^p_{21} (Active to Passive State)	Social Usage Model θ^s_{11} (Passive to Passive State)	 θ^s_{21} (Active to Passive State)
Intercept	5.776**	2.446**	5.554**	2.940**
PR Professionals	−0.249	0.939**	0.834	−0.385
Academics & Researchers	−0.397*	0.319	0.602*	0.226
Feedback When Joining	−0.174	−0.541**	−0.876**	0.110
Invited Others	−0.512**	−0.470*	0.228	−1.719**

* The 90% HPD intervals do not include zero.

** The 95% HPD intervals do not include zero.

Figure A.1: Run Test Results for Personal and Social Usage

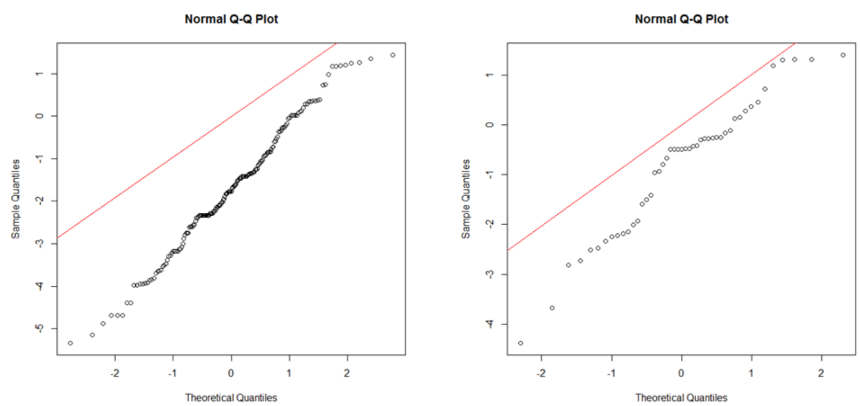

Figure A.2: Dynamic Effect of Inbound Sharing

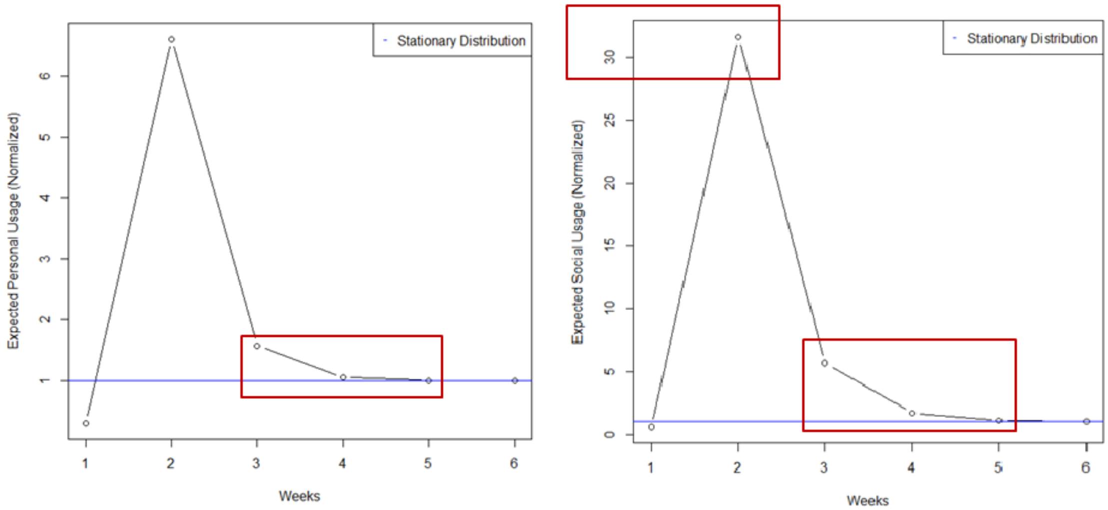

Table A.6: Effect of Cust.-to-Cust. Communication on Transition Probabilities

	Baseline			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	98.0 %	0.74 %	1.22 %	0.04 %
	(97.7% - 98.4%)	(0.48% - 0.97%)	(0.97% - 1.47%)	(0.02% - 0.07%)
Passive, Active	83.31%	15.43%	0.68%	0.58 %
	(80.1% - 87.5%)	(11.15% - 18.62%)	(0.46% - 0.87%)	(0.43% - 0.78%)
Active, Passive	88.04%	0.55%	11.18%	0.23%
	(86.11% - 90.94%)	(0.35% - 0.72%)	(8.28% - 13.23%)	(0.14% - 0.34%)
Active, Active	76.38%	12.22%	7.61 %	3.80%
	(73.13% - 80.11%)	(8.18% - 15.0%)	(5.19% - 9.25 %)	(2.71% - 4.77%)

	Inbound Sharing			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	26.44%	61.03%	1.57%	10.97%
	(5.18% - 48.16%)	(34.93% - 78.88%)	(0.14% - 4.00%)	(3.51% - 18.89%)
Passive, Active	6.19%	81.28%	0.17%	12.36%
	(0.02% - 16.36%)	(68.99% - 93.66%)	(0.00% - 0.62%)	(3.96% - 20.59%)
Active, Passive	25.30%	56.54%	2.71%	15.46%
	(4.80% - 46.95%)	(34.59% - 76.19%)	(0.41% - 6.31%)	(5.44% - 26.06%)
Active, Active	5.94%	75.90%	0.42%	17.74%
	(0.020% - 15.56%)	(57.80% - 89.64%)	(0.00% - 1.85%)	(6.14% - 28.67%)

*Parenthesis denotes the 95% HPD intervals.

Table A.7: Effect of Firm-to-Cust. Communication on Transition Probability

	Baseline			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	98.0 %	0.74 %	1.22 %	0.04 %
	(97.7% - 98.4%)	(0.48% - 0.97%)	(0.97% - 1.47%)	(0.02% - 0.07%)
Passive, Active	83.31%	15.43%	0.68%	0.58 %
	(80.1% - 87.5%)	(11.15% - 18.62%)	(0.46% - 0.87%)	(0.43% - 0.78%)
Active, Passive	88.04%	0.55%	11.18%	0.23%
	(86.11% - 90.94%)	(0.35% - 0.72%)	(8.28% - 13.23%)	(0.14% - 0.34%)
Active, Active	76.38%	12.22%	7.61 %	3.80%
	(73.13% - 80.11%)	(8.18% - 15.0%)	(5.19% - 9.25 %)	(2.71% - 4.77%)
	Blog Posts			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	97.87%	0.75%	1.33%	0.045%
	(97.22% - 98.48%)	(0.31% - 1.26%)	(0.98% - 1.72%)	(0.015% - 0.08%)
Passive, Active	81.80%	16.83%	0.71%	0.67%
	(72.87% - 89.30%)	(8.57% - 25.18%)	(0.46% - 1.04%)	(0.41% - 0.94%)
Active, Passive	87.69%	0.56%	11.51%	0.24%
	(84.33% - 90.83%)	(0.24% - 0.91%)	(8.45% - 14.79%)	(0.078% - 0.40%)
Active, Active	74.90%	13.35%	7.61%	4.14%
	(67.62% - 82.06%)	(6.55% - 20.74%)	(5.14% - 10.47%)	(2.57% - 6.15%)
	Tweets			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	98.06%	0.68%	1.22%	0.038%
	(97.73% - 98.40%)	(0.46% - 0.90%)	(0.99% - 1.46%)	(0.018% - 0.062%)
Passive, Active	83.27%	15.47%	0.68%	0.59%
	(80.41% - 87.44%)	(11.28% - 18.35%)	(0.47% - 0.86%)	(0.44% - 0.77%)
Active, Passive	87.90%	0.50%	11.39%	0.21%
	(85.89% - 90.58%)	(0.33% - 0.66%)	(8.62% - 13.42%)	(0.13% - 0.31%)
Active, Active	76.21%	12.19%	7.74%	3.86%
	(73.12% - 79.80%)	(8.42% - 14.89%)	(5.44% - 9.40%)	(2.81% - 4.75%)

*Parenthesis denotes the 95% HPD intervals.

A.1.2 Cloud-Based File Sharing Service Dataset

Table A.8: Descriptive Statistics

Key characteristics	Average/Percentage
Overall Observations (Customer-Weeks)	60,211
Overall Maximum Number of Weeks	206
Overall Number of Customers	1,200
Word-of-Mouth Customers	394
Others	806
Overall Average Number of Weeks	50.2
Word-of-Mouth Customers	53.8
Others	48.4
Overall Weekly Average Personal Usage (Num. of Files Synced)	88.22
Word-of-Mouth Customers	89.73
Others	87.48
Overall Weekly Average Social Usage (Num. of Files Shared)	57.48
Word-of-Mouth Customers	59.55
Other	56.47
Perc. Customers Who Received Inbound Sharing	95 %
Perc. of Customers Who Invited Others	38 %

Table A.9: Bivariate HMM Model Estimates

Parameter	Personal Usage Estimates	Social Usage Estimates
	State Dependent Usage Covariates	
β_{01} state dependent intercept	0.377	1.21
- Others (Passive State)	(0.366, 0.389)	(1.20, 1.22)
β_{02} state dependent intercept	1.59	1.36
- Others (Active State)	(1.59, 1.59)	(1.35, 1.36)
WOM (Passive State)	-0.336	0.19
	(-0.356, -0.316)	(0.18, 0.20)
WOM (Active State)	0.075	0.0463
-	(0.0688, 0.0809)	(0.0402, 0.0532)
	Transition Probability Covariates	
θ_{11} threshold (Passive State)	4.778	3.768
	(1.113, 7.609)	(1.179, 6.122)
θ_{21} threshold (Active State)	-2.572	1.818
	(-4.457, 0.752)	(0.0163, 3.411)
$Var(\theta_{11})$	1.305	0.963
	(0.281, 2.118)	(0.245, 1.577)
$Var(\theta_{21})$	1.340	0.799
	(0.342, 1.950)	(0.318, 1.212)
Inbound Sharing (Passive State)	0.031	0.128
	(0.002, 0.064)	(0.0948, 0.162)
Inbound Sharing (Active State)	-0.032	0.022
	(-0.075, 0.005)	(-0.0207, 0.0626)
Blog Posts (Passive State)	0.024	0.0552
	(-0.043, 0.082)	(0.0104, 0.0988)
Blog Posts (Active State)	-0.079	-0.0378
	(-0.124, 0.037)	(-0.0760, -0.006)
Tweets (Passive State)	-0.005	-0.0615
	(-0.067, 0.054)	(-0.107, 0.0128)
Tweets (Active State)	0.147	-0.0416
	(0.093, 0.196)	(-0.107, 0.0282)

Transition Probability Covariates should be interpreted as the effect of moving a customer from the specified state to a passive state.

Table A.10: Mean Estimates for Observed Heterogeneity Parameters

Parameter	Personal Usage Model θ^p_{11} (Passive to Passive State)	Personal Usage Model θ^p_{21} (Active to Passive State)	Social Usage Model θ^s_{11} (Passive to Passive State)	Social Usage Model θ^s_{21} (Active to Passive State)
Intercept	5.683**	3.251**	4.456**	2.225**
Invited Others	−2.366**	−1.774**	−1.799**	−1.062**

** The 95% HPD intervals do not include zero.

Table A.11: Effect of Cust.-to-Cust. Communication on Transition Probabilities

	Baseline			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	89.21%	5.66%	4.47%	0.66%
	(88.79% - 89.64%)	(5.36% - 5.98%)	(4.22% - 4.74%)	(0.59% - 0.73%)
Passive, Active	77.27%	17.61%	3.39%	1.73%
	(75.68% - 79.59%)	(15.45% - 19.21%)	(3.16% - 3.64%)	(1.53% - 1.91%)
Active, Passive	78.91%	4.89%	14.77%	1.43%
	(77.67% - 80.20%)	(4.61% - 5.17%)	(13.57% - 16.02%)	(1.30% - 1.56%)
Active, Active	68.80%	15.00%	11.86%	4.34%
	(66.89% - 71.16%)	(13.08% - 16.40%)	(10.59% - 12.96%)	(3.79% - 4.88%)

	Inbound Sharing			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	88.52%	6.25%	4.49%	0.74%
	(88.00% - 89.05%)	(5.87% - 6.66%)	(4.25% - 4.78%)	(0.65% - 0.82%)
Passive, Active	76.97%	17.80%	3.45%	1.78%
	(75.37% - 79.25%)	(15.61% - 19.33%)	(3.19% - 3.70%)	(1.58% - 1.97%)
Active, Passive	78.69%	5.44%	14.32%	1.55%
	(77.60% - 80.23%)	(5.12% - 5.81%)	(12.91% - 15.36%)	(1.40% - 1.70%)
Active, Active	68.86%	15.28%	11.56%	4.30%
	(66.95% - 71.27%)	(13.35% - 16.55%)	(10.13% - 12.57%)	(3.75% - 4.84%)

*Parenthesis denotes the 95% HPD intervals.

Table A.12: Effect of Firm-to-Cust. Communication on Transition Probabilities

	Baseline			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	89.21%	5.66%	4.47%	0.66%
	(88.79% - 89.64%)	(5.36% - 5.98%)	(4.22% - 4.74%)	(0.59% - 0.73%)
Passive, Active	77.27%	17.61%	3.39%	1.73%
	(75.68% - 79.59%)	(15.45% - 19.21%)	(3.16% - 3.64%)	(1.53% - 1.91%)
Active, Passive	78.91%	4.89%	14.77%	1.43%
	(77.67% - 80.20%)	(4.61% - 5.17%)	(13.57% - 16.02%)	(1.30% - 1.56%)
Active, Active	68.80%	15.00%	11.86%	4.34%
	(66.89% - 71.16%)	(13.08% - 16.40%)	(10.59% - 12.96%)	(3.79% - 4.88%)

	Blog Posts			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	88.92%	5.87%	4.51%	0.69%
	(88.52% - 89.36%)	(5.52% - 6.22%)	(4.25% - 4.81%)	(0.62% - 0.78%)
Passive, Active	77.62%	17.17%	3.48%	1.72%
	(76.03% - 79.67%)	(15.16% - 18.77%)	(3.20% - 3.74%)	(1.54% - 1.92%)
Active, Passive	79.22%	5.13%	14.21%	1.44%
	(78.00% - 80.52%)	(4.78% - 5.46%)	(12.87% - 15.38%)	(1.31% - 1.56%)
Active, Active	69.57%	14.78%	11.54%	4.11%
	(67.54% - 71.72%)	(12.98% - 16.17%)	(10.45% - 12.75%)	(3.60% - 4.70%)

	Tweets			
	t			
t - 1	Passive, Passive	Passive, Active	Active, Passive	Active, Active
Passive, Passive	89.42%	5.43%	4.51%	0.64%
	(88.84% - 89.91%)	(5.07% - 5.76%)	(4.24% - 4.86%)	(0.57% - 0.71%)
Passive, Active	77.78%	17.08%	3.45%	1.70%
	(75.66% - 80.47%)	(14.49% - 19.11%)	(3.18% - 3.72%)	(1.44% - 1.92%)
Active, Passive	77.96%	4.60%	15.97%	1.47%
	(76.43% - 79.40%)	(4.30% - 4.94%)	(14.45% - 17.32%)	(1.33% - 1.63%)
Active, Active	68.31%	14.25%	12.92%	4.52%
	(66.07% - 71.04%)	(11.96% - 16.00%)	(11.78% - 14.40%)	(3.81% - 5.24%)

*Parenthesis denotes the 95% HPD intervals.

Bibliography

Ajzen, I., M. Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Prentice-Hall.

Aral, S., M. Van Alstyne. 2011. The diversity-bandwidth tradeoff. *American Journal of Sociology* (Forthcoming).

Aral, S., D. Walker. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* **57** 1623–1639.

Aronson, E., J. Mills. 1959. The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology* **59** 177–181.

Ascarza, E., B. Hardie. 2011. A joint model of usage and churn in contractual settings. *Working Paper* (Second round review at Marketing Science).

Atchade, YvesF. 2006. An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. *Methodology and Computing in Applied Probability* **8**(2) 235–254. doi:10.1007/s11009-006-8550-0. URL `http://dx.doi.org/10.1007/s11009-006-8550-0`.

Bajari, Patrick, C Lanier Benkard, Jonathan Levin. 2007. Estimating dynamic models of imperfect competition. *Econometrica* **75**(5) 1331–1370.

Bakos, Y. 1997. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science* **43** 1676–1692.

Berger, Paul D, Nada I Nasr. 1998. Customer lifetime value: marketing models and applications. *Journal of interactive marketing* **12**(1) 17–30.

Bezanson, Jeff, Stefan Karpinski, Viral B. Shah, Alan Edelman. 2012. Julia: A fast dynamic language for technical computing. *CoRR* **abs/1209.5145**.

Biyalogorsky, Eyal, Eitan Gerstner, Barak Libai. 2001. Customer referral management: Optimal reward programs. *Marketing Science* **20**(1) 82–95.

Bolton, R.N., K.N. Lemon. 1999. A dynamic model of customers' usage of services: Usage as an antecedent and consequence of satisfaction. *Journal of Marketing Research* **36** 171–186.

Brown, J., P. Reingen. 1987. Social ties and Word-of-Mouth referral behavior. *Journal of Consumer Research* **14** 350–62.

Buttle, Francis A. 1998. Word of mouth: understanding and managing referral marketing. *Journal of strategic marketing* **6**(3) 241–254.

Chan, Tat Y, Chunhua Wu, Ying Xie. 2011. Measuring the lifetime value of customers acquired from google search advertising. *Marketing Science* **30**(5) 837–850.

Chellappa, Ramnath K, Shivendu Shivendu. 2005. Managing piracy: Pricing and sampling strategies for digital experience goods in vertically segmented markets. *Information Systems Research* **16**(4) 400–417.

Chevalier, J., D. Mayzlin. 2006. The effect of Word-of-Mouth on sales: Online book reviews. *Journal of Marketing Research* 345–354.

Ching, Andrew T, Susumu Imai, Masakazu Ishihara, Neelam Jain. 2012. A practitioner's guide to bayesian estimation of discrete choice dynamic programming models. *Quantitative Marketing and Economics* **10**(2) 151–196.

Choi, J., D. Bell, L. Lodish. 2011. Traditional and IS-enabled customer acquisition on the Internet. *Management Science* (Forthcoming).

D. Kahneman, J. L. Knetsch, R. H. Thaler. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives* 193–206.

Davis, F.D. 1986. A technology acceptance model for empirically testing new end-user information systems: Theory and results. *Doctoral Dissertation* (Sloan School of Management, MIT).

Davis, F.D., Richard P. Bagozzi, Paul Warshaw. 1989. User acceptance of computer technology: A comparison of two theoretical models. *Management Science* **35** 982–1003.

Eddelbuettel, Dirk. 2013. *Seamless R and C++ Integration with Rcpp*. Springer, New York. ISBN 978-1-4614-6867-7.

Eddelbuettel, Dirk, Romain François. 2011. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* **40**(8) 1–18. URL `http://www.jstatsoft.org/v40/i08/`.

Fader, Peter S, Bruce GS Hardie, Ka Lok Lee. 2005. Rfm and clv: Using iso-value curves for customer base analysis. *Journal of Marketing Research* 415–430.

Frank, Ronald E. 1962. Brand choice as a probability process. *The Journal of Business* **35**(1) 43–56.

Friestad, M., P. Wright. 1994. The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research* **21** 1–31.

Friestad, M., P. Wright. 1995. Persuasion knowledge: Lay people's and researcher's beliefs about the psychology of advertising. *Journal of Consumer Research* **22** 62–74.

Gelman, Andrew, Donald Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* **7** 457–511.

Ghose, A., S. Yang. 2008. An empirical analysis of search engine advertising: sponsored search in electronic markets. *Proceedings of the ACM International Conference on Web Search and Datamining Conference* (WSDM 2008).

Girolami, Mark, Ben Calderhead. 2011. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(2) 123–214. doi:10.1111/j.1467-9868.2010.00765.x. URL `http://dx.doi.org/10.1111/j.1467-9868.2010.00765.x`.

Griewank, Andreas, David Juedes, Jean Utke. 1996. Algorithm 755: Adol-c: a package for the automatic differentiation of algorithms written in c/c++. *ACM Transactions on Mathematical Software (TOMS)* **22**(2) 131–167.

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

Gupta, Sunil, Carl Mela, Jose Vidal-Sanz. 2006b. The value of a free customer. *Harvard Business School Working Paper* .

Hanemann, W. Michael. 1984. Discrete/continuous models of consumer demands. *Econometrica* 541–61.

Heiman, Amir, Bruce McWilliams, Zhihua Shen, David Zilberman. 2001. Learning and forgetting: Modeling optimal product sampling over time. *Management Science* **47**(4) 532–546.

Heiman, Amir, Eitan Muller. 1996. Using demonstration to increase new product acceptance: Controlling demonstration time. *Journal of Marketing Research* 422–430.

Herr, P., F. Kardes, J. Kim. 1991. Effects of Word-of-Mouth and product-attribute information on persuasion: an accessibility-diagnosticity perspective. *Journal of Consumer Research* **17** 454–62.

Hogan, J.E., K.N. Lemon, B. Libai. 2003. What is the true value of a lost customer? *Journal of Service Research* **5**(3) 196–208.

Homans, G. 1958. Social behavior as exchange. *American Journal of Sociology* **63** 597–606.

Hotz, V. J., R. A. Miller. 1993. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies* **60** 497–529.

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

Imai, Susumu, Neelam Jain, Andrew Ching. 2009. Bayesian estimation of dynamic discrete choice models. *Econometrica* **77**(6) 1865–1899.

Iyengar, R., A. Ansari, S. Gupta. 2007. A model of consumer learning for service quality and usage. *Journal of Marketing Research* 529–544.

Jain, Dipak, Vijay Mahajan, Eitan Muller. 1995. An approach for determining optimal product sampling for the diffusion of a new product. *Journal of Product Innovation Management* **12**(2) 124–135.

Kirkpatrick, David. 2010. *The Facebook Effect*. Simon & Schuster.

Laibson, David. 2001. A cue-theory of consumption. *The Quarterly Journal of Economics* **116**(1) 81–119.

Lehmann, Donald R, Mercedes Esteban-Bravo. 2006. When giving some away makes sense to jump-start the diffusion process. *Marketing Letters* **17**(4) 243–254.

Lemon, K.L., T.B. White, R.S. Winer. 2002. Dynamic customer relationship management: incorporating future considerations into the service retention decision. *Journal of Marketing* **66** 1–14.

Liechty, J., R. Peters, M. Wedel. 2003. Global and local covert visual attention: evidence from a Bayesian hidden Markov model. *Psychometrika* **68** 519–541.

Michelangeli, Valentina. 2008. Does it pay to get a reverse mortgage. *Doctoral Thesis Manuscript. Department of Economics, Boston University.* .

Montgomery, A., S. Li, K. Srinivasan, J. C. Liechty. 2004. Predicting online purchase conversion using web path analysis. *Marketing Science* **23** 579–595.

Montoya, R., O. Netzer, K. Jedidi. 2010. Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Science* **29** 909–924.

Muller, E., R. Peres, V. Mahajan. 2010. Innovation diffusion and new product growth models: a critical review and research directions. *International Journal of Research in Marketing* **27** 91–106.

Nam, S., P. Manchanda, P. K. Chintagunta. 2010. The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service. *Marketing Science* **29** 690–700.

Needleman, S., A. Loten. 2012. When freemium fails. *The Wall Street Journal* .

Netzer, Oded, James Lattin, V. Srinivasan. 2008. A hidden Markov model of customer relationship dynamics. *Marketing Science* 185–204.

Niculescu, Marius, DJ Wu. 2013. Economics of free under perpetual licensing: Implications for the software industry. *Available at SSRN 1853603* .

Pakes, Ariel, Michael Ostrovsky, Steven Berry. 2007. Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics* **38**(2) 373–399.

Plott, C. R., K. Zeiler. 2005. Subscribed content the willingness to pay – willingness to accept gap, the endowment effect, subject misconceptions, and experimental procedures for eliciting valuations. *The American Economic Review* **95** 530–545.

Raftery, Adrian E, Michael A Newton, Jaya M Satagopan, Pavel N Krivitsky. 2007. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian Statistics* (8) 1–*45.

Reichheld, Frederick F. 2003. The one number you need to grow. *Harvard business review* **81**(12) 46–55.

Rogers, E.M. 2003. *Diffusion of Innovations*. Free Press.

Roos, Jason MT. 2012. Hyper-media search and consumption. Ph.D. thesis, Duke University.

Rust, J. 1987. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica* **55** 999–1033.

Rutz, O., R. Bucklin. 2011. From generic to branded: a model of spillover dynamics in paid search advertising. *Journal of Marketing Research* **48** 87–102. (Forthcoming).

Ryan, Stephen P. 2012. The costs of environmental regulation in a concentrated industry. *Econometrica* **80**(3) 1019–1061.

Ryu, Gangseog, Lawrence Feick. 2007. A penny for your thoughts: Referral reward programs and referral likelihood. *Journal of Marketing* **71**(1) pp. 84–94. URL `http://www.jstor.org/stable/30162131`.

Schweidel, David A, Eric T Bradlow, Peter S Fader. 2011. Portfolio dynamics for customers of a multiservice provider. *Management Science* **57**(3) 471–486.

Silverman, George. 1997. How to harness the awesome power of word of mouth. *DIRECT MARKETING-GARDEN CITY-* **60** 32–37.

Song, Inseong, Pradeep K Chintagunta. 2007. A discrete: Continuous model for multicategory purchase behavior of households. *Journal of Marketing Research* 595–612.

Su, Che-Lin, Kenneth L Judd. 2012. Constrained optimization approaches to estimation of structural models. *Econometrica* **80**(5) 2213–2230.

Timmins, Christopher. 2002. Measuring the dynamic efficiency costs of regulators' preferences: Municipal water utilities in the arid west. *Econometrica* **70**(2) 603–629.

Trusov, M., A. Bodapati, R. Bucklin. 2010. Determining influential users in internet social networks. *Journal of Marketing Research* **47** 643–658.

Trusov, M., R. Bucklin, K. Pauwels. 2009. Effects of Word-of-Mouth versus traditional marketing: findings from an internet social networking site. *Journal of Marketing* **73** 90–102.

Van den Bulte, Christophe, Yogesh V Joshi. 2007. New product diffusion with influentials and imitators. *Marketing Science* **26**(3) 400–421.

Venkatesh, V., M. G. Morris, P. L. Ackerman. 2000. A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organizational Behavior and Human Decision Processes* **83** 33–60.

Verhoef, P., B. Donkers. 2005. The effect of acquisition channels on customer loyalty and cross-buying. *Journal of Interactive Marketing* **19**.

Villanueva, J., S. Yoo, D. M. Hanssens. 2008. The impact of marketing-induced versus Word-of-Mouth customer acquisition on customer equity growth. *Journal of Marketing Research* **45** 48–59.

White, T., K. Lemon, J. Hogan. 2007. Customer retention when the customer's future usage is uncertain. *Psychology and Marketing* **24** 849–870.

Wickham, Hadley. 2014. *Advanced R*. CRC Press. URL `http://adv-r.had.co.nz/`.

Wojnicki, A., D. Godes. 2012. Signaling success: Strategically-positive word of mouth. *Working Paper* (3rd Round Review at Journal of Marketing Research).

Wood, Wendy, David T Neal. 2007. A new look at habits and the habit-goal interface. *Psychological review* **114**(4) 843.

Wood, Wendy, David T Neal. 2009. The habitual consumer. *Journal of Consumer Psychology* **19**(4) 579–592.

Yao, S., C. Mela. 2011. A dynamic model of sponsored search advertising. *Marketing Science* **30** 447–468.