Essays on Telecommunications Management: Understanding Consumer Switch, Search and Purchase Behaviors

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Engineering of Engineering and Public Policy

Baojiang Yang

B.S., Electrical Engineering, Shanghai Jiao Tong University
M.S., Electrical Engineering and Computer Science, Northwestern University
M.S., Engineering and Public Policy, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

May, 2018

Abstract

Digitization has been pervasively reshaping the landscape of home-based telecommunication industries. The massive disruptive challenges call for telecom companies to react with efficient strategic managerial policies. Meanwhile, how consumer decision makings and welfare are impacted by such policies often remains complicated and non-transparent to policy makers. My thesis aims at leveraging large-scale empirical data to investigate the impacts of several prevalent firm initiated strategies on both sides of the market, i.e. consumers and firms. The thesis is comprised of three studies focusing on consumer *switch, search* and *purchase* behaviors.

The first study, centering at consumer switching behaviors, investigates the impact of lock in shortening policies on both firm profits and consumer welfare in home-based telecommunication service market. Using household level data from a large telecommunications service provider, we show that a market level policy that shortens the lock-in period from the status quo can decrease the profits on the firms side more than it increases consumer surplus. This is majorly caused by the substantial acquisition costs associated with user switching and service initiation. As a result if regulators shorten lock-in periods but then firms respond by collaboratively increasing prices to recover their rate of return, the consumers, as the analyses indicate, may be worse off compared to the world in which lock-in periods do not change. Therefore lock-in reduction policy need to be paired with a policy precluding operators from increasing prices too much.

The later two studies jointly examine consumer's search and purchase behaviors in social environment. With a wide scope of services, telecommunication service providers can often leverage their knowledge on consumer's social environment to reshape consumer choices. We aim to understand how consumers combine different sources of social information, one from friends versus one from the crowd, as a function of how close they are to the point of conversion. We developed a dynamic structural econometric model that jointly describes consumer information search and product purchase while taking into account sequential arrival of information and non-negligible search costs. The model is then instantiated on two connected yet distinct empirical contexts, where consumers shop for movies to watch on home screens.

The first empirical context (discussed in the second study in this thesis) lies in an observational setting, where we studied individual level clickstream and transactional data from a Video-on-Demand service platform operated by a large telecommunication service provider. Later in the third study, we created an artificial movie market and leveraged a randomized web experiment to further study the research questions with more solid identification support. We find that, in both contexts, consumers seem to start by browsing products they heard about from friends. The popularity signals become more relevant when consumers getting closer to the point of purchase. The results have important managerial implications to online vendors by suggesting a reasonable strategy of providing the most valuable social information at the right time to enhance consumer shopping experiences.

Acknowledgements

I acknowledge the financial support provided by Portuguese Foundation for Science and Technology through the Carnegie Mellon Portugal Program under Grant *SFRH/BD/52611/2014*. I also acknowledge the support from the Portuguese Foundation of Science and Technology with grant reference UID/GES/00407/2013 provided through Catlica Lisbon Research Unit in Business and Economics (CUBE). I would also like to express my appreciations to our industrial partner for their collaboration and data during my research.

I would like to thank my advisor Professor Pedro Ferreira, for all his precious support and guidance, for all the time dedicated to this research, for the insightful discussions, and for the immense help to my life in both Pittsburgh and Lisbon. I thank my co-advisor, Professor Miguel Godinho de Matos, for his countless support and detailed guidance throughout my research, and for the knowledge he passed to me and inspired me to learn. I am also indebted to Professor Mario Silva, for his longstanding support and help for both my research and life during the years in Portugal.

I thank the members of my Ph.D. committee: Professor Pedro Ferreira (chair), Professor Miguel Godinho de Matos, Professor Alex Davis, and Professor Rodrigo Belo, for their interests in my research, for the time dedicated throughout the journey, and for the suggestions and comments that have greatly helped to improve the quality of this thesis.

My appreciation goes to all the professors in EPP and Heinz College at Carnegie Mellon, Catlica Lisbon School of Business and Economics-UCP, and Instituto Superior Tcnico, whom I had opportunities to know in person, for all their help and support whenever needed, and for all the lively discussions that have immensely helped my research work.

I would like to leave a special word to my colleagues and friends during my PhD study. We made the office a lovely place like home; we had wild brainstorms that were not limited by our majors and specialties; and continuously enhanced the tie strength without having Granovetter's theory in mind. Not forgetting many others, my special word goes to Rodrigo, Filipa, Antonio, Qiwei, and Xiaochen in my research group, Nandi, Fan, Andrew (Hamann), Hassan, Alex (Ligo), Yu and Paul in my department, along with Pu, Hang, Shanghang, Lorena, Ana Paula, and Tatiana Overseas.

Most of all, I would like to thank my wife, Haylee, who has been present during the good and bad times along this physical and mental journey, and provided countless support to me in virtually everything along my PhD life. Finally, my respectful gratitude goes to my lovely parents, Yingguo Yang and Qingli Ge, for their unconditional support and encouraging belief in me.

Contents

1 Introduction **1**

2 The Effect of Shortening Lock-in Periods in Telecommunication Services **8**

2.1 Introduction 8

2.2 Literature Review 11

2.3 Empirical Context 12

2.4 Empirical Approach 15

2.4.1 The Choice Model 16

2.4.2 Churn rates, Consumer Surplus and Firm Profits 18

2.5 Econometric Estimation 22

2.6 Policy Simulations 23

2.6.1 Unchanged Prices 25

2.6.2 Increased Prices 29

2.6.3 Reduced Prices 31

2.7 Conclusions 34

3 The Interplay of Information from Friends Versus Crowds in the Consumer Shopping Journey: An Observational Study **37**

3.1 Introduction 37

3.2 Related Works 42

3.2.1 Influence From Popularity Information 42

3.2.2 Influence From Friends 43

3.2.3 Peer Influence In Movie Industry 44
3.2.4 The Conversion Funnel and Consumer Information Search 45
3.3 Data 47
3.4 Model and Estimation 51
3.4.1 A Structural Model for Search-Purchase Decision Makings 51
3.4.2 Estimation 54
3.5 Results 59
3.5.1 Impact of Social Signals on Search vs Purchase Decisions 59
3.5.2 Valence of Signals from Friends 63
3.6 Discussions and Conclusions 64

4 Value of Social Signals in Consumer Search and Purchase: An Randomized Web Experiment 69

4.1 Introduction 69
4.2 Related Works 72
4.2.1 Identification of endogenous social effect 72
4.2.2 Incentive Design in Lab Experiment 74
4.3 Experiment Design 75
4.3.1 *MoviePlatform* 75
4.3.2 Randomization 78
4.3.3 Social Signals Impression 80
4.3.4 Experiment Implementation 81
4.4 Data 85
4.5 Model and Estimation 88
4.5.1 Model Revisit 89
4.5.2 Empirical Specifications 90
4.6 Results 91
4.6.1 Estimation Results 93
4.6.2 Comparing With The Observational Study 94

4.7 Discussion and Conclusion 96

5 Conclusions **100**

A Appendix for Chapter 2 **112**

A.1 Simulation Details 112

A.1.1 Churn Rate Estimation 112

A.2 Complementary Policy Simulations 115

A.2.1 Price changes to maintain Consumer Surplus 115

A.3 Additional Statistics 116

B Appendix for Chapter 3 **118**

B.1 Distribution of Uncertainty 119

B.2 Calculating Reservation Utility 120

B.3 Estimation Algorithm 121

B.4 Model Performance and Alternative Models 122

B.5 Additional Descriptive Statistics 125

C Appendix for Chapter 4 **128**

C.1 Ethics Statement 129

C.2 Experiment Implementation 132

C.2.1 Survey Structure 132

C.2.2 Reward calculation 133

C.3 Alternative Models: Logistic regression 135

C.4 Supplementary survey questions 137

C.5 Demographic Characteristics of Experiment Participants 140

List of Figures

2.1 Density of changes inside the carrier and churn as a function of time to lock-in expiry 16

2.2 Flow chart of our simulation approach 17

2.3 Difference in firm profits collected from an existing subscriber's monthly payment ($top - left$), breaching penalties($top - right$), total earnings ($bottom - left$) and in consumer surplus ($bottom - right$) in dollars when the lock-in period reduces from 24 months. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using the estimation results in column (1) of Table 2.3. 27

2.4 Difference in firm profits ($left$) defined as Eq (2.15) and in market welfare ($right$) in dollars when the lock-in period reduces from 24 months. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using the estimation results in column (1) of Table 2.3. 29

2.5 Percentage change in price ($left$) and in change consumer surplus ($right$) when the lock-in period reduces and the firm reacts to keep the same level of profit, comparing to the status quo of 24-month lock-in periods. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using estimation results in column (1) of Table 2.3. 30
2.6 Effect of price decrease (in percentages) on firm overall profit (left column), consumer surplus (middle column), and market welfare (right columns) when lock-in was reduced from 24 months to 16 months. The numbers are comparing to status quo 24 months with unchanged prices. Plots are based on simulations using estimation results in column (1) of Table 2.3. A set of contour plots showing how prices and lock-in reduction may jointly affect the above metrics can be find in Figure 2.7 in appendix. 32
2.7 Contour plots of joint effect of price decrease (in percentages) and lock-in reduction (in months) on firm overall profit (left column), consumer surplus (middle column), and market welfare (right columns) in a four-provider market. Combinations between two firm RoRs and three consumer discount rates were shown. Plots are based on simulations using estimation results in column (1) of Table 2.3. . . . 33

3.1 Digital Shopping Journey in a Social Environment 47
3.2 Number of friends distribution 50
3.3 Model Framework 55
3.4 Impact of Social Signals as a Function of Price 62
3.5 MRS as Function of Price 63
3.6 Impact of social signals as a function of price when distinguishing the valence of signals from friends 65

4.1 Example of catalog page in *MoviePlatform* 76
4.2 Example of movie landing page in *MoviePlatform*. 77

4.3 Example of social network displayed to participants 82
4.4 A participant's experience in the HIT 84
4.5 Instructions on study reward calculation used in the *Qualtrics* survey 85
4.6 Number of different movies and number of hypothetical friends browsed 86
4.7 Time spent on different stages of the study 87
4.8 Marginal rates of substitutions between social signals in the observational study vs experimental study 95
4.9 Dollar values of social signals in the observational study vs experimental study . . 96

A.1 Probability of churn (Top) and probability of survival ($Bottom$) when the length of the lock-in period varies according to the estimation results in column (1) of Table 2.3. Two market local competition scenarios were represented: solid curves correspond to market with one competitors (two service providers market) and dashed curves correspond to one with three competitors (four service providers market). 114
A.2 Percentage change in price ($left$) and percentage change in firm profit ($right$) when the lock-in period reduces and the firm reacts to increase price up to making consumers indifferent to the status quo of 24 months lock-in period. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using estimation results in column (1) of Table2.3. 116
A.3 Boxplots of prices of same products for consumers before and after contract expiry. The plots were broken down to each service products and the boxplots show the 5%, 50% and 95% quantiles. 117

B.1 Comparison of the ROC curves 123
B.2 Distribution of "Number of Likes" for movies when in front of consumers. The x-axis is in Log-scale. 125

B.3 Distribution of number of movies (among the 1771 available movies) that had at least one friend rentals when search session begins for all the users studied. 126

B.4 Distribution of number of "Friends Rental Signals" that consumers could observe at the beginning of the search session. The histogram is quite similar to Figure B.3 as the friend rental signals were dispersed among movies. 126

B.5 Cumulative Density Plot for the proportion of movie streamed by consumers who rented. The dashed line corresponds to 85% of the movie runtime being streamed (cumulatively after rentals). 127

C.1 Left shows how lottery is embedded in Qualtrics survey. Right shows how final payment is calculated. 134

C.2 Quiz to evaluate whether participants understood the reward calculations. The participants were asked "Suppose you rent movie "Life of Pi" priced at $6.00 at MoviePlatform, you win the lottery and you complete your task. What would you get as a reward for your participation in the study?". Five options were provided in random order for participants to choose from (respective to the five options listed vertically in the graph). (1) I will just get the HIT task completion fee of $0.50; (2) I will get HIT task completion fee of $0.50 plus $10.00. In total, I get $10.50; (3) I will get HIT task completion fee of $0.50 plus a voucher to watch "Life of Pi". In the end, I get $0.50 plus the movie voucher; (4) I will get HIT task completion fee of $0.50, a voucher to watch "Life of Pi" and the balance to $10.00. In the end, I get $4.50 ($0.50 + $4) plus the movie voucher; (5) I will get HIT task completion fee of $0.50, a voucher to watch "Life of Pi" and $10.00. In the end, I get $10.50 plus the movie voucher. Each participant can choose a single answer. Here Option (4) here is the right answer. When a participant missed the right answer, we reaffirm the calculation by giving the explanation of the right answer. 138

C.3 In this survey question we asked whether the participants had relied on any of the social information displayed for a movie, including the number of likes and number of friends rentals. We showed an example of where these signals appear on *MoviePlatform* to remind the participants their shopping experiences and in particular direct them to the signals appeared specifically on the platform. Then we asked "While browsing movies at MoviePlatform, did you rely on information about how many people liked a movie and how many of your friends have rented a movie?" Four options were provided in randomized orders to each participant: (1) No, I didn't rely on any of them; (2) Yes, sometimes seeing the number of likes for a movie affected my decision to browse or rent the movie; (3) Yes, sometimes seeing ¡strong¿a friend rented ¡/strong¿a movie affected my decision to browse or rent the movie, and (4) Yes, I relied on both. The options are exclusive. The plot shows the number of selections to each options respectively. 139
C.4 Distribution of participants' educational background 140
C.5 Distribution of participants genders 140
C.6 Distribution of participants household annual income in USD 141
C.7 Distribution of participants races 141

List of Tables

2.1 Summary statistics on Triple-Play packages offered by our industrial partner . . . 14

2.2 Interpretation of coefficients for discrete choice models 19

2.3 Estimation results of discrete choice models 24

3.1 Descriptive Statistics 48

3.2 Information Updates of Search 53

3.3 Interpretations of Variables in Empirical Model 58

3.4 Estimation results: Impact of social signals on search vs purchase decisions 60

3.5 Estimation results when distinguishing the valence of signals from friends 65

4.1 Distributions for displayed randomized volume of social signals and prices for movie j observed by user i 79

4.2 Correlations between movie characteristics with number of likes, number of friend rentals, and price charged for movies on *MoviePlatform*. The brackets below the correlation estimate show the P-Values the Pearson correlation coefficients 80

4.3 Descriptive statistics for user behaviors on *MoviePlatform* 86

4.4 Measures of study transparency and clarity 88

4.5 Characteristics for movies displayed on *MoviePlatform* 88

4.6 Main Estimation Results Following the Sequential Search Framework 92

4.7 Comparing the observational study and experiment study: Marginal rates of substitution between signals and signal Dollar values 94

B.1 Alternative Model: Logit 124

B.2 Comparison in predictive metrics: sequential search model versus logistic regression model 125

C.1 Alternative Model: Logistic Regression 136

Chapter 1

Introduction

Digitization is profoundly reshaping the landscape of home-based telecommunication industries. Incumbent telcos are facing increasingly fierce competitions from various innovative business models and technologies in the new digital age. Consumer preferences on traditional communication products are rapidly changing at the same time. McKinsey (2015) predicts that even the overall communications activity grows, the total consumer spends on traditional services has entered a period of slow decline.

To survive in the digital economy and remain profitable, incumbent telecommunication service providers can leverage their unique market power in response to these challenges. From a protector's perspective, incumbent telcos operate large installed customer base. Effective customer relationship management instruments can help firms maintain a substantial amount of profitable subscribers (Netzer et al., 2008). From the attacking perspective, traditional telcos usually provide a broad umbrella of services in high volume to households. They can therefore take advantages of economies of scale and economies of scope to strategically provide customized services to consumers so as to protect the market status (Prince and Greenstein, 2014).

On the other side of the market, consumers are directly influenced by service provider's strategic actions. With firms being profit maximizers, profit-driven strategies can be either beneficial or detrimental to consumers. When considering the necessity of particular policy interventions, policy makers need to comprehensively evaluate the role such profit-driven strategy plays on both sides of the market, balancing firm's profitability and consumer welfare. The first yet fundamental step for such evaluation is to understand how firm initiated strategies may interfere consumer decision makings.

My dissertation aims to enhance the understanding on three types of consumer decision making behaviors, *switching, searching*, and *purchasing*, under the interferences of several firm-initiated managerial strategies. The first study (introduced in Chapter 2) examines the impact of lock in contracts associated with services on consumer switching behaviors, while the second (introduced in Chapter 3) and third study (introduced in Chapter 4) jointly researched consumer's search and purchase behaviors in a social environment where service providers are capable to manipulate the salience of various social signals.

Chapter 2 focuses on the relationship between the lock-in periods and consumer switching behaviors in home-based telecommunication service services. Lock-in periods are a common practice employed by telecommunication providers to reduce the risk of failing to cover the significant capital expenditure needed for building the network in the first place and upgrading it over time. In this type of markets, consumers can not terminate contracts before the lock-in period is over and leave to a competitor – an action called churn – without paying significant financial penalties. Lock-in periods are a particular case of switching costs, which, in essence, include any mechanism by which firms enforce to reduce the incentive for consumers to leave.

Telecommunication regulators have been concerned about the effect of lock-in periods on consumer welfare. For example, in the European Union, the Telecommunications Law enforces that service contracts that lock-in consumers cannot exceed 24 months and operators must offer at least one alternative with a lock-in period shorter than 13 months (European-Union, 2009). Recent regulation by the FCC in the US (CTIA, 2014) requires mobile providers to allow consumers to unlock phones for free (and thus change provider and keep their device) once they stayed with the carrier for 24 months or paid the contracted financial penalty to leave.

Previous literature suggests that the effect of lock in on consumer welfare is more complicated than its first look. On one hand, longer lock-in periods reduce the consumers' freedom to change telecommunications provider and entrants have a hard time to steal consumers that are locked in to the market leader. On the other hand, low switching costs reduce the firms' incentives to attract consumers in the first place, which may result in higher starting prices (Dubé et al., 2009). Meanwhile, in markets where service switching generates additional cost to the market, low switching costs cause more frequent user switching that may render the market less efficient (Gans, 2001). Therefore, regulators are mostly interested in learning the impact of reducing the lock-in period on consumer welfare and, necessarily, on firm profit too, in order to anticipate how firms are likely to respond to potential changes in the current policy.

In the first study (Chapter 2), we measure the welfare implications of shortening the lock-in period to less than the current status quo of 24 months. We study the market for triple-play services, which is now dominant both in the US and in the EU (OVUM, 2015). We use data from a large provider between April and October 2013. For each household over time we have information on the service bundle they subscribe, when households change service within the company and when they churn, along with information on similar bundles from competitors and on the prices they charge in the local market.

We fit a multinomial logit model to this data to study how consumers make service choices with switching costs at play. We find empirical evidence of the existence of switching cost for both inside provider service switching and between service providers service switching. We then conducted a series of policy simulations to study the potential impact of lock in reduction policy on both firm profits and consumer welfare. Interestingly, we find that shortening the lock-in period from the current status quo of 24 months decreases the profit of the firm more than it increases consumer surplus. The intuition is that when lock in reduces, the market would observe earlier and more frequent user switching, which generates more substantial acquisition costs associated with service initiation to the firms side. Our simulations indicate that a consequence of this result is that consumer surplus may instead reduce if firms react to shorter lock-in periods by increasing prices to keep their profit levels. Therefore, our study shows that regulators enforcing shorter lock-in periods must also regulate prices to avoid significant loss to consumers.

Our study is the first to measure the switching costs associated to lock-in periods in the context of triple play services in telecommunications markets. Furthermore, we do so at the household level and we measure the effect of shortening the lock-in period on consumer surplus and firm profit, therefore going a step further relative to the prior literature that studied only the effect of switching costs on churn rates and, sometimes, on prices. We expect this work to provide novel insights to telecom regulators at a time when regulating the length of lock-in periods in being heavily debated in the industry, and in particular on the importance of using several policies in tandem to achieve the desired goals.

In Chapter 3 and Chapter 4, we turn our focus to consumers' search and purchase behaviors in a representative service market, i.e., video-on-demand (VoD) services, where consumers shop movies to watch from their home screens. VoD systems are usually provided by home-based telecom companies and combined with their pay-TV services to compete with the Over-The-Top (OTT) services.

Consumer shopping process often starts way ahead of the point of purchase. Instead, the final conversion may be determined by a longer consideration process, which is usually termed by marketers as the *conversion funnel* (Edelman, 2015; Kotler and Armstrong, 2010). The typical conversion funnel consists of several stages that usually happen chronologically: awareness, information search, before purchase evaluation, purchase, and post-purchase activities (**?**).

We focus on the central part of the conversion funnel starting from information search and ending with the decision of purchase or no purchase. Consumers leverage information search to resolve the uncertainty related to product utility (Feinberg and Huber, 1996). But search is often costly, such that they cannot explore all the options. The typical search to purchase process begins with consumers sample a subset of goods for further exploration to form a reasonable consideration set (Stigler, 1961; Ke et al., 2016). Later, they choose a product to buy from this consideration set or abandon the market without buying (Mehta et al., 2003).

Consumers online shopping decisions can be influenced by multiple sources of information at different stages (Branco et al., 2012). Among the intricate sources of relevant product information, an important source arises from the social environment that consumers are embedded in. Social influence literature suggests that social signals play important roles in personal outcomes such as occupation (Calvo-Armengol and Jackson, 2004), health (Christakis and Fowler, 2007) and product consumption(Aral and Walker, 2011b; Bapna and Umyarov, 2015), etc. Service providers, therefore, can take advantage of the knowledge on the consumer social networks and historical consumptions to strategically manipulate the display of social information, in order to interfere consumer decision makings and achieve business objectives. Consumers, in the meantime, can be affected by mixed mechanisms associated to the vendor's strategical display of social information. On the one hand, consumers receive valuable quality references for their peer consumers or socially connected individuals, which may in return guide them to make reasonable choices with less search costs. On the other hand, firm's profit-driven information manipulation may distort consumer decision makings in a way that harms consumer welfare, for example, making them purchase something they don't actuarially need.

Our research is designed to understand whether firm's strategical display of social information can affect consumer decision makings. We are interested social signals from two distinct sources, *popularity information assessed by the crowd*, and *information from friends in the social network*. Our research questions are: *(1) whether consumers combine such information when they are making search decisions and purchase decisions? (2) If yes, do they assign different weights to these two sources of information when they are at different stages of their shopping journey?*

We answer these questions by first developing a dynamic structural econometric model to jointly describe the series of decisions consumers make starting from information search and ending with determining whether to purchase a product. The model combines a discrete choice model with an optimal stopping framework following the theory of sequential search (Weitzman, 1979). The main idea is that consumers start the shopping process by sampling the product for which the information search is most valuable, and keep sampling sequentially until the marginal search cost of searching an additional product exceeds the expected marginal benefit of searching. We assume consumers are rational and forward-looking such that the decisions they make at any stage maximize the expected future return given the information available at that time. Our model takes into accounts the sequential arrival of information revealed by search and non-negligible search costs.

Chapter 3 and Chapter 4 then used two highly connected yet different empirical context to instantiate the structural model and study the above research questions. In Chapter 3, we study a set of observational data obtained from the VoD platform of a large multinational telecommunication services provider. Our primary dataset includes individual level click stream data and data on transactions in the VoD system from 117 thousand randomly sampled users. We construct the social graph between users using their Call Detail Records (CDRs) for cell phone communications provided by the same service provider between August and October 2015.

Noticing that merely using observational data may induce identification challenges if we aim to identify social influence from other underlying mechanisms (Manski, 1993), in Chapter 4, we created an artificial movie market by constructing and operating a VoD system and leveraged a randomized web experiment to further study the research questions.

Interestingly, in both empirical context, we find very similar results in regard of the research questions. The results show that consumers seem to use the information from friends to find the set of movies for which they would like to conduct information search and form a consideration set. In a later stage, popularity information becomes more relevant to consumers when they are making purchase decisions, conditional on the consideration set formed through information search. One possible explanation to the findings is that at the start of information search, consumers face high uncertainties on their shopping objectives, such that they are more willing to rely on friends' opinions as they might be seen as more trustworthy and diagnostic to their own tastes. On the other perspective, popularity signals are more relevant to quality assessments as they reflect the wisdom of crowds. When consumers are making purchase decisions, quality measures become more relevant and the reliance on information from the crowds increases.

Our findings have several valuable managerial implications. For online marketplace of digital experience goods such as online videos, books, and music where both popularity and friend influence might be at play, vendors can leverage their knowledge about the social networks to strategically display social information in order to obtain better user monetization value. If the objective is to enhance consumer shopping experiences and improve user engagement in the long term, a reasonable strategy would be to highlight customized friends information from consumers own social network earlier in their shopping process, to guide them through the large log of products, but later highlight the popularity information, to assist their product quality assessments. If the vendors still want the friends information to play a role at the point of purchase, they may consider enrich it by combining it with more salient valence measures from whom the signals were generated. Finally, although the two sources of social information exhibit different influences along the conversion funnel, both exert positive impact on the overall purchase intentions. vendors should be encouraged to explore more informative popularity and social network information to display, meanwhile incentivize the creation of social network and encourage conversations between friends.

The rest of the dissertation is organized as follows: Chapter 2 expands the discussion on contractual lock-in by elaborating the major results from the study; In Chapter 3, I first develop the structural econometric model on consumer search and purchase, and then introduce the observational study on how consumers combine social signals when doing information search and product purchase; The experimental study focusing on the same research question is provided in Chapter 4; Chapter 5 summarizes the conclusions.

Chapter 2

The Effect of Shortening Lock-in Periods in Telecommunication Services

2.1 Introduction

In industries with significant upfront capital expenditure for customer acquisition, firms typically opt for subscription-based business models locking-in consumers into long-term contracts. These contracts, often termed lock-in periods in the industry, aim at ensuring that consumers stay with the firm enough time so that the cumulative of their monthly bills over their tenure with the company covers not only the costs with maintenance and service provision, but also the initial costs of consumer acquisition and service deployment (Farrell and Klemperer, 2007)[1].

In this type of markets, consumers can still terminate contracts before the lock-in period is over and leave to a competitor – an action called churn. However, to do so, they need to pay financial penalties established in the contractual agreement with the firm. These penalties are typically set in a way that allows the firm to cover its upfront deployment costs to service the households which are often substantial.

Starting from the early 2000s, in the telecommunications sector, several firms offered lengthy contracts that locked-in consumers for long periods of time. In the UK, Orange had deliber-

[1]Examples of industries that share these characteristics include the Telcommunication's industry, the Home Security and Surveillance industry and several utility firms that need to deploy equipment on the premises of the consumer that allow accessing/measuring the service that is provided (e.g. set-top-box, internet modems, security cameras, measurement equipments, etc.)

ately encouraged customers to sign up for 24 month contracts in its stores (Capgemini, 2009). In Canada, Rogers Communications, BCE, and Telus, sold three-year contracts for mobile phone service (CRTC, 2012) and in Asian countries such as South Korea, Japan and China firms offered 24 month service contracts around the same time.

From 2011 onwards there was a regulatory push in the opposite direction. In the EU, since October 2011, the Telecommunications Law banned three-year contracts and limited the maximum lock-in period to 24 months (European-Union, 2009). Similar trends followed in other regions. For example, in 2013, the Canadian Radio-television and Telecommunications Commission (CRTC) changed mobile phone subsidisation rules blocking companies from charging customers early cancellation frees on fixed-term contracts that exceed the value of a device subsidy and limited the length of the subsidy to 24 month (CRTC, 2013). More recently, in the US, President Obama signed the "Unlocking Consumer Choice and Wireless Freedom Act" that requires US carriers to unlock their devices when customers request which allows switching provider keeping the same handset (Congress, 2013).

Lock-in periods are a particular case of switching costs and they may hurt consumers surplus because they reduce the freedom to choose service providers (Klemperer, 1987, 1995). However, a number of complex dynamics makes the effect of switching costs on overall consumer surplus and welfare ambiguous (Dubé et al., 2009; Villas-Boas, 2015). For example, when firms cannot exploit existing consumers they are less likely to compete for them in the first place, which may increase the average price for all consumers (Cabral, 2009)[2].

The complex dynamics that switching costs introduce in market outcomes suggest that measuring the impact of switching costs on welfare is essentially an empirical question. Our study contributes to this line of research, and we provide an empirical analysis of switching costs in the

[2]When the Telecommunications Law was transposed into Belgium Law in August 2012, Belgacom, the telecommunications incumbent, acted as first mover and took a step further than what was dictated by the new regulatory landscape and started offering contracts as non-binding. After this move, there was a large increase in mobile communications churn in the overall Belgium market, firm margins (Earnings Before Interests, Depreciation and Amortizations) declined, and mobile handset subsidization stopped (Tefficient, 2013). So in the end, while services prices declined, it is unclear what was the welfare effect of the policy in the market.

context of the telecommunications Industry.

The contribution of our paper is to study, and measure, the welfare implications of shortening the lock-in period to less than the current status quo of 24 months. We focus on the market for triple-play services, which is now dominant way of consuming communication and media services both in the US and in the EU (OVUM, 2015).

Our paper uses a dataset from a large triple-play telecommunications provider to study what happens to consumers and to the firm when the lock-in period reduces. We use this dataset to fit a multinomial logit model in which households can choose to keep the same bundle, change service bundle inside the carrier or churn. This model allows us to measure the switching costs associated to changes of service inside the carrier and to churn in dollar terms. This model also allows us to simulate how changes in the length of the lock-in period affect both consumer surplus and profits. In our empirical context, we find that the average switching cost to change bundle inside the carrier is \$165 whereas the average switching cost associated to churn is roughly \$215 if lock-in free and more than \$218 when the contract is still active depending on how far away from lock-in expires.

We also find that the firm loses more profit than what consumers gain in surplus when the lock-in period reduces. This happens after we've accounted for the cross switching of subscribers between the service providers. Consequently, our work shows that regulators need to be very careful when considering shortening the length of lock-in periods. Shortening the lock-in period increases consumer surplus, however, it also decreases the profit of the firm, and regulators need to ponder these two opposing forces in a way that provides consumers flexibility but also ensure that operators have sufficient incentive to be in business and maintain, or even upgrade, the quality of the services provided.

The remainder of our paper is organized as follows. Section 2.2 reviews the related work. Section 2.3 describes our empirical context and provides descriptive statistics. Section 2.4 introduces our model and empirical strategy. Section 2.5 presents results of our econometric model and section 2.6 describe the results or our policy simulations. We conclude in section 2.7.

2.2 Literature Review

Our paper is mostly related with the empirical literature of switching costs measurement which is closely linked to the active monitoring of switching costs that regulatory authorities carry out. Regulatory authorities oversee switching costs and suggest legislation that Governments may limit them. This is a complex task because regulators have to deal with the trade-off between consumer surplus and market welfare, the latter defined as the sum of consumer surplus and firm profits. The regulator's task is not just one of fairly splitting welfare between consumers and firms but also one of looking for ways to maximize overall well-being (Gans, 2001).

Conventional wisdom suggest that low switching costs are likely to increase consumer surplus (Klemperer, 1995). However, and at the same time, low switching costs provide little incentive for firms to provide the service in the first place, which can reduces both consumer surplus and welfare (Farrell and Klemperer, 2007).

In general, regulators work to limiting switching costs because when they are substantial, market leaders are likely to enjoy a significant advantage that allows them to sustain a large market share (Lieberman and Montgomery, 1998; Bijwaard et al., 2008). With high switching costs entrants have a hard time to steal consumers that are locked-in to the market leader. High switching costs also allow firms to exploit consumers that are locked-in by charging them higher prices, a strategy called bargain-then-rip off (Klemperer, 1987, 1995) and a number of empirical studies found support for the theoretical argument (Sharpe, 1997; Shy, 2002; Stango, 2002; Viard, 2007).

However, low switching costs may also have an adverse effect for consumers (Cabral, 2009). For example, when firms cannot exploit existing consumers, they have fewer incentives to attract them in the first place (Dubé et al., 2009; Shin and Sudhir, 2008; Doganoglu, 2010), which may result in higher starting prices for consumers.

Additionally, in markets where service switching generates additional cost to the market, low switching costs, which will lead to more frequent switching from consumers, may render the market less efficient (Gans, 2001).

Although the theory of switching costs is rich in economic models, papers measuring them are scarcer. Exceptions include Borenstein (1991) measured the magnitude of switching cots in the US retail gasoline market, Knittel (1997) showed how the presence of significant switching costs led to little change in the prices of long distance phone calls in the US after the divestiture of AT&T in 1984, Viard (2007) studied the introduction of number portability for toll-free numbers in the US and found that switching costs had an ambiguous effect on prices for firms that could not discriminate between existing and new consumers, Epling (2002) studied competition in the long distance telephony in the US after the Telecom Act of 1996 and found that consumers subject to higher switching costs paid higher prices, Grzybowski (2008) found significant switching costs in the mobile sector in the UK after the turn of the century. Using discrete choice experiment, Confraria et al. (2017) found that in a similar European country consumers are willing to pay 1.3 Euro per month to reduce the lock-in period from 1 year to 6 months in their mobile plans. Closer to our work, Shcherbakov (2016) studied switching costs in the TV industry in the US between 1997 and 2006. He found that these amounted to $200 and $244 for cable and satellite systems, respectively. These estimates are remarkably close to the ones we find in our paper.

2.3 Empirical Context

We use a transactional dataset from large telecommunications triple-play provider (hereafter referred as TELCO) covering the period between April and October 2013. Our partner is a major provider of telecommunication services in the country we analyze where triple play service penetration is above 50%. In 2013, when we obtained our dataset, about 70% of TELCO customers subscribed to triple-play a service and the number has grown since then.

The triple play service TELCO provides includes TV, Internet and fixed telephony. For each household and each month, this dataset contains information on bundle subscriptions and prices charged. For each bundle offered by TELCO we have bundle-specific characteristics such as number of TV channels, the maximum Internet speed, premium features such access to Video-on-Demand and whether mobile service was included. We also obtained information on service bundles that were provided by competing providers in each household's geographical market during the period of analysis. Specifically, for each household of TELCO, we know how many service providers were offering similar bundles in the household's zip code and the lowest price available for the service that households subscribe to in the local geographic market they reside.

Finally, the dataset provides household level variables such as the service contract's detail (including information of lock-in), the number of months that the household has been using each service from TELCO, that is, the households' tenure with each service and aggregated information on the monthly usage of these services, such as Internet traffic (uploads and downloads) and the number of fixed-to-fixed phone calls.

We obtained a random sample of 100,000 triple play households of which we discarded 2,772 that did not have bundle plan information available and our results are genetared from the remaining 97,228 TELCO's households.

Table 2.1 summarizes the service bundles that TELCO comercialized at the time of our data collection. Differences across bundles include the number of TV channels offer, the internet speeds and the availability of advanced features for watching TV. For example, bundles marked premium offer advanced features such as video-recording and video-on-demand. One bundle offered mobile service. The table reports the average price charged for each bundle alonside its standard deviation. The same bundle may be charged different prices to different households depending, for example, on when each household signs up, the negotiation between household and firm, and marketing campaigns, etc.

TELCO offered non triple play packages during the period we study. We focus only on triple-play subscribers because this allows us to know the entire home-based telecommunication portfolio in which context we can define churn clearly.

Table 2.1 provides information on the lowest avaialble price for each bundle in the *best price* column. This was calculated from the competitive information avaialble in each household's zip code. During the period of analysis, the average number of services provided per local market was 2.88 and the median was 3.

During our period of analysis consumers could change bundle inside the same carrier or switch carriers. A consumer whose lock-in period ends will commit to a 12 months when changing bundle inside the carrier. A consumer that is more than 12 months away from lock-in expiry experiences

Table 2.1: Summary statistics on Triple-Play packages offered by our industrial partner

No.	Share	N.channels	Internet	Telephony	Premium	Mobile	Avg.Price	Sd.Price	Best Price
1	0.20	$\simeq$ 120	$\simeq$ 30mbps	Yes	No	No	57.62	13.42	52.59
2	0.13	$\geq$ 160	$\geq$100mbps	Yes	Yes	No	64.73	5.71	62.57
3	0.12	$\geq$ 160	$\geq$100mbps	Yes	Yes	No	71.65	12.36	67.75
4	0.11	$\geq$ 160	$\simeq$ 30mbps	Yes	Yes	No	57.57	5.77	53.63
5	0.07	$\geq$ 160	$\geq$100mbps	Yes	Yes	No	73.41	4.04	69.51
6	0.07	$\geq$ 160	$\geq$100mbps	Yes	Yes	No	76.78	10.67	71.65
7	0.05	$\simeq$ 120	$\simeq$ 10mbps	Yes	No	No	58.68	4.77	51.08
8	0.04	$\simeq$ 120	$\simeq$ 10mbps	Yes	No	No	54.31	2.54	53.36
9	0.04	$\simeq$ 150	$\simeq$ 10mbps	Yes	No	No	57.16	4.49	54.59
10	0.03	$\geq$ 160	$\geq$100mbps	Yes	Yes	No	73.01	12.10	70.50
11	0.03	$\simeq$ 150	$\geq$100mbps	Yes	No	No	54.98	11.46	52.92
12	0.02	$\simeq$ 30	$\simeq$ 10mbps	Yes	No	No	51.99	0.10	51.99
13	0.02	$\simeq$ 30	$\simeq$ 1mbps	Yes	No	No	45.93	1.23	45.48
14	0.02	$\geq$ 160	$\geq$100mbps	Yes	Yes	Yes	101.97	11.40	100.31
15	0.02	$\simeq$ 150	$\simeq$ 10mbps	Yes	No	No	55.82	6.78	53.36
16	0.02	$\simeq$ 120	$\simeq$ 30mbps	Yes	No	No	58.67	3.85	57.65

(a) All money values are in 2013 US Dollars. (b)*Premium* is a dummy variable indicating whether the product contains premium features. (c) *Best.Price* stands for the lowest price available in the local market for a bundle with similar features offered by other service providers.

no change in the lock-in duration if she changes bundle inside the carrier. A consumer who is less than 12 months away from lock-in expiry gets her lock-in period reset to 12 months if she changes bundle inside the carrier. New consumers face always a lock-in period of 24 months so if a consumer switches providers then a lock-in period of 24 months is enforced by the new provider.

Between April and October 2013 two types of service changes can occur in our data. Households can change bundle inside the carrier or churn. On average per month, around 1% of the consumers churn, 4% of them change bundle inside the carrier and 1% of the consumers are new to the provider.

Figure 2.1 shows the density of changes inside the carrier and churn as a function of time to lock-in expiry. The x-axis shows the number of months to lock-in expiry. Negative values indicate the number of months after the lock-in period expires. Rates of change in these figures are small within the first 12 months of a 24 month lock-in period. Otherwise, change happens when lock-in periods expire, in particular significant churn occurs around month 24. Sometimes, consumers churn when there is still 1 month to contract expiry because competitors cover this financial penalty to steal consumers (by offering up to three months of free service). Changes inside the carrier happen within the second part of 24-month lock-in periods. These changes set back the lock-in period to 12 months. A peak of changes within the carrier occurs at around 10 months into the lock-in period, which may be related to the firm's proactive marketing strategies that are, in part, aimed at ensuring that lock-in periods remain far from expiry. During our period of analysis 54% of the households in our sample were within a lock-in period. Furthermore, during our period of analysis, all major service providers that compete with TELCO adopted similar contract policies.

2.4 Empirical Approach

To study the impact of changing lockin on consumer surplus and firm profit we follow the approach laid out in figure 2.2.

First we use the dataset that TELCO provided to estimate consumer demand for the different product bundles that consumers can choose from at any point in time. We use discrete choice models to do so. Second, we use the results from the econometric model to simulate the churn

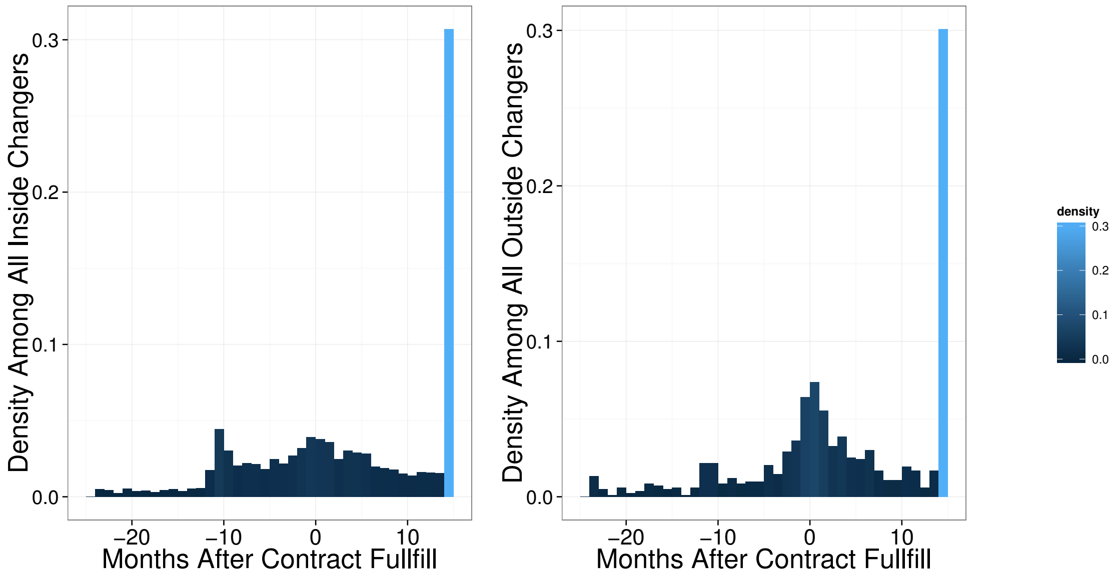


Figure 2.1: Density of changes inside the carrier and churn as a function of time to lock-in expiry

rates and consumer surplus as bundle characteristics change. Finally, firm profits are determined using the predicted market shares of different products, their prices and the rates of churn projected in our model that allow us to compute the revenue flows over the expected lifetime of consumers.

In the next sections we provide a detailed overview of how we model consumer behavior, firm profit and consumer surplus.

2.4.1 The Choice Model

We model household behavior using a multinomial logit model. Households choose among $J+2$ alternatives: J triple play bundles at TELCO (denoted as the 1^{st} to J^{th} option), an option to deduct service by choosing an non-triple-play service at TELCO [3] (denoted as the $(J+1)^{th}$ option) , or

[3]TELCO offered several non-triple play service packages during the observational period all of them included TV service. In addition to triple play bundles, during the period of analysis, households could choose to subscribe to ”TV-only”, ”TV+Internet”, and ”TV+Voice” bundles. For a current triple play service subscriber, he/she can choose to stay with the current subscription, switch to another triple play bundle, churn to a competitor, or switch to a non-triple play bundle without changing service provider. In our data, on average 0.9% triple play users choose to move to a non-triple play package with TELCO and we account for this possibility in our model. Triple play households who

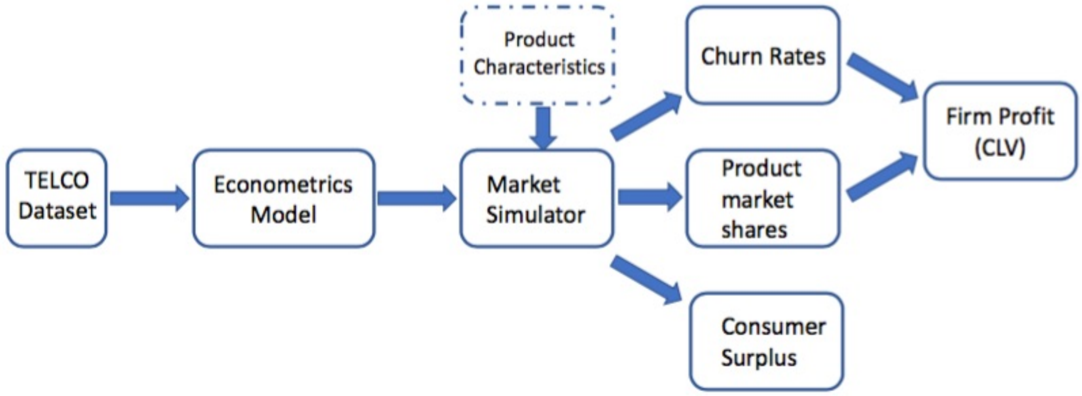


Figure 2.2: Flow chart of our simulation approach

churn (the $(J+2)^{th}$ option).

When a household churns we assume it does so to subscribe a similar triple play bundle from a competitor with lowest price. The prices offered by competitors are set as described in section 2.3. In this setting the utility of household h from choosing alternative j at time t, represented by u_{hj}^{t}, is given by:

$$u_{hj}^{t}(\mathbf{X_j}, p_{hj}^{t}, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t) = V_{hj}^{t}(\mathbf{X_j}, p_{hj}^{t}, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t) + \epsilon_{hj}^t \tag{2.1}$$

where $V(\cdot)$ represents the observable part of utility, which depends on a vector of bundle specific characteristics $\mathbf{X_j}$, the monthly bill p_{hj}^t, the household's choice of bundles up to the previous time period $\mathbf{a_h^{t-1}} = \{a_h^1, ..., a_h^{t-1}\}$ where a_h^τ represents the choice of household h at time τ, a vector of demographic time-varying characteristics $\mathbf{z_h^t}$, the remaining lock-in period l_h^t and the original length of the last lock-in period (potentially the current one if still active), L_{hj}^t. ϵ_{hj}^t represents the idiosyncratic error term, which we assume follows an i.i.d. Type I extreme value distribution. At any given time t, households had $J+2$ options to choose from. The probability that household h chooses alternative j at time t, is given by:

reduced the number of services were removed from our sample when they reduced their level of service.

$$P(a_h^t = j \mid \{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2}) = \frac{\exp u_{hj}^t(\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t)}{\sum_{k=1}^{J+2} \exp u_{hk}^t(\mathbf{X_h}, p_{hk}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hk}^t)} \quad (2.2)$$

We use a linear functional form for $V(\cdot)$ to estimate switching costs both to change bundle inside TELCO as well as to churn. We define:

$$\begin{aligned} V_{hj}^t(\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_h^t) =& \mathbf{X_j}\alpha - \beta p_{hj}^t + \mathbf{z_h^t}\mu_j - \gamma_1 C_{hj}^t I_j \\ &+ \gamma_2 C_{hj}^t O_j \mathbb{1}(l_h^t \leq 1) + \gamma_3 C_{hj}^t O_j \mathbb{1}(l_h^t > 1) + \gamma_4 C_{hj}^t O_j \mathbb{1}(l_h^t > 1) l_h^t \\ &+ \gamma_5 C_{hj}^t O_j \mathbb{1}(l_h^t \leq 1) L_{hj}^t + \gamma_6 C_{hj}^t O_j \mathbb{1}(l_h^t > 1) L_{hj}^t \\ &+ \gamma_7 C_{hj}^t O_j \mathbb{1}(l_h^t \leq 1) Tenure_h^t + \gamma_8 C_{hj}^t O_j \mathbb{1}(l_h^t > 1) Tenure_h^t \\ &+ \gamma_9 C_{hj}^t I_j N_h^t + \gamma_{10} C_{hj}^t O_j \mathbb{1}(l_h^t \leq 1) N_h^t + \gamma_{11} C_{hj}^t O_j \mathbb{1}(l_h^t > 1) N_h^t \end{aligned} \quad (2.3)$$

where $I_j = \mathbb{1}(j \neq "churn")$ and $O_j = \mathbb{1}(j = "churn")$ indicate whether alternative j is a bundle inside TELCO or churn, respectively. $C_{hj}^t = \mathbb{1}(a_h^{t-1} \neq j)$ indicates whether household h changes bundle at time t. $Tenure_h^t$ indicates the tenure of household h at time t with TELCO and $\mathbf{z_h^t}\mu_j$ represent interactions between household characteristics and dummies for each alternative. Additionally, N_h^t represents the number of competitors in the region that offers similar (competitive) services to h's current subscribed service. All coefficients in this expression have economic meaning and their ratio to β provide interpretations in dollar terms. Table 2.2 describes the different coefficients, their meaning and the signs that we expect to observe.

2.4.2 Churn rates, Consumer Surplus and Firm Profits

Churn Rates: are determined based on the market share of the churn alternative in the multinomial choice model. To simulate different chur rates one changes the characteristics of the bundles and then measures how the churn rate changes in response to such changes because.

Consumer Surplus: is determined based on the surplus of the representative household which is given by the utility of the best alternative in the multinomial choice, that is:

Table 2.2: Interpretation of coefficients for discrete choice models

Label	Interpretation	Hypothesis
β	Effect of price on product utility	Negative sign. Utility reduces with price.
γ_1	Switching cost associated to changing bundle inside TELCO.	Negative sign. Switching costs negatively affect utility.
γ_2	Switching cost associated to churn when there is at most 1 month to the end of the current lock-in period	Negative sign. Same reason as above.
γ_3	Switching cost associated to churn when there is more than 1 month to the end of the current lock-in period	Negative sign. Same reason as above.
γ_4	Change in γ_3 with one more month of lock-in remaining	Negative sign (same as γ_3). The more month lock-in remaining in the contract, the higher the switching cost.
γ_5	How γ_2 changes with the original length of the current lock-in period	Positive sign. Consumers may get tired of a longer contract and be more likely to churn when the contract has expired
γ_6	How γ_3 changes with the original length of the current lock-in period	No hypothesis made.
γ_7	How γ_2 changes with household tenure	Negative sign. Time filtered out loyal customers. Users with longer tenure the less likely to churn.
γ_8	How γ_3 changes with household tenure	Negative sign. Time filtered out loyal customers. Users with longer tenure the less likely to churn.
γ_9	How γ_1 changes with one more competitor in the local market	Positive sign. More intensive competition likely leads to less switching costs.
γ_{10}	How γ_2 changes with one more competitor in the local market	Positive sign. More intensive competition likely leads to less switching costs.
γ_{11}	How γ_3 changes with one more competitor in the local market	Positive sign. More intensive competition likely leads to less switching costs.

$$CS_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2}) = \frac{1}{\beta} \max_{j=1,...J+2} \{u_{hj}^t(\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t)\} \tag{2.4}$$

$$E[CS_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2})] \approx \frac{1}{\beta} \ln(\sum_{j=1}^{J+2} \exp(V_{hj}^t(\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t))) + C \tag{2.5}$$

where the approximation for the expected value is obtained from integrating over the distribution of the error term and C is an unknown constant that is irrelevant for comparison purposes and therefore usually ignored for policy analysis (Train, 2009). The expected cumulative surplus of the representative household is therefore given by

$$E[CS_h] = \sum_{t=0}^{\infty} \frac{E[CS_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2})]}{(1+\delta)^t} A_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2}) \tag{2.6}$$

$$A_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_j^t, L_{hj}^t\}_{j=1}^{J+2}) = \prod_{\tau=0}^{t-1} 1 - P(a_h^t = "churn" \mid \{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2}) \tag{2.7}$$

where $A_h^t(\cdot)$ represents the survival probability of household h at time t. δ represents the household monthly discount rate.

Firm profits: are measured using the representative household's lifetime value to the firm, which is a function of the discounted net future earnings and costs. Earnings include the monthly payments for the service and contract breaching penalties that customer pay when contracts are terminated before expiry.

The firm's expected earnings for a household h in a certain month t denoted by $\Pi_{h,t}^{MP}$, is determined by the following equation:

$$\Pi_{h,t}^{MP} = \sum_{k=1}^{J} s_k^t(p_{hk}^t - c_{hk}^t) A_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2}) \tag{2.8}$$

where $s_k^t(a_h^t) = P(a_h^t = k)$ represent the (within TELCO) market share of alternative k and c_{hk}^t represents the marginal operational cost of alternative k for h. r denotes the firm's rate of return (RoR). Expected earnings from the financial penalties at month t, represented by $\Pi_{h,t}^{Pen}$, are determined by:

$$\Pi_{h,t}^{Pen} = \sum_{k=1}^{J} s_k^{t-1} \mathbb{1}(L_{hj}^t - t \geq 1) p_{hk}^{t-1}(L_h^t - t) B_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^J) \tag{2.9}$$

where $B_h^t(\cdot)$ is the churn probability of household h at time t

$$\begin{aligned} B_h^t(\{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2}) =& P(a_h^t = \text{"}churn\text{"} \mid \{\mathbf{X_j}, p_{hj}^t, \mathbf{a_h^{t-1}}, \mathbf{z_h^t}, l_h^t, L_{hj}^t\}_{j=1}^{J+2}) \\ & \times \prod_{\tau=0}^{t-1} 1 - P(a_h^\tau = \text{"}churn\text{"} \mid \{\mathbf{X_j}, p_{hj}^\tau, \mathbf{a_h^{\tau-1}}, \mathbf{z_h^\tau}, l_h^\tau, L_{hj}^\tau\}_{j=1}^{J+2}) \end{aligned} \tag{2.10}$$

The expected earnings of the firm from an average household h, represented by Π_h is defined as:

$$\Pi_h^t = \Pi_{h,t}^{MP} + \Pi_{h,t}^{Pen} \tag{2.11}$$

Finally, in addition to network maintenance and operation costs per household, telecommunication providers pay non-negligible initial costs of consumer acquisition and setup (Farrell and Klemperer, 2007). These costs occur at the beginning of subscriptions in the form of equipment such as internet modems and set-top-box devices, network deployment in the customer house or apartment, and a specialized technician visit, etc.

Let AC_h represent the acquisition costs associated with offering services to household h, and r be the firm's monthly weighted average cost of capital[4], the expected lifetime profit from household h is given by

[4]Sometines denoted by the firm's Expected Rate of Return, or discount rate.

$$\Pi_h = \sum_{t=0}^{\infty} \frac{\Pi_h^t}{(1+r)^t} - AC_h$$
$$= \Pi_h^{MP} + \Pi_h^{Pen} - AC_h$$

(2.12)

2.5 Econometric Estimation

In this section we summarize the estimation results of the multinomial logit model described in the previous section. Table 2.3 shows the results obtained. Column (1) corresponds to Eq (2.3) while columns (2) and (3) are presented as robustness checks.

The negative coefficients γ_1, γ_2 and γ_3 indicate that switching costs reduce the probability of switching and the ratio of these coefficients to the coefficient on price allows interpreting switching costs in dollar terms.

Looking at our main model provided in column (1) we see that in a market with at least two service providers (the focal firm and one competitor), if the lockin period is over, the switching cost to an external firm (churn) is $210.1 ((-8.62+0.215)/-0.04) and reduces by $5.4 (0.215/-0.04) per additional competitor in the market.

In the last moth of their lockin period, the average switching cost to churn is $212.6 ((-8.73+0.227)/-0.04) which reduces by $5.7 (0.227/-0.04) per addtional competitor. This number increases by $0.95 (-0.037/-0.040) per additional month of contract outstanding.

The average switching cost to change bundles inside the carrier in a market with at least two service providers (the focal firm and one competitor) is $161.9 ((-6.60+0.122)/-0.04). In this case, one more competitor in the local market reduces within provider switching costs by $3.1 (0.122/-0.04).

In column (2) we interact household demographic characteristics with product dummies to control for potential specific demographic effects. The demographic variables available to us are the (standardized) age of the household member paying the bill, and the household's (standardized) intensity of usage for Internet and voice services respectively. All estimates are qualitatively and quantitatively similar to those shown in column (1).

Finally, column (3) provides a model-based check for the assumption of independence of irrelevant alternatives (IIA) that is implicit in the multinomial logit model (Keane, 1992). We compare the estimation results of the multinomial logit to those of the mixed logit model which does not impose IIA (Cheng and Long, 2007). Results show that the mean effects in mixed logit model are qualitatively and quantitatively consistent with that of the multinomial logit model and provice strong evidence that the simplest model is unlikely to violate the IIA assumption. Since model in column (1) is computationally tractable and model (3) is not, we rely on the former for our policy simulations in section 2.6.

2.6 Policy Simulations

We use our market simulator to study how shortening lock-in periods changes market outcomes. To make the metrics comparable, we measure the expected firm profits and consumer welfare in a representative household's lifetime. We discuss how these metrics change relative to the status quo of 24 months as lock-in period lengths shorten.

We make several simplifying assumptions that allow us to study how consumer surplus changes, firm profits change and market level welfare changes in response to exogenous changes in the lockin period firms can impose to consumers.

First, we assume that changes to the length of the lock-in period are applied to all bundles and to every operator in the market.

Second, we assume that the overall size of the market is fixed - reducing lockin will not change the total number of households that pay for triple play. We believe this is realistic because the penetration of private household-based telecom service is close to 90% in the country we analyze.

Third, consistent with the previous assumption we assume that consumer switching between the firms is symmetric, firms that churn from one firm will chose a competitor and will not drop out from the market.

Fourth, we assume that firm market shares will remain constant.

Table 2.3: Estimation results of discrete choice models

	(1)	*(2)*	*Mixed Logit*	
	Mean Effect	Mean Effect	Mean Effect	Heterogeneity
Price (β)	-0.040^{***}	-0.039^{***}	-0.048^{***}	
	(0.001)	(0.001)	(0.001)	
Change Inside(γ_1)	-6.600^{***}	-6.611^{***}	-7.411^{***}	3.186 ***
	(0.057)	(0.020)	(0.052)	(0.306)
Change Outside $\times$ Contract Free(γ_2)	-8.616^{***}	-8.778^{***}	-9.646^{***}	3.193 ***
	(0.210)	(0.212)	(0.322)	(0.153)
Change Outside $\times$ Contract Active(γ_3)	-8.729^{***}	-8.933^{***}	-9.805^{***}	2.012^{***}
	(0.383)	(0.386)	(0.459)	(0.239)
Change Inside $\times$ N Competitors(γ_9)	0.122^{***}	0.122^{***}	0.130^{***}	
	(0.018)	(0.018)	(0.030)	
Change Outside $\times$ Contract Free $\times$	0.215^{***}	0.226^{***}	0.358^{***}	
N Competitors (γ_{10})	(0.045)	(0.045)	(0.048)	
Change Outside $\times$ Contract Active $\times$	0.227^{***}	0.243^{***}	0.341	
N Competitors (γ_{11})	(0.089)	(0.090)	(0.093)	
Change Outside $\times$ Contract Active $\times$	-0.037^{***}	-0.037^{***}	-0.036^{***}	
Month-to-contract-expiry(γ_4)	(0.016)	(0.016)	(0.016)	
Change Outside $\times$ Contract Free $\times$	0.023^{**}	0.023^{**}	0.042^{**}	
Length-previous-contract(γ_5)	(0.005)	(0.005)	(0.006)	
Change Outside $\times$ Contract Active $\times$	0.010	0.010	0.013	
Length-previous-contract(γ_6)	(0.008)	(0.009)	(0.008)	
Change Outside $\times$ Contract Free $\times$ Tenure(γ_7)	-0.006^{***}	-0.005^{***}	-0.010^{***}	
	(0.001)	(0.001)	(0.001)	
Change Outside $\times$ Contract Active $\times$ Tenure(γ_8)	-0.010^{***}	-0.009^{***}	-0.010^{***}	
	(0.002)	(0.002)	(0.002)	
Number of channels	0.002^{***}	0.003^{***}	0.005^{***}	0.000
	(0.000)	(0.000)	(0.000)	(0.003)
Internet speed	-0.004^{***}	-0.004^{***}	-0.005^{***}	0.000
	(0.001)	(0.001)	(0.001)	(0.002)
Premium features	1.868^{***}	1.868^{***}	2.176^{***}	0.012
	(0.042)	(0.043)	(0.060)	(0.149)
Mobile	1.424^{***}	1.422^{***}	3.118^{***}	0.076
	(0.057)	(0.059)	(0.064)	(0.122)
Demographic Controls (μz)	No	Yes	No	
Observations ($97,228$ $Households$)	$535,656$	$535,656$	$535,656$	
Log-Likelihood	$-179,386$	$-177,927$	$-176,955$	
Mcfadden R^2	0.900	0.901	0.901	

Note: $\cdot$p<0.1; *p<0.5; **p<0.01; ***p<0.001

Standard errors were robust clustered within households

Finally, we estimate changes in the expected acquisition costs using data provided in the annual report of TELCO which states that initial service deployment of a new households cost on average $390 (in 2013 USD) and includes the equipment to install in consumer premises and the trip of the technical team to the customer house/apartment to install and activate the service.

Beyond the outcomes of the simulations that we present below, appendix A.1.1 summarizes the changes in churn rate that we obtain when we use the multinomial choice model to simulate the market as the lockin period changes.

2.6.1 Unchanged Prices

Our first simulation assumes that bundle prices remain the same when lockin periods reduce. We use Eq.2.6 and Eq. 2.12 to estimate consumer surplus and the lifetime profit of a household.

For each existing consumer, firm profits decline when lockin periods reduce. The acquisition costs in equation 2.12 are zero and are not affected by the policy change, but the earnings, Π_h^{MP} and Π_h^{Pen}, reduce because (a) households are more likely to churn and (b) breaching costs reduce as lock-in length shortens. Figure 2.3 shows how these components of firm profit, i.e. Π_h^{MP} (top-left) and Π_h^{Pen} (top-right), and consumer surplus (bottom-right) change if the lock-in period reduces from the status quo of 24 months.

The plot shows firm profits for 3 different levels of the yearly Rate of Return (RoR) set around the typical rates experienced in the telecommunication sector (9%, 11% and 13% according to Damodaran (2015)). For consumer surplus we surveyed the literature to determine how much households discount the future in their financial decisions and we selected discount rates that frange from 0.075 to 3.76 [5].

The simulation shows that with constant prices, shortening the lock-in period leads to larger

[5] Empirical dynamic model literature usually uses 0.995 as a weekly discount factor, corresponding to 0.98 monthly and 0.77 yearly (Erdem and Keane, 1996; Sun, 2005). Alternatively, another reasonable way to determine the consumer discount factor is to back induct from annual interest rate. This number usually ranges from 1% to 8% (World-Bank, 2013). Recently, using Chinese cellphone data, Yao et al. (2012) suggested the discount factor may be way below the above numbers ($0.86 \sim 0.91$ weekly). In our sensitivity analysis, we used a fairly wide range of yearly consumer discount rate, from 0.075 to 3.76, corresponding to annual discount factor ranging from 0.21 to 0.93.

reductions in firm's earnings ($bottom\text{-}left$) than the gains in consumer surplus ($bottom\text{-}right$). For example, in a market where four different service providers offer similar services ($blue$ curves), if lock-in periods reduce to 16 months, the present value of expected lifetime consumer surplus increases $\$2 \sim \$22$ (blue-round curves in $bottom\text{-}right$ plot) depending on the consumer discount rates, whereas the present value of the expected earnings that the firm enjoys reduces more than \$50. This loss in earnings corresponds to 1.5% of the expected present value of the profits the firm would get when the lock-in period is 24 months. Firm profit results are qualitatively similar across the rates of return we simulate. As for consumer surplus, higher discount factors ("squares") are associated with smaller changes in consumer surplus as lock-in length shortens because the benefits a consumer enjoys in the future are substantially discounted.

Competition also affects the results. In more competitive local market, consumers encounter more alternatives and are more likely to churn to different service providers and both firm profits and consumer surplus are more sensitive to contract length reduction.

As detailed in the preamble of this section, we assume that consumers that leave one firm do so because they swap into a competing provider. Together with our assumption of stable market shares, firm swapping implies that the earnings loss due to churn will be compensated by new consumers joining. However, these changes generate service initiation costs that firms will have to pay, that would not occur with less frequent switching behavior.

Let Π_{fh} represent the profit of firm f gained from an arbitrary representative subscribing household hs lifetime, which includes the expected monthly payments Π_{fh}^{MP}, expected breaching penalties from when contracts are terminated before expiry Π_{fh}^{Pen}, and the marginal costs associated to consumer acquisition C_{fh} that firms pay for every new customer that they acquire.

$$\Pi_{fh} = \Pi_{fh}^{MP} + \Pi_{fh}^{Pen} - C_{fh} \tag{2.13}$$

A change in the lock in period associated to triple-play services induces a change in this firms profit determined by:

$$\Delta\Pi_{fh} = \Delta\Pi_{fh}^{MP} + \Delta\Pi_{fh}^{Pen} - \Delta C_{fh} \tag{2.14}$$

When the lock in period reduces, a household is more likely to churn (earlier) in her lifetime,

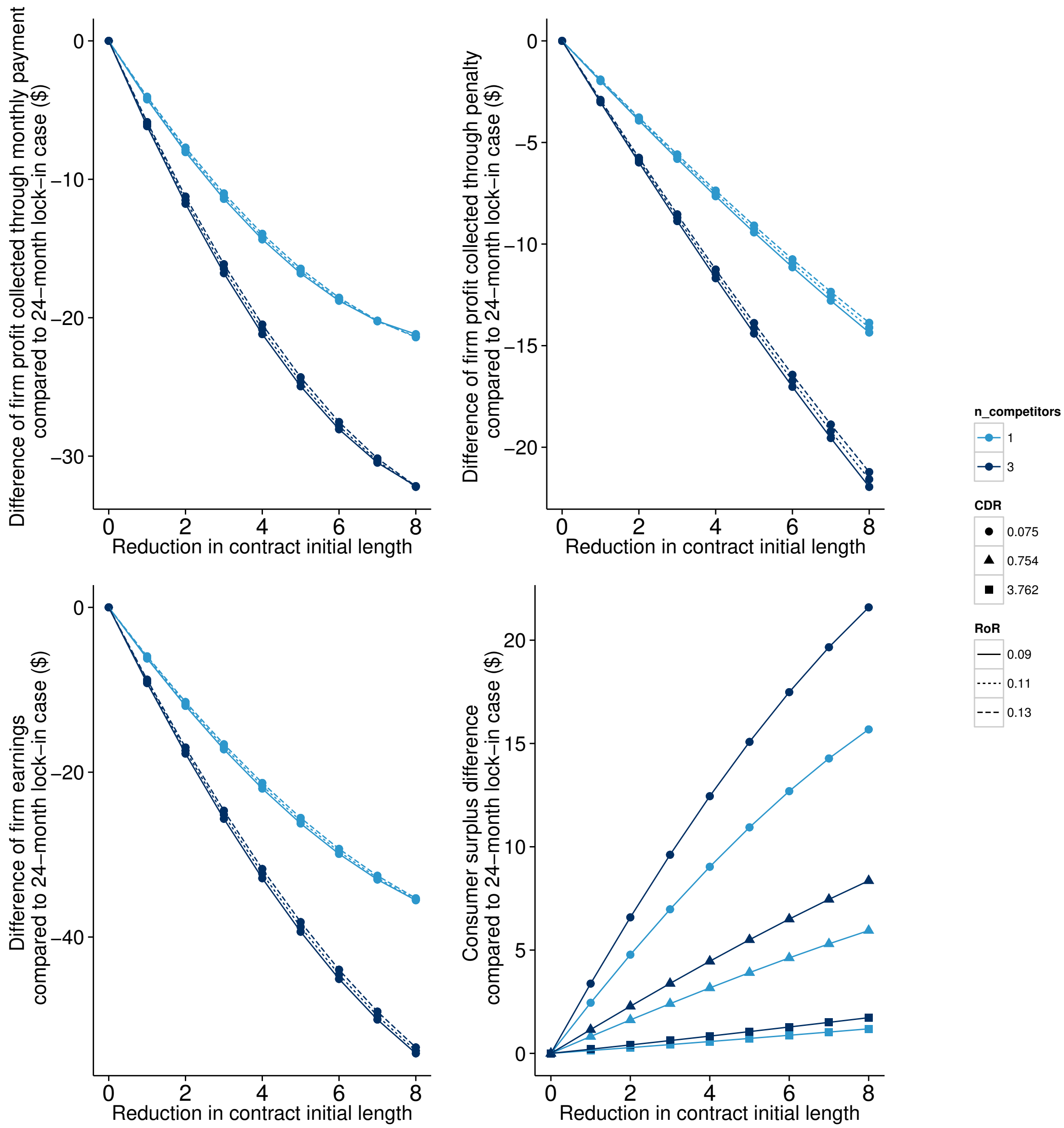


Figure 2.3: Difference in firm profits collected from an existing subscriber's monthly payment ($top - left$), breaching penalties($top - right$), total earnings ($bottom - left$) and in consumer surplus ($bottom - right$) in dollars when the lock-in period reduces from 24 months. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using the estimation results in column (1) of Table 2.3.

leading to negative $\Delta\Pi_{fh}^{MP} < 0$ and negative $\Delta\Pi_{fh}^{Pen} < 0$. The expected change in consumer acquisition cost, ΔC_{fh}, is zero because C_{fh} was burned only at the beginning of the service provision.

Without loss of generality, assume that if household h churns from firm f, she signs up with firm f'. Then firm f' starts to gain profit collected through monthly payment from household h. Meanwhile, due to service switching, firm f' also needs to pay the corresponding acquisition cost $C_{f'h}$. These are the expected acquisition costs that firm f' pays for initiating service to h and they include the equipment to install in consumer premises (internet modems and set-top-box devices) and the trip of the technical team to the customer house/apartment to install and activate the service.

From the market perspective, when lock in period reduces, the part of firm profits from household monthly payments are roughly unchanged given the assumption of stable market (it only transferred from f to f'). However, the more frequent (and earlier) switching of household resulted in higher expected acquisition costs for firms, $\Delta C_{f'h} > 0$. Under these assumptions, the changes to profits for the firms side in the market from a representative household consist of the following two parts:

$$\Delta\Pi_{f,f'} = \Delta\Pi_{fh}^{Pen} - \Delta C_{f'h} \tag{2.15}$$

We plot the corresponding changes in profits for the firms side and consumer surplus in figure 2.4.

The left panel shows how firm profit changes if the lock-in period reduces. The results show that even when taking business stealing into consideration, firms still lose more than consumers gains because of the additional acquisition costs they have to pay. In our simulations, in a market with three competitiors, a firm may lose as much as $40 per household lifetime if lock-in length was reduced to 16 months from the status quo 24 months.

The right panel of Figure 2.4 shows the market welfare (over customer lifetime), defined as the sum of firm profit and consumer welfare, i.e. $\Delta\Pi + \Delta CS_h$. Welfare decreases as lock-in length shortens. The loss in firm profits supersedes the gains in consumer surplus, resulting in a decrease of market welfare.

In a three-competitor (four providers) market, when consumers discount the future less (corresponding to higher consumer welfare gain, curves with ”round” dots), the market will lose about \$20 per household lifetime. The market welfare loss is higher if consumers have a high discount rate (curves with ”triangle” and ”square” dots).

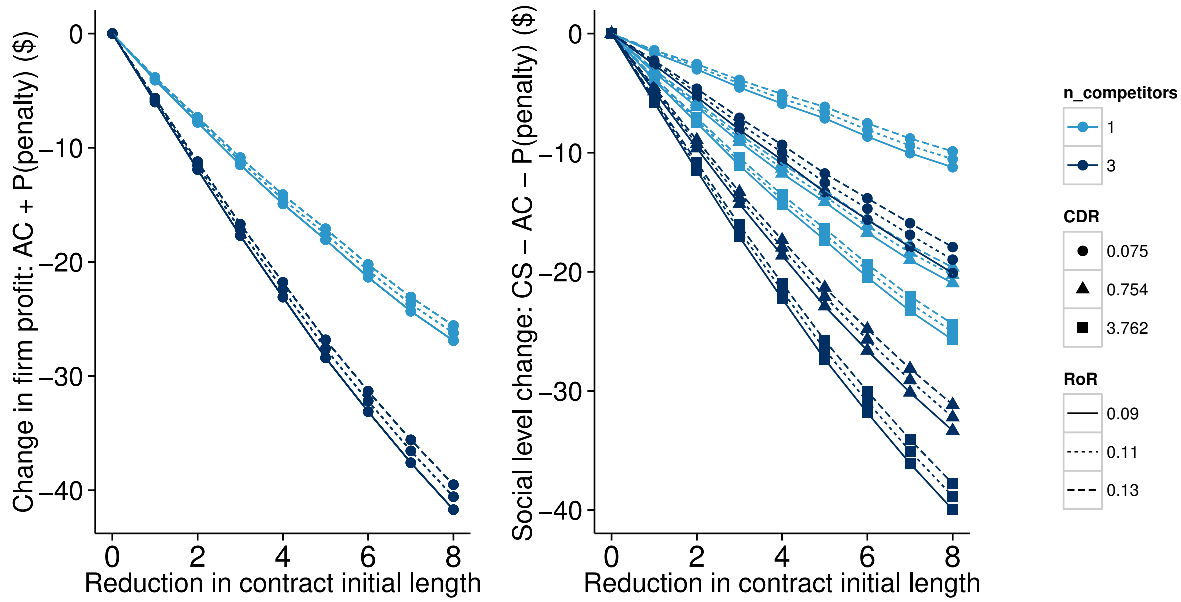


Figure 2.4: Difference in firm profits ($left$) defined as Eq (2.15) and in market welfare ($right$) in dollars when the lock-in period reduces from 24 months. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using the estimation results in column (1) of Table 2.3.

2.6.2 Increased Prices

Our second set of simulations tests what happens if all firms increase prices to compensate for the loss in profit due to reduced lock-in periods enforced by regulatory authorities, a phenomenon similar to the ”water-bed” effect in telephony industry (Genakos and Valletti, 2011).

This can happen because triple play markets are composed of few firms who often form oligopolistic markets being a key characteristics of such markets the presence of strategic interactions among firms [6].

In this section, we make the additional assumption that firms increase prices to maintain the same level profitability (RoR) before the change in lock-in periods. An additional assumption we make in this section is that changes in prices are similar, in percentage terms, for all products offered by TELCO. The results of our analysis are depicted in Figure 2.5.

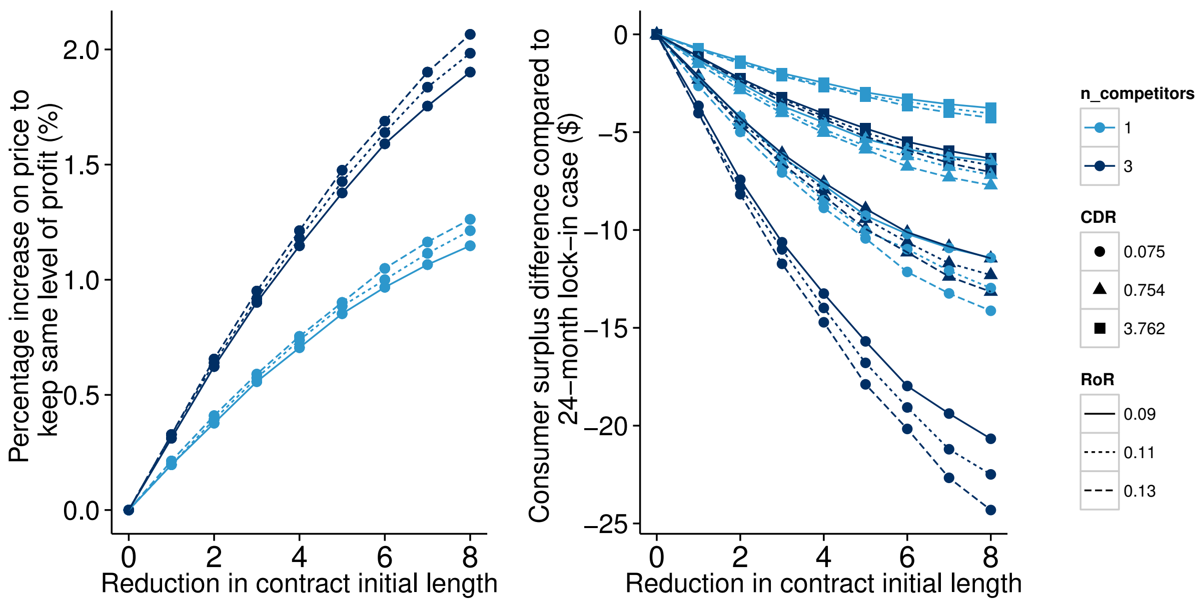


Figure 2.5: Percentage change in price ($left$) and in change consumer surplus ($right$) when the lock-in period reduces and the firm reacts to keep the same level of profit, comparing to the status quo of 24-month lock-in periods. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using estimation results in column (1) of Table 2.3.

For example, in a three-competitor market, to keep the same profit the firm increases price by roughly 1.5% to counter a reduction in the lock-in period of 8 months. The right plot in Figure

[6] If there are few providers in the market, the elasticity of market demand depends on the output of other firms. Firms therefore have incentives to coordinate and signal their behavior to others.

2.5 shows what happens to consumers (and the market) if the firm decides to increase prices when lock-in periods reduce to keep the same profit. As a result of the increased price, consumers lose around $\$4 \sim \$15$ in surplus depending on the discount rate and therefore they would have been better off if the regulatory authority did not reduce the lock-in period in the first place without preventing firms from increasing prices to counter the effect of the policy on their profits. These results show how reductions in the lock-in period must be paired with price regulation, otherwise firms are likely to increase prices to counter the losses in profits arising from lock-in reductions leaving consumers even worse.

2.6.3 Reduced Prices

Another possibility is that prices decrease in response to a more competitive environment. Our third set of simulations study how the decrease in price may affect firm overall profit, consumer surplus, and market welfare, respectively.

Figure 2.6 shows an example in which lock-in period lengths reduce from 24 months to 16 months. As before, we assume that price changes apply to every player and every product in the market. We also assume that the discount applied to all products is the same (in percentage terms).

When the prices decrease by 1% to 5%, firms lose profit and consumers gain welfare. The magnitude of the change depends on firm RoR and on the consumer discount rates (CDR).

When firm RoR and consumer discount rate are comparable (for example when ”CDR=0.075”), price reduction leads to similar changes in magnitudes for firm profit and consumer welfare, leaving the market welfare relatively stable. However, when firm discount the future less than consumers, which the literateure has established to be more likely the case in the telecommunication sector (Yao et al., 2012), firms are more sensitive to the price drop. Therefore when market prices decrease in response to lock-in reduction policy, market welfare worsens in addition to the effect of lock-in reduction, albeit consumers are beneficiary.

In addition to the results of figure 2.6 we provide simulations of price decreases in appendix that span different discount rates, and changes of the length of the lockin period in appendix A.

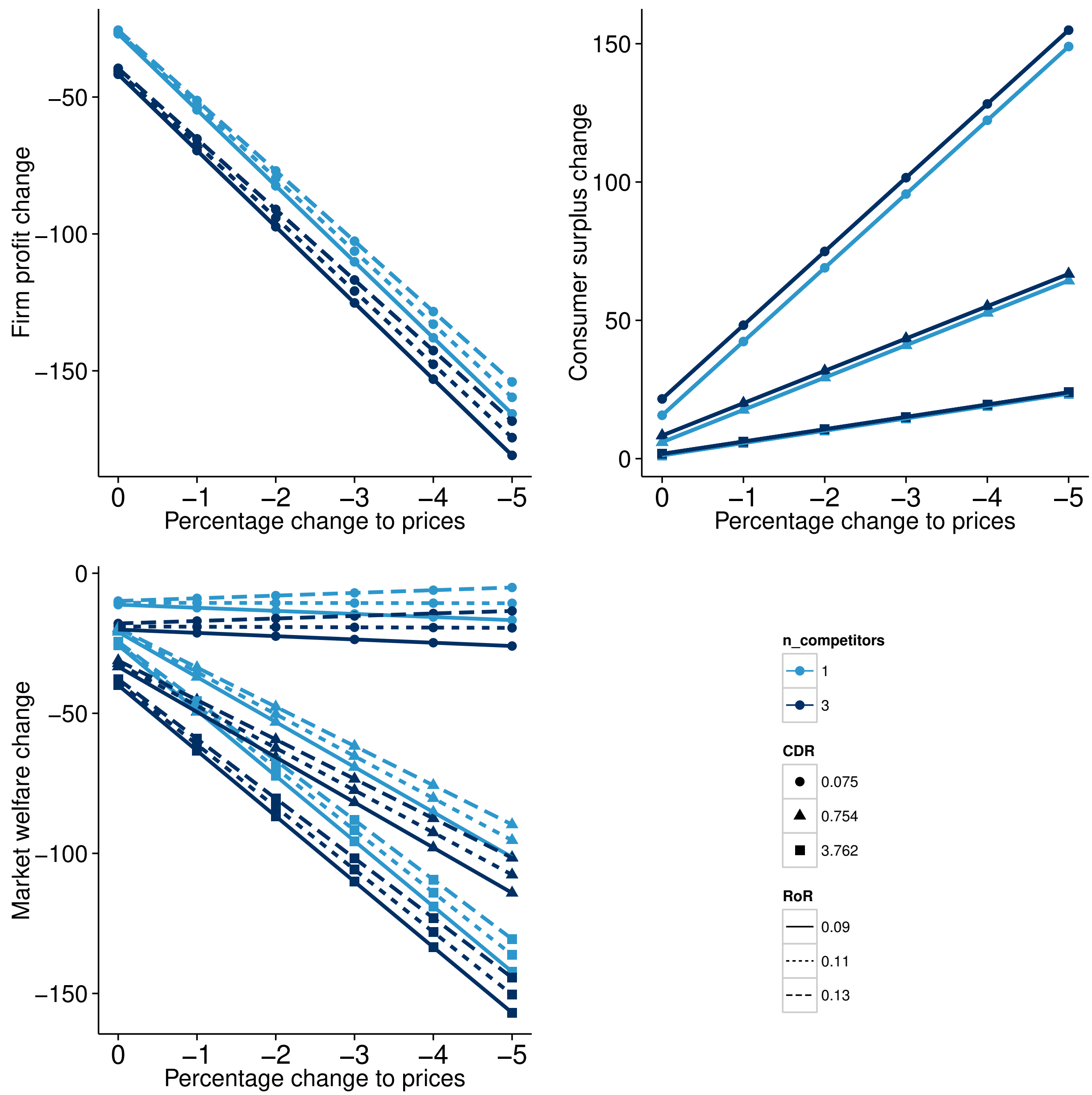


Figure 2.6: Effect of price decrease (in percentages) on firm overall profit (left column), consumer surplus (middle column), and market welfare (right columns) when lock-in was reduced from 24 months to 16 months. The numbers are comparing to status quo 24 months with unchanged prices. Plots are based on simulations using estimation results in column (1) of Table 2.3. A set of contour plots showing how prices and lock-in reduction may jointly affect the above metrics can be find in Figure 2.7 in appendix.

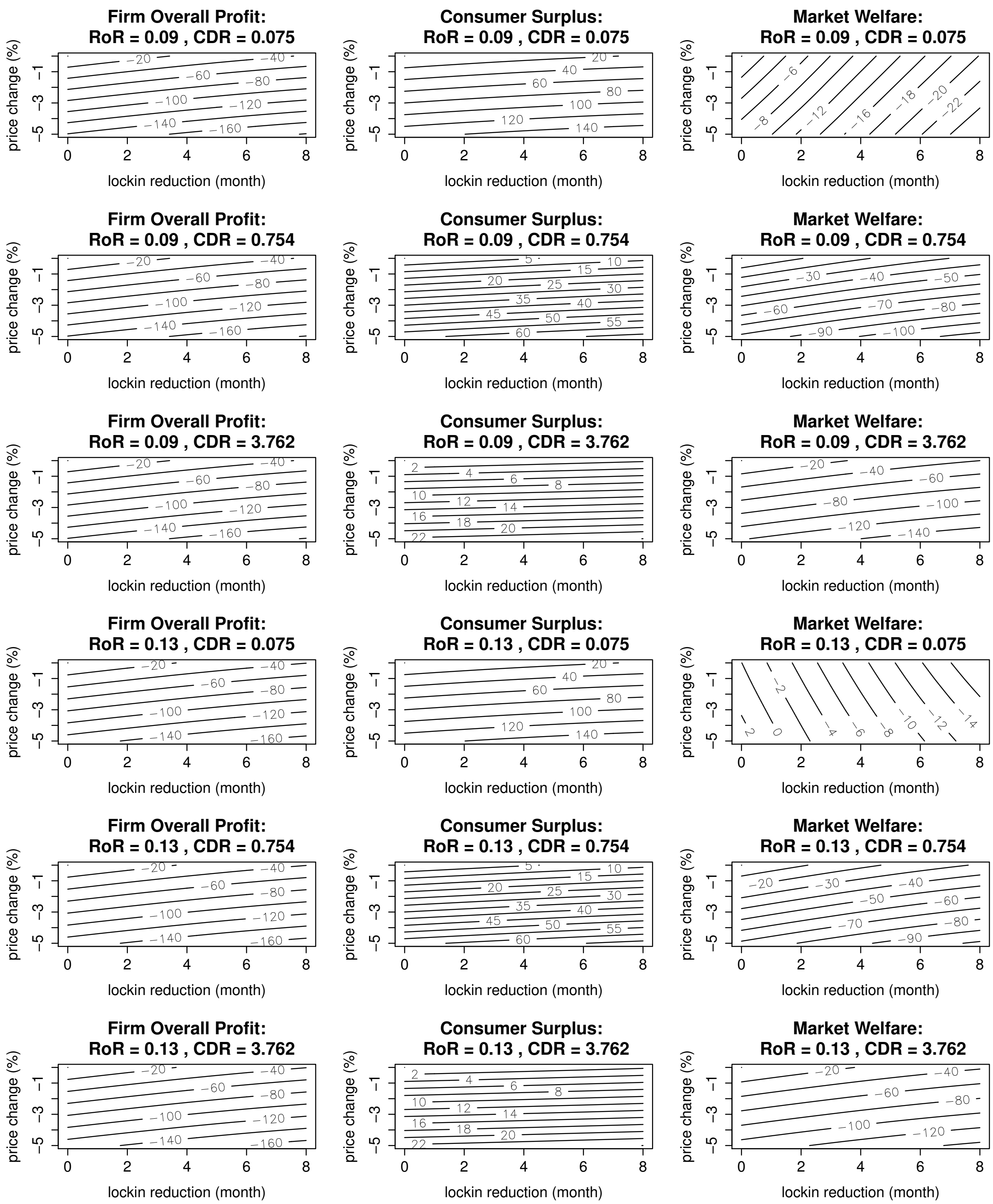


Figure 2.7: Contour plots of joint effect of price decrease (in percentages) and lock-in reduction (in months) on firm overall profit (left column), consumer surplus (middle column), and market welfare (right columns) in a four-provider market. Combinations between two firm RoRs and three consumer discount rates were shown. Plots are based on simulations using estimation results in column (1) of Table 2.3.

2.7 Conclusions

Lock-in periods in telecommunications services are a common practice employed by telecommunication providers to reduce the risk of failing to cover the significant capital costs needed to build the network in the first place and upgrade it over time. In short, operational revenues need to cover all operational costs and all investments in network upgrades as well as the initial cost with the network. The current practice is to lock-in consumers for periods of 24 months, which reduces the uncertainty for the firm. Consumers pay financial penalties if they breach contracts while lock-in periods are still active. Telecommunication firms have been operating with these type of contracts for a number of years now.

Telecommunication regulators have been studying the effect of lock-in periods on consumer welfare. Lock-in periods are a particular case of switching costs and they may hurt consumers surplus because they may reduce the consumers' freedom to change telecommunications provider. Researchers have shown that the effect of switching costs on overall consumer surplus and welfare is ambiguous. Therefore, regulators are mostly interested in empirically learning the impact of reducing the lock-in period on consumer welfare and, necessarily, on firm profit too, in order to anticipate how firms are likely to respond to potential changes in the current policy.

Our paper uses a dataset from a large triple-play telecommunications provider to study what happens to consumers and to the firm when the lock-in period reduces from the current status quo of 24 months. We find that the firm loses more profit than what consumers gain in surplus when the lock-in period reduces. This happens even when we accounted for the cross exchange of subscribers. This arises because reduced lock in increases churn and every time a consumer switches provider there is a new (non-trivial) cost to set her up at the new provider. More specifically, the welfare generated by one consumer to the market is given by the difference between: i) her willingness to pay for service and, ii) the cost incurred to provide her with it.

With respect to i) the willingness to pay for service is unlikely to change with reductions in lock in periods. Nowadays, when a consumer terminates triple-play service at one provider she usually signs up for a similar service from another provider in the same market. The value that she associates to the service that she obtains from the new provider is, therefore, similar to the value that she associates to the service that she used to obtain from the older provider. The reason why she churns is, most likely, because she can obtain service from the latter provider at a cheaper price. As for ii), the operational cost to provide triple-play service is also unlikely to change due to a reduction in the lock in period. This cost is essentially determined by the technology used to provide service. Therefore, a consumer that takes advantage of a reduction in the lock in period to switch from one provider to another is likely to generate the same welfare at the new provider, expect for the fact that setting up the consumer at the new provider involves a cost, namely to deploy new equipment at the consumers premise, such as a new set-top-box and a new Internet modem, to wire and cable cabling as well as technical onsite support. As a consequence, welfare reduces with reductions in the lock in period because consumers take advantage of the new policy and churn more than they otherwise would triggering additional setup costs.

This study shows that reducing lock in periods may be insufficient to improve consumer well-being. Such a policy needs to be paired with price controls to further protect consumers. Our analyses don't assert that regulators should not reduce lock in periods. In fact, doing so increases consumer surplus if no other market variable changes. However, regulators need to ponder between such policy and the additional costs that firms incur to set up the consumers that churn. Firms are likely to increase prices to cover such costs, which they are likely to pass on to consumers. This may end up hurting consumers more than benefiting them. Regulators need to consider policies in tandem to achieve the desired objectives. We believe that this insight, in the particular case of regulating lock in periods, is new to the literature and may be of great value to telecom regulators now that a number of them are looking at changing lock in policies.

Our paper has several limitations. First, we limit the discussion of lock-in length in a fairly short range. We are conservative in generalizing the trend to the cases where lock-in period exceeds 24 months because of data limitations. In our data we only observe contracts that fell in the range between 12 months and 24 months. The variation in lock-in lengths allows us to estimate the effect of lock-in length on consumer utilities and welfare. However, the lack of observation on lock-in lengths more than 24 months limits our ability to understand what would happen when lock-ins were instead increased. Moreover, we do not model a fully dynamic and competitive environment. We study how firms react to the regulators change in policy together by collaboratively raising prices but we do not take into account how firms may react differently in a competitive environment. Future research can study various equilibria in different stages of a collusion game. Finally, although our estimation accounted for customized service prices for households, we do not observe situations in which TELCO offered discounts to households who then still decided to churn. Because we fail to see these cases, switching costs may be overestimated in our paper. However, with smaller switching cost the simulated market would generate higher rates of churns (resulting in higher acquisition cost for firms) and less salient improvement for consumer welfare when lock-in shortening policy takes place. This would not contradict the main conclusion of the paper.

Chapter 3

The Interplay of Information from Friends Versus Crowds in the Consumer Shopping Journey: An Observational Study

3.1 Introduction

Most consumers search for product information prior to making purchasing decisions. They read product descriptions, seek opinions from other consumers, observe what others buy, and in the meantime adjust their preference beliefs to products (Branco et al., 2012). Information search is beneficial to consumers as it resolves the uncertainty related to product utility (Feinberg and Huber, 1996). Because acquiring product information is costly, a consumer can not inspect all options. The search process stops when it reaches a point where the marginal benefit of searching for additional product information does not exceed the cost of search (Stigler, 1961; Weitzman, 1979). She then evaluates the products in her consideration set and chooses the one with the highest expected utility (Wang and Sahin, 2017). The final consumption usually represents a collective decision that incorporate multiple channels of information throughout the shopping journey (Salganik et al., 2006).

During the past decade, the volume and scope of information accessible to consumers have been pervasively increased. With the fast development of information systems and growing omnipresence of social information, consumers limited attentions are competed by more complex and diversified opinions from more distributed sources during their shopping journey (Chen et al., 2011; Cheung and Lee, 2012; Zhu and Dukes, 2017). For example, shopping platforms such as

Amazon.com, *Netflix.com*, and *Airbnb.com* leveraged rich consumer data to provide customized user intelligence through their recommender systems, which dramatically improved user search experiences. Meanwhile, the crowd-sourced review platform such as Yelp.com and TripAdvisor.com attracts millions of reviews and ratings every year that in turn reshape future consumers' quality assessments and decision makings.

The information competition also brings managerial challenges to the vendors. Because of the decentralized nature of user generated opinions, companies can find it more difficult to keep track of consumer sentiments and promote profitable products. Such excess of information can also hinder consumers from efficiently inferring the product quality and eventually reduce their purchase intentions (Iyengar and Lepper, 2000). As a result, being able to screen and filter useful information to consumers in a timely manner becomes critically important for business success (Edelman, 2015).

Among the intricate sources of product information, an important source arises from the social environment that consumers are embedded in. Usually, consumers are exposed to information from both the peer consumers who they don't know personally and from their own social network (Lee et al., 2015; Dewan et al., 2017). The two sources of information may appear at different stages of the shopping journey and play different roles in the purchase intentions. For instance, conversations with friends can increase awareness of products, encouraging consumers to externally search for product information (Branco et al., 2012). Meanwhile, crowds generated social signals, in forms of ratings, reviews and comments can interfere a consumer's own assessment of product quality and eventually affect her purchase decisions (Tucker and Zhang, 2011).

In this research, we aim to understand how consumers combine social signals from different channels as a function of how close they are to the point of purchase. We focus on two types of social information generated from two distinct sources: *popularity assessed by the crowds*, and *information from friends in the social network*. Our research questions are: *Do consumers combine the two sources of information during their search and purchase process? If yes, how do they assign weights to these signals when they start information search, and how they adjust the weights as getting closer to the point of conversion after more knowledge gained through information search?*

Understanding the relative importance of signals at various distances to the point of purchase may have valuable implications for marketers. It is widely accepted that in online marketplaces the vendors can strategically provide a search environment to reshape consumers' shopping paths (Dukes and Liu, 2015). When the goal is to promote contents to consumers so that they are more aware of their options, the platform can make the signal that relevant to the search stage more salient to consumers. If the goal is to encourage transactions, then the signal more relevant to the purchase stage should be highlighted. If the overall effect of crowd signals is more important, it would be a good idea for vendors to highlight aggregated popularity statistics from the proper user segments. Alternatively if social information from the consumer's own social network is more valuable, then the platform should incentivize the creation of social network, encourage conversations between friends, and make the friends generated signals more observable.

In this study, we are particularly interested in how consumers combine these social signals in an empirical context where they shop for movies on a home-based Video-on-Demand (VoD) platform. The movie context is ideal to answer the research questions for several reasons. First, movies are experience goods embedded in a cultural market. Consumers potentially value others' opinions to evaluate the quality of the content and let them guide their own consumptions. Second, as hedonic goods, movie contexts often appear in conversations between friends, triggering excessive word of mouth. Moreover, movie distribution platforms have changed dramatically during the past decade. Thanks to the development of over-the-top technologies and transitions of traditional TV industry, it becomes more flexible and less difficult nowadays to integrate social information in the distribution platform in real-time and personalized fashion. Finally, such transformation also makes the movie shopping experience similar to that of other home-based online goods. Understanding the social influence in the home-based movie industry will likely help us learn what could also happen from other media consumptions, such as music and books.

We developed a econometric structural model to describe the consumer search and purchase decisions. The typical search to purchase process starts with consumers sampling a subset of goods for further exploration to form a reasonable consideration set (Stigler, 1961; Ke et al., 2016). Later, they choose a product to buy from this consideration set or abandon the market without buying

(Mehta et al., 2003; Kim et al., 2010). We model the search-to-purchase process using a dynamic structural model that combines a discrete choice model with an optimal stopping framework based on the sequential search theory (Weitzman, 1979). The main idea of the model is that consumer keep searching products sequentially until a condition is reached that the marginal search cost of searching an additional product exceeds the expected marginal benefit of search. Our model takes into accounts the sequential arrival of information revealed by search and positive search cost. Consumers are assumed to be rational and forward-looking. They have a stream of decisions to make during this process, and each decision maximizes the expected future return given the real time level of information.

We instantiate the model with data from the VoD platform of a large multinational provider of telecommunication services. The primary dataset includes household-level clickstream data and transactional data for approximately 261 thousand search-purchase sessions from more than 117 thousand households during a six-month period (from June to December of 2015). We also approximate the social graph between users using their Call Detail Records (CDRs) for cell phone communications provided by the same service provider between August and October 2015. Two users are connected in our social graph if they made reciprocal phone calls during this period.

In our empirical context, we regard the number of likes for a movie as the signal of popularity assessed by the crowds, and proxy the friends signal by counting the number of friends rentals for the movie before it was browsed by the focal user. Following the search literature, we define a ”search” as the behavior of clicking through a movie cover to access the movie landing page, where more detailed movie information displays (Chen and Yao, 2016). In our setting, a consumer can observe the number of likes prior to search but the movie rental price is only revealed after clicking-through the movie landing page.

Our analyses find that positive signals from the crowds and from friends are associated with increases in the consumers evaluation of the product. In our context, friend signals seemed to have higher explanatory power on consumer search decisions to formulate the consideration set. *Likes* became more relevant to consumers when they are making purchase decisions, after they have collected information through information search. For instance, one additional *like* from the mean has a monetary value that worth 0.7 cents when deciding which movie to search, but increased to 3.7 cents when making purchase decisions. Meanwhile, the average value of a friend rental signal doesnt seem to change much when moving from search to purchase, worthing $2.6 when searching, and $2.9 when purchasing. In this particular market, on average, a friend rental would worth about 370 additional *likes* when making search decisions, but this value reduces to 80 when deciding which movie to buy.

The importance of the two social signals was moderated by product price. We find that when a consumer was evaluating an expensive product she would increase her reliance on friend rental signals comparing to when she was evaluating a cheaper option. Our study also showed that the effectiveness of a friend signal could be different depending on whether that friend enjoyed watching the movie, a measure of the signal valence. If the friend who rented the movie also enjoyed it, the corresponding signal to the focal user was associated with higher chance of purchase, and vice versa.

Our study provides new evidence of how consumers think throughout the shopping process. In our case, consumers seem to start by browsing products they heard about from friends and popularity information represented by the number of likes seem to only play a role closer to purchase. Our findings have valuable managerial implications to online marketplaces. For example, the result shows that a good way to improve consumer search environment and help consumers navigate large logs of products might be to first show them products that their friends bought and then later highlight the popularity information.

The rest of the chapter is organized as follows. In Section 3.2 we discuss the related works in the literature. We describe the empirical context and data in Section 3.3. The development of the structural model is detailed in Section 3.4, along with the design of estimation strategy and empirical models. In Section 3.5.1 we discuss the estimation results and in Section 3.6 we summarize the conclusions.

3.2 Related Works

This study is related to several streams of research works. First, it relates to the literature studying social influence on consumer behaviors. These works studied the impact of social information on product sales and reviews with applications in various empirical contexts. Our work also closely relates to the literature that models consumer decision makings along the conversion funnel, with focus on the product information search and product purchase. In particular, our work belongs to the stream of empirical studies that adopted the concept of sequential search framework.

3.2.1 Influence From Popularity Information

The averaged judgements from a large social group can be smarter and more reasonable when compared with that from individuals. This phenomenon is often referred to as the "wisdom of crowd effect" (Lorenz et al., 2011; Salganik et al., 2006). The social psychology literature suggests that when people become aware of the crowds decisions, they may adjust their own decisions makings for reasons including suspecting others may have better information (Barabási and Albert, 1999), partially following the wisdom of the crowds (Janis, 2008), or conformity concerns (Lorenz et al., 2011).

The influence of popularity information may usually be attributed to a hybrid types of social interactions and information exchanges between consumers (Chen et al., 2011). One primarily type among them is termed word-of-mouth (WoM), which is generally referred to as the dissemination of personal opinions through communications (Chevalier and Mayzlin, 2006). Examples of WoM from the crowds in online environment are consumer product reviews and ratings. Researchers have studied the impact of product reviews on consumer decision makings and obtained interesting findings such as negative reviews are more influential than positive reviews (Chevalier and Mayzlin, 2006), previous user ratings may shape future user ratings (Muchnik et al., 2013), the relative importance of volume and valence of WoM signals differs in various contexts (Chevalier and Mayzlin, 2006), and biased herding effect may converge to true qualities in long term (Godinho de Matos et al., 2016).

Another type of social interaction that related to the crowd influence is observational learning. The information cascade theory suggests that consumers tend to infer product quality from others' choices and follow them (Tucker and Zhang, 2011). Sometimes this happens even without knowledge about their peers preferences and purchase intentions (Banerjee, 1992). One example of observational learning information is aggregated market performance statistics (e.g. "sales rank", "number of subscribers", etc.). Researchers find evidence of observational learning in different market contexts, including kidney transplant (Zhang, 2010), wedding products (Tucker and Zhang, 2011), online music (Salganik et al., 2006), and computer software (Duan et al., 2009).

In many empirical context, however, it's impractical to separate the effect of WoM and observational learning (Dewan et al., 2017). No exception in our setting as we are interested in the crowd signal represented by the number of like votes for a movie. This metric can be viewed as a simplified consumer WoM signal if viewed as a product review with abstract positive votes. It also represents the overall popularity of a movie among peer audience, therefore influence consumers quality measure through observational learning. Similar popularity information metrics were studied in the recent works, for example, the number of favorites for a song in (Dewan et al., 2017), and the number of Facebook likes for Groupon vouchers (Li and Wu, 2013).

3.2.2 Influence From Friends

The theory of social influence in forms of word-of-mouth and observational learning are relevant for friends generated information as well, albeit the underlying mechanisms can be different. In theory, friends connected in a social network are socially closer that their shared experiences may lead to similar valuations, beliefs, and behaviors (Jussim and Osgood, 1989). Examples of friend WoM include product related conversations within the social network either online or offline (Li and Wu, 2013), as well as ratings and reviews from friends with author information identifiable to focal consumers (Dewan et al., 2017). Observational learning from friends can happen when a consumer see her friends purchasing a product and infer the product could be of good fit for herself as well.

Information from friends can be combined with external information about the social connections, therefore increasing the richness and validity of the signals. A consumer may pertain personal knowledge about her friend's taste, criteria, or purchase intent (Huang et al., 2014; Dewan et al., 2017). It has been suggested that information from socially closer others can be more reliable when people make decisions under high uncertainty (Huang et al., 2014; Galaskiewicz and Shatin, 1981). Lee (2014) also argued that consumers may be willing to trust information from closer friends than that from further ones as it is perceived to be more diagnostic and less uncertain than that from unfamiliar others.

3.2.3 Peer Influence In Movie Industry

A number of works studied the social influence in the context of the movie industry. Prior to the age of online peer generated reviews, studies focused on professional movie critics as they were the most prominent public opinion resources. Litman (1983) found that reviews of movie critics are positively correlated with box office sales. In a later paper, Eliashberg and Shugan (1997) questioned the causal influence of critics on movie box office performance, and with weekly box office data and movie critics data from 1990 to 1993, the authors reported that movie critics don't serve as good predictors for first week box office sales. Nevertheless, Reinstein and Snyder (2005) used a difference in difference model to find that movie critics reviews influenced box office sales during the opening week, and positive reviews had more salient impact.

A more recent stream of researches studied amateur generated movie reviews in the online context. The conclusions seem mixed but suggest that the valence, volume and dispersion of popularity signals may play different roles in movie box office performances. Zhang and Dellarocas (2006) applied a diffusion model and find that the valence measures significantly influenced movie box office sales, but the volume and variance measures didn't. In a later study, Dellarocas et al. (2007) find that with information diffusion model, volume, variance and dispersion of user generated reviews all have significant explanatory power. Likewise, Liu (2006) studied the influence of WoM from Yahoo Movies discussions board. He finds that volume of online conversations about a movie positively predicts movie box sales but valence doesn't. However, applying a simultaneous equation mechanism, Duan et al. (2008) finds that the volume of WoM significantly predicts movie box office sales but the valence, represented by user ratings didn't. Finally, Chevalier and Mayzlin (2006) studied the ratings for new TV shows during the 1999 2000 season. The researchers find that the dispersion of WoM was positively correlated with user ratings on these shows. Our study is relevant to the discussion on the relative importance of volume and valence of social signals to movie consumption. In our context, the number of likes can be considered as a partial measure of both volume and valence, while the number of friend rentals can be considered more as the volume than valence of social signals.

While most works focus on the social information from the crowds, few studies investigate the role of social influence from the personal network in the movie market. One notable exception is Lee et al. (2015). The authors used observational data primarily from a social movie site to compare the role of peer influence from both the crowds side and the friends side on consequent movie ratings. They find that while friends' ratings always positively correlate with the focal user's rating, the effect of crowds ratings was moderated by the popularity of movies. The influence from the crowds' ratings may reduce with presence of ratings from friends. While their focus was on information cascade after the point of purchase, our work studies an earlier part of consumer decision making process that is more relevant to consumption. We focus on the process that begins with consumer information search and ends to point of transaction.

3.2.4 The Conversion Funnel and Consumer Information Search

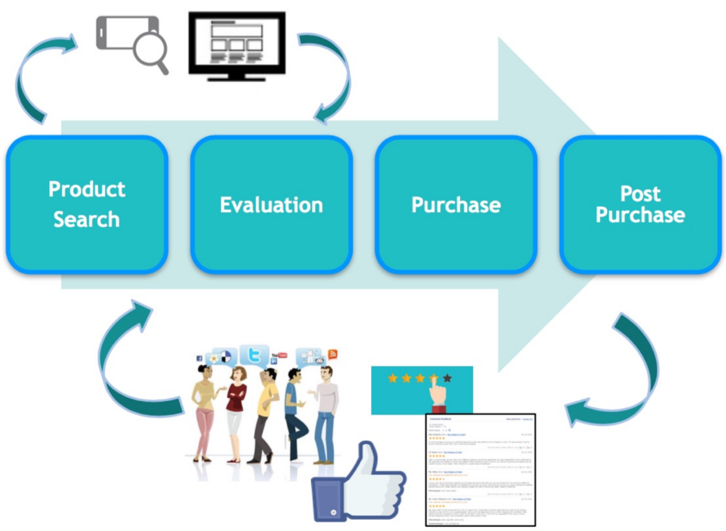
Our study is related to the literature where consumer decision making processes are modeled through a conversion funnel (Edelman, 2015; Kotler and Armstrong, 2010). The conversion funnel models suggest that consumer decision making involves a multi-stage process of ”awareness”, ”information search”, ”before purchase evaluation”, ”purchase”, and ”post-purchase activity” (Jansen and Schuster, 2011).

We zoom in by focusing on a central part of the conversion funnel that starts with the processes of information search and ends with purchase decision makings. The primary goal of consumer information search studied in this paper is to reduce uncertainty about product quality and validate their beliefs rather than become aware of the products (Edelman, 2015; Ke et al., 2016). We model consumers sample a subset of products sequentially to collect more information in order to form a reasonable consideration set. Conditionally, they choose a product to purchase from this consideration set or abandon the market without buying (Mehta et al., 2003; Kim et al., 2010). With the stock of information updated through sequential arrival of information through search, a consumer adjusts the weights she puts on various product features and then make a collective decision (Branco et al., 2012).

Understanding consumer search behavior has long been the focus of studies theoretically (Stigler, 1961; Weitzman, 1979; Branco et al., 2012; Ke et al., 2016) and empirically (Mehta et al., 2003; Honka and Chintagunta, 2016; De los Santos et al., 2012; Kim et al., 2010; Koulayev, 2014; Ghose et al., 2018). Researchers model the search behaviors as an information accumulation process that consists of a series of dynamic decision makings with discrete points of information arrivals. Nevertheless, the literature divides into two different views when modeling the stopping rule of the search process: simultaneous search and sequential search. The former assumes that decision makers initiate search by first determining a fixed number of alternatives they would like to sample, and then select which alternatives to sample. The seminal theoretical work is Stigler (1961) and a list of recent empirical works adopted this strategy. For example, Honka and Chintagunta (2016) studied the US auto insurance industry using inquiry data and Mehta et al. (2003) studied the liquid detergent market with scanner data.

The later stream argues that the consumer information search in the online evironment should follow a sequential process. Engaging in search is costly and consumers do not predetermine the size of the consideration set. They start by sampling the product for which the information search is most attractive, and keep sampling until the marginal cost of searching an additional product exceeds the expected marginal benefit of searching (Mortensen, 1970; Weitzman, 1979; Reinganum, 1982). The seminal works that serve as the theoretical foundations for sequential search models are Weitzman (1979) and Reinganum (1982). Our paper joins a number of recent empirical works that applied sequential search models to study the online information search behaviors (Kim et al., 2010; Bronnenberg et al., 2016; Koulayev, 2014; Ghose et al., 2018; Chen and Yao, 2016; Ursu, 2018). Kim et al. (2010) studied the online camcorders market and developed a method to estimate the sequential search model using aggregated search data from Amazon.com. Koulayev (2014) studied the hotel market using search clickstream data and found that dynamic search models yield different price elasticities than static models, where the later could be biased due to endogeneity of search-generated choice sets. Ghose et al. (2018) also studied the hotel industry. The authors argue that the sequential search model can be combined with social media information to improve search engine performance. Chen and Yao (2016) focused on the refinement tools on hotel search engine. They find that refinement tools have significant impact on consumer decision makings and lead to less concentrated market structures. Finally, Ursu (2018) studied the impact of exogenous product ranking on hotel search engine on consumer search and purchase behaviors.

Figure 3.1: Digital Shopping Journey in a Social Environment



3.3 Data

The products we study in this paper are movies provided on a Video-on-Demand (VoD) platform. Our data is from a household-based VoD service platform (hereafter called *VoDMedia*) operated by a large multinational telecom company. In addition to its traditional triple play (TV, Internet, telephony) telecommunication services, the company provides VoD services through cable to approximately 1 million households. Our first dataset contains household level clickstreams and transactions for the period from June to December 2015. A typical search clickstream starts with a user launching onto the VoD platform, and consists of user initiated events such as page views, purchases, streaming, and like votes, etc. All events were recorded with timestamps of initiation and termination. The transactional data includes precise timestamp and transaction price for each movie purchased. Events from a same household were identified with an anonymous identifier, which allows us to recover a consumers comprehensive footprints at the VoD platform.

We regard the entire month of December in 2015 as the period of interest and use the data from the previous 6 months to calculate historical information. We identify a search-purchase session by clustering the stream of events that happen with less than 24 hours apart by the same household. We ended up having 261,407 search-purchase sessions from 117,455 households, who had browsed the VoD system at least once during the observational period. Consumers can search a same movie multiple times in the same or different sessions. On average, a household was engaged in around 1.9 search-purchase sessions per month, 24.8% of which ended up with at least one transaction. Within a session a consumer searched about 2.8 different movies and spent around 10 minutes before either renting a movie, or leaving the system without any transaction.

Table 3.1: Descriptive Statistics

Variable	**min**	**max**	**median**	**mean**	**sd**
Movie					
imdbRating	1.10	9	6.40	6.27	1.21
imdbVotes	5	$1,529,615$	$27,207$	$94,864$	$164,877$
year since release	0	74	5	6.51	6.84
price	1.49	19.99	2.99	4.43	3.34
perpetual VoD (PVoD)	0	1	0	0.14	0.35
number of likes when searched	1	$6,929$	110	391.60	698.30
Consumer search					
number of movies searched per session	1	62	2	2.822	3.517
number of movies with friends' purchases	0	239	5	10.940	15.820
number of purchases in session	0	1	0	0.239	0.427
number of search sessions in a month	1	10	1	1.719	1.175
number of friends with VoD service	0	311	12	16.4	20.9

VoD Interface

When a consumer navigated the VoD system, she was displayed a page with highlighted movies. This page contains multiple menus of suggested movies chosen by an editorial team without personalization. The menus were named headers, such as Suggested Movies and Recent Releases, etc. With the menus aligned vertically on the screen, movies within a menu were displayed horizontally. Despite the user TV configurations, the interface fit the screen with exactly 2 menus at a time and households may use a cursor, that highlights one movie at a time, to move across menus and movies within a menu.

Clicking on a movie cover in the highlight page leads to a movie landing page. When a consumer lands the cursor on a movie cover but hasnt clicked into its landing page, she is displayed with the concurrently accumulated number of like votes from peer VoD users for the movie by the time of browsing. She doesnt reveal the price to rent the movie until after clicking through the cover and entering the movie landing page. Along with price, she is also displayed on the landing page with additional movie features such as length of play, casts, directors, year of release, IMDB rating and IMDB votes, etc. From the landing page, she can either proceed to rent the movie with payment, click like/dislike button, or get back to the menu page. Most leased movies have an expiration period of 48 hours after rental, while 15% of the them can be kept perpetually within the VoD system.

We obtain the movie display information by processing raw system configuration log provided by the service provider. We obtained accurate real time movie display information, including exact positions of menus and movies, which is important for search cost measures, and displayed features on a movie landing page, etc., which is critical to understand what information consumers reveal from searching a movie. During the observation period the service did not personalize content positions but may dynamically change them per decisions from the editorial team.

Social Graph

We construct an approximated social graph between households using Call Detail Records (CDRs) for cell phone communications served by the same service provider. Each entry of the CDR contains the anonymized phone numbers of both callers and timestamps of the call. This dataset contains over 193 million calls that allow us to define an undirected social graph across households. We matched the anonymized phone numbers to their VoD accounts and kept the records of only calls made between two calling parties whom we identify having access to VoD services. An edge between two households is included in the social graph they had reciprocal calls between each other. The resulting social graph contains 474,617 households and 3,858,889 edges. The median and average degree are 11 and 16 respectively. Figure 3.2 shows the distribution of number of friends households had.

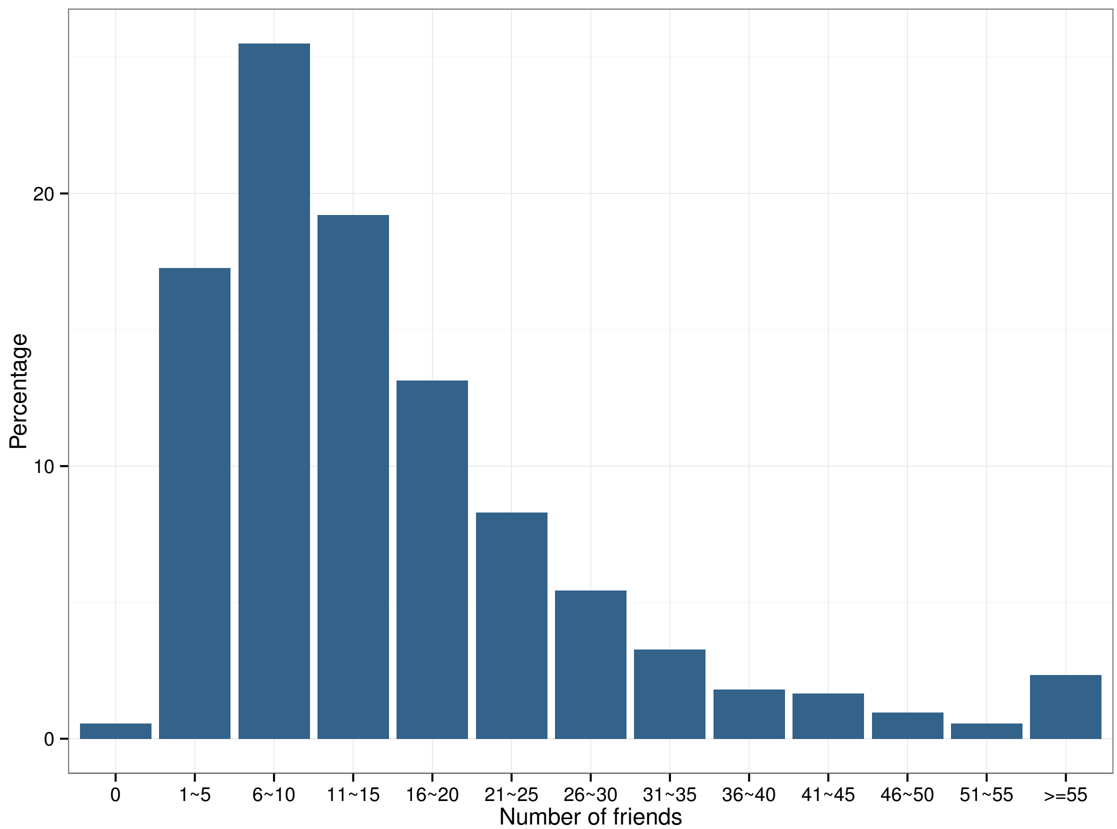


Figure 3.2: Number of friends distribution

Using the social graph, we calculated the household-movie-time specific social information that would be observed. We calculated the concurrently cumulative number of likes and concurrently cumulative number of friends rentals associated with a certain movie, to the point of time a consumer searches. The number or likes would be the number the consumer observed on screen and the friend rentals were also calculated as she would have obtained at the point of search. For example, if two friends Mike and Roger both rented a movie but Mike rented at least a day before Roger, Roger would be assumed to have rental signal from Mike but Mike would not observe a rental signal from Roger. Also awaring that the number of friends events may be biased to movies that have longer age in the VoD system, we used a moving time window of 6 months such that only related friends rentals happened within half a year were accounted.

3.4 Model and Estimation

3.4.1 A Structural Model for Search-Purchase Decision Makings

The online movie shopping process involves a series of decision makings. It starts with a consumer launching to the VoD system, proceeds with browsing and searching for movie information, and concludes with a monetary transaction or abandon the market without buying.

The search process alone consists of a stream of decision makings itself. With a zapper in hand, she needs to decide whether to move the cursor to a particular movie and click into the movie landing page to learn more about the movie. Then, she needs to decide when to stop searching and proceed to checkout. During the process the decisions she make are conditional on the cumulative information she has collected. We approach the dynamic search and purchase decision making by constructing a structural model.

We model the consumer search process following the sequential search framework (Weitzman, 1979). A forward-looking consumer builds her consideration set through a sequence of information search and ends the search process when the marginal benefit of searching another product is not worth the search cost. Then the purchase decision would be made conditional on the products she has searched, i.e., from her consideration set.

Utility of Movies and Search Costs

There are N consumers indexed by $i \in \mathbb{I} = \{1, ..., N\}$ and J movies indexed by $j \in \mathbb{J} = \{1, 2, ..., J\}$. A consumer has some individual-movie specific knowledge X_{ij} about a movie j prior to launching the movie landing page. By clicking into a landing page she can obtain more relevant information about the movie. As mentioned earlier, we define the click-through behavior as *search*.

When movie j's is searched, some additional movie-specific information z_j would be displayed on the screen and get completely disclosed to the consumer. Following the literature we assume that prior to search consumers do not know the particular value of z_j, but the distribution of z_j is public knowledge (Kim et al., 2010; Koulayev, 2014; Ghose et al., 2018).

We represent consumer i's utility function of renting movie j as follows:

$$u_{ij} = X_{ij}\beta + z_j\gamma + \epsilon_{ij} \tag{3.1}$$

where the stochastic component ϵ_{ij} represents the unknown idiosyncratic stochastic error, which we assume has Extreme Type I distribution ($\epsilon_{ij} \sim TypeIEV(0, 1)$). Prior to search a consumer doesn't observe the true values of z_j but knows its distribution. Therefore, her utility estimates were calculated by substituting the expected value of z_j in Eq (3.1), along with an uncertainty term (ξ_{z_j}). With a little abuse of notation we distinguish the prior search utility function with superscript 0:

$$u_{ij}^0 = X_{ij}\beta + \mathbb{E}(z_j)\gamma + \xi_{z_j} + \epsilon_{ij} \tag{3.2}$$

The goal of search for a consumer is to eliminate the uncertainty prior to search, i.e. to reveal $\xi_{z_j} + \epsilon_{ij}$.

Engaging in search is costly. We assume the search costs are generally known to consumers. We model consumer search cost as a function of the products appearance on the VoD system. The individual-movie-specific[1] search cost takes an exponential specification to ensure the switching

[1] The search cost for a movie is individual-movie-specific because the catalog page configurations were consistently changing over time. Therefore the search cost for a movie depends on when a consumer conducts search.

cost is always non-negative (Kim et al., 2010):

$$c_{ij} = exp(w_{ij}\eta) \tag{3.3}$$

Table 3.2: Information Updates of Search

	Consumers	Researchers
Known Before Search	X_{ij}	X_{ij}, z_j
Uncertain Before Search	z_j, ϵ_{ij}	ϵ_{ij}
Realized After Search	X_{ij}, z_j, ϵ_{ij}	X_{ij}, z_j
Uncertain After Search	$n.a$	ϵ_{ij}

Optimal Searching

Define an arbitrary search stage (after k movies were searched) with partition of movies $\mathbb{J} = S_{ik} \cup \bar{S}_{ik}$, where S_{ik} containing all searched movies ($|S_{ik}| = k$) representing the concurrent consideration set and $\bar{S}_{ik}$ containing all unsearched movies ($|\bar{S}_{ik}| = J - k$). A consumer would choose from two directions: (i) continuing to search another movie, or (ii) stop searching and choose a movie among the consideration, given that leaving the system without purchasing anything is also an alternative, which we regard as the outside option. We assume consumers are forward looking and rational, such that they make choices to optimize the expected return.

The benefit of searching a movie origins from the reduced uncertainty. Assume that the uncertain component of utility to a consumer before her search $z\gamma$ has a publicly known distribution with $c.d.f.$ denoted as $f(\cdot)$. We discuss the determination of $f(\cdot)$ in Appendix B.

Denotes the highest utility from one of the searched movies as $u^*_{ik} = max_{j \in S_{ik} \cup \{0\}} u_{ij}$. The expected benefit from searching an additional movie $j \in \bar{S}_{ik}$ is:

$$B_{ij}(u^*_{ik}) = \int_{u^*_i}^{\infty} (u^0_{ij} - u^*_{ik}) f(u^0_{ij}) du^0_{ij}. \tag{3.4}$$

A consumer will keep searching as long as there is at least one unsearched movie j' with marginal search cost less than the marginal benefit $c_{ij'} < B_{ij'}(u^*_{ik})$.

For each movie j there is a value R_{ij} that makes consumer i indifferent between keeping searching this movie. The value R_{ij} is formally referred as *reservation utility*, which is can be interpreted intuitively as the *attractiveness of search* (Koulayev, 2014) or a "Rate-of-Return" from searching (Ghose et al., 2018). R_{ij} solves the following equation and is unique given a utility function:

$$B_{ij}(R_{ij}) = c_{ij} \tag{3.5}$$

The seminal work in sequential searching Weitzman (1979) showed that a forward-looking consumer's optimal search strategy is to start by searching items associated with large attractiveness of search, until reaching a point that searching an additional content would always be associated with higher marginal search cost than marginal search benefit, i.e. $c_{ij'} \geq B_{ij'}(u^*), \forall j' \in \bar{S}_{ik}$. This is shown equivalent to $R_{ij'} \leq u^*_{ik}, \forall j' \in \bar{S}_{ik}$ (Weitzman, 1979).

Essentially, the optimal stopping rule would always be reached at some point. This is because that the marginal benefit of search $B_{ij}(u^*_{ik})$ is a monotonically decreasing function of u^*_{ik}. As a consumer keeps searching more movies, u^* is non-decreasing, while the highest reservation utility among unsearched items is a non-increasing function of k.. Search will reach a level such that $c_{ij'} \geq B_{ij'}(u^*_{ik}), \forall j' \in \bar{S}_{ik}$.

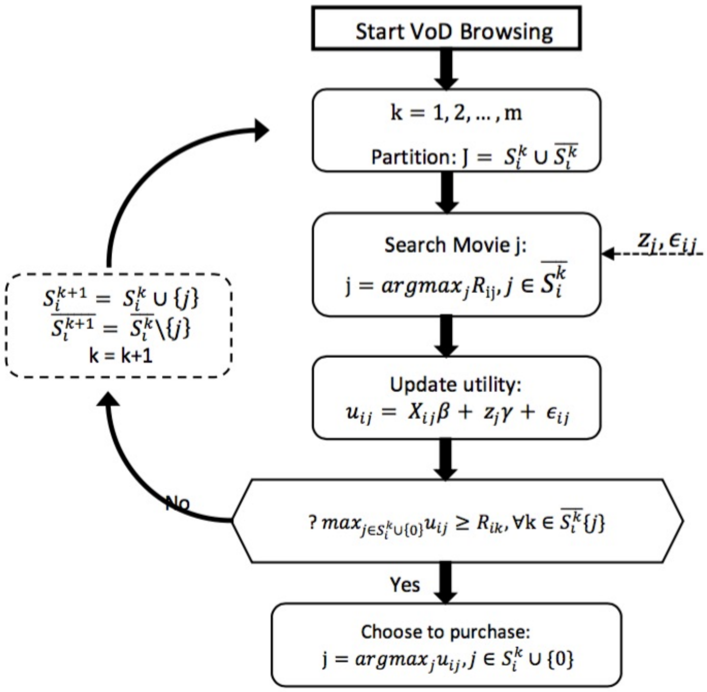
More formally the optimal sequential search and purchase strategy can be summarized to following steps: (1) Calculate the reservation utilities of movies and rank them in descending order, $R_{i1}, R_{i2}, ..., R_{iJ}$. (2) at any search stage $S_{ik} \cup \bar{S}_{ik}$, check the stoping rule to see if $R_{i(k+1)} > u^*_{ik}$. If true, keep searching the next movie with highest rank (the movie with R_{ik}). Otherwise, stop searching by choosing[2] $j = argmax_{j \in S_{ik} \cup \{0\}} u_{ij}$. A comprehensive illustration of decision making process in our model is shown in Figure 3.3.

3.4.2 Estimation

We apply the method of maximum simulated likelihood to estimate the set of parameters that determine product utility and search cost. The calculation of likelihood of a search-purchase session follows directly from the optimal search model above.

[2]The outside option, namely purchasing nothing, is associated with 0 search cost, therefore the reservation utility is infinity and it always exists in the consideration set.

Figure 3.3: Model Framework



Derive the Likelihood Function

Rank the movies in a descending order of their reservation utilities $r(1), r(2), ..., r(J)$ such that $R_{r(1)} \geq R_{r(2)} \geq ... \geq R_{r(J)}$. At an arbitrary stage when k products were searched (i.e. with partition $S_{r(k)} \cup \bar{S}_{r(k)}$), the $(k+1)^{th}$ product would be searched if its reservation utility exceeds the highest utility revealed. Denote $\rho_{r(k+1)}$ as the probability that $r(k+1)$ (for $k > 0$) is searched:

$$\begin{aligned}
\rho_{r(k+1)} =& Prob\{r(k+1)\ being\ searched\} \\
=& Prob\{R_{r(k+1)} > max_{j \in S_{r(k)} \cup \{0\}} u_j\} \\
=& Prob\{R_{r(k+1)} > max_{j \in S_{r(k)} \cup \{0\}} (V_{r(k)} + \epsilon_{r(k)})\}
\end{aligned} \tag{3.6}$$

According to the optimal stopping rule, the probability that search stops after $r(m)$ is searched can be represented as follows:

$$\begin{aligned} s_m =& Prob\{search\ stops\ after\ r(m)\ is\ searched\} \\ =& Prob\{max_{\tau \in S_{r(m)} \cup \{0\}} u_\tau > max^J_{\phi=m+1} R_{r(\phi)}\} \\ =& Prob\{max_{\tau \in S_{r(m)} \cup \{0\}} (V_{r(\tau)} + \epsilon_{r(\tau)}) > max^J_{\phi=m+1} R_{r(\phi)}\} \end{aligned} \tag{3.7}$$

The purchase decision can be modeled with a discrete choice model conditional on the consideration set determined by the search process. The conditional probability of choosing alternative j given $j \in S(m)$ is:

$$\begin{aligned} \omega_{j|S(m)} =& Prob\{alternative\ j\ has\ the\ highest\ utility\ in\ S(m) \cup \{0\}\ \} \\ =& Prob\{u_{ij} \geq u_{ij'}, \forall j' \neq j \in S_m \cup \{0\}\ \} \end{aligned} \tag{3.8}$$

Denote the final consideration set of consumer i's λ^{th} search session as $S_{i\lambda} = \{r(1), r(2), ..., r(m_{i\lambda})\}$. The likelihood of the entire search-purchase session can be calculated in the following way:

$$\begin{aligned} L_{i\lambda} =& Prob\{S_{i\lambda}\ finalizes\ the\ consideration\ set\ and\ r(k) \in S_{i\lambda}\ is\ chosen.\} \\ =& L^{search}_{i\lambda} \cdot L^{purchase}_{i\lambda} \\ =& (\prod_{j=1}^{m_{i\lambda}} \rho_{r(j)}) s_{m_{i\lambda}} \cdot \prod_{k \in S_{m_{i\lambda}} \cup \{0\}} (\omega_{k|S_{m_{i\lambda}}})^{y_{ik}} \end{aligned} \tag{3.9}$$

where y_{ik} is an indicator that consumer i purchased product j in her λ^{th} search-purchase session during the period.

Model Specifications

We consider X_{ij} to include the movie-individual-time specific characteristics that a consumer know prior to clicking into a specific movie page, for example number of likes on a movie (N_likes_{ij}) and number of rentals accomplished by friends ($N\_FrdRental_{ij}$), etc. N_likes_{ij} may be different for different users at different time since it is accumulating and time sensitive. We control previous searches using the number of previous search sessions involving a certain movie (but not purchase) before the current search-purchase session, $N\_SearchesBefore_{ij}$, to account for variation in knowledge a consumer carries before searching a movie. $N\_SearchesBefore_{ij}$ was known before search. Other movie characteristics we control in the utility function include IMDB rating score ($IMDBrating_j$ (out of ten), number of IMDB votes, $IMDBvotes_j$, and age of movie ($YearSinceRelease_j$), etc.

The vector of movie features z_{ij} contain knowledge that a consumer learned from searching the movie by clicking into the movie landing page. In our context, the most important movie feature that was revealed through information search is the price to rent a movie, i.e. $Price_j$.

With respect to search cost, we consider w_j to contain an intercept to capture the baseline search cost, a movie's appearance frequency in different menus within the system (N_path_{ij}) and the average sort order in menus it appeared ($SortOrder_{ij}$). We keep the subscript i in search cost variables to account for the change of movie locations when different consumers browse the system at different time.

The variables used in the model specifications above are summarized in Table 3.3.

We have the empirical model for utility of consumer i on movie j as follows:

$$
\begin{aligned}
u_{ij} =& \alpha - \beta Price_j + \mu X_j + \gamma_{11} log(N\_likes_{ij}) + \gamma_{21} N\_FrdRental_{ij} \\
&+ \gamma_{12} log(N\_likes_{ij}) \times PostSearch_{ij} \\
&+ \gamma_{22} N\_FrdRental_{ij} \times PostSearch_{ij} \\
&+ \gamma_{31} log(N\_likes_{ij}) \times PostSearch_{ij} \times Price_j \\
&+ \gamma_{32} N\_FrdRental_{ij} \times PostSearch_{ij} \times Price_j \\
&+ \epsilon_{ij}
\end{aligned}
\tag{3.10}
$$

and the search cost model as follows:

$$
c_{ij} = exp(\eta_0 + \eta_1 N\_path_{ij} + \eta_2 SortOrder_{ij}) \tag{3.11}
$$

We expect the price coefficient β to be negative because higher price reduces utility. All the coefficients represented by γ are related to the two channels of social signals. Because of the information updates happened through search, the values consumers had for different movie characteristics were different before and after search (recall in the previous section consumers substitute the expected values for variables they need to reveal after search). Consumers may also adjust the weights they assigned for movie characteristics (Branco et al., 2012). We estimate the weights adjustment by adding a dummy variable, $PostSearch_{ij}$, which equals 1 only when entering the utility function after a movie was searched, and 0 when the utility function represents the expected utility before search.

Table 3.3: Interpretations of Variables in Empirical Model

Variable	Interpretation	Revealed	Transformation
$Price_j$	Price to rent movie j when a movie was searched.	After	Normalized
$N_{likes_{ij}}$	Number of likes displayed for movie j when searched by i.	Before	Log-transformed
$N_{FrdRental_{ij}}$	Number of friend rentals for movie j when searched by i.	Before	NA
$IMDBrating_j$	IMDB Rating displayed for movie j.	After	Normalized
$IMDBvotes_j$	Number of IMDB votes displayed for movie j.	After	Normalized
$YearSinceRelease_j$	Movie j's age as of December 2015.	After	NA
$PostSearch$	Dummy variable, =1 when enters utility function after a movie is searched.	After	NA
N_path_{ij}	Number of appearance in different locations in the system for movie j when searched by i	Before	NA
$SortOrder_{ij}$	Average sort order within a menu of movie j when searched by i	Before	Normalized

We expect γ_{11} and γ_{21} to estimate the average weights a consumer had for one unite change of $log(N\_likes_{ij})$ and $N\_FrdRental_{ij}$ respectively before search. Correspondingly, coefficients for the interaction terms with $PostSearch_{ij}$, γ_{12} and γ_{22} were designed to capture the weight updates of N_likes_{ij} and $N\_FrdRental_{ij}$ respectively after information revelation. Consequently, $\gamma_{11}+\gamma_{12}$ represents the overall coefficient for $log(N\_likes_{ij})$ after search, and $\gamma_{21}+\gamma_{22}$ represents the overall coefficient for $N\_FrdRental_{ij}$ after search.

γ_{31} and γ_{32} were used to estimate the moderating effect of movie price that a consumer reveals after searching the movie. The interactions with price only enters the post-search utility function as price information is only realized after searching and not included in the utility function above.

In the search cost function, we use η_1 to capture the average change to consumer's search cost when a movie was displayed one more time in a different location. We expect η_1 to be negative because higher frequency of appearance makes users easier to search, reducing the search cost. η_2 was designed to capture the average change in search cost if a movie was displayed one more position to the right within the same menu. Since consumers may need to move the cursor to click movies that exist further to the right, which may be associated with higher cost of search, we expect η_2 to be positive.

Estimation Strategy

The model parameters $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}\}$ were estimated using method of maximum simulated likelihood. We derive the overall likelihood function of consumers searching and purchasing movies in the data from Equation 3.9 as follows:

$$Likelihood(\Theta) = \prod_{i=1}^{N}\prod_{\lambda=1}^{n_i} L_{i\lambda} \tag{3.12}$$

and the overall log-likelihood function is

$$LL(\Theta) = \sum_{i=1}^{N}\sum_{\lambda=1}^{n_i} ln(L_{i\lambda}) \tag{3.13}$$

We coded the estimator based on the likelihood function above. The overall log likelihood function was calculated in each update of parameters and then maximized through heuristic search[3]. Deciding to search a movie can either because of high utility or low search cost. Our identification strategy is based on the fact that consumer preferences enter both the search process and decision making process, while the search cost enters only in the consideration set formation. The covariates that enters only the search cost model but not the utility model serve as the exclusion restrictions for identification (Chen and Yao, 2016). The coefficients of covariates were identified in a way similar to the traditional discrete choice models, which depends on the variation in frequency of products with various features being searched and purchased.

3.5 Results

3.5.1 Impact of Social Signals on Search vs Purchase Decisions

Table 3.4 shows the estimation results of the model specified in Eq.3.10 and Eq.3.11. The "Baseline" model corresponds to the model without interaction terms. The "Search and Purchase" model

[3]Since the overall likelihood function in Equation(3.13) doesn't have a closed form expression and is non-smooth, we applied non-gradient based optimization methods, e.g. *Nelder-Mead* and also checked with other common optimization methods including *BFGS*, *BFGS with L1 regularization* and *NLM*. In our case different methods provides consistent results.

corresponds to the model that adds two interaction terms between the post search dummy and the two channels of social signals to capture the effect before and after search. The "Interacting Price" model corresponds to the model with additional three-way interaction terms to assess the moderating effect of movie price after it is revealed. For each model, the left column displays the point estimates and the right column shows the standard errors.

Table 3.4: Estimation results: Impact of social signals on search vs purchase decisions

(Utility)	Baseline Model		Search and Purchase		Interacting Price	
Intercept	-1.550***	(0.031)	-1.569***	(0.031)	-1.569***	(0.031)
Price	-0.113***	(0.003)	-0.108***	(0.003)	-0.096***	(0.003)
Number of likes	0.080***	(0.001)	0.024***	(0.002)	0.025***	(0.002)
N Friend Purchase	0.094***	(0.005)	0.070***	(0.009)	0.071***	(0.009)
N Searches Before	0.018***	(0.001)	0.017***	(0.001)	0.018***	(0.001)
N Likes × PostSearch			0.105***	(0.002)	0.103***	(0.002)
N Friend Purchase × PostSearch			0.015	(0.011)	0.013	(0.011)
N Likes × Price × PostSearch					-0.039***	(0.002)
N Friend Purchase × Price × PostSearch					0.070***	(0.010)
Other Controls (IMDB ratings and votes, movie age, movie age in platform, etc.)	Yes		Yes		Yes	
(Search Cost)						
Intercept	-1.279***	(0.013)	-1.678***	(0.031)	-1.688***	(0.032)
Appearance frequency	-0.599***	(0.008)	-0.461***	(0.009)	-0.459***	(0.009)
Slot within menu on screen (increasing left to right)	0.099***	(0.002)	0.090***	(0.002)	0.090***	(0.002)
Menu dummies	Yes		Yes		Yes	
(Log Likelihood)	-172217.0		-171005.7		-170867.6	

*** $p < 0.005$, ** $p < 0.01$, * $p < 0.05$

The price coefficient is negative and statistically significant as utility decreases with cost. All three models indicate that both positive signals from crowds (more number of likes) and from friends (more friend rentals) were associated with a consumer's higher evaluation on a movie. The coefficient for $N\_SearchesBefore$, the number of times the movie was searched before by the user, is positive and statistically significant, indicating that if the user searched a movie multiple time probably she's interested in the movie and may have higher expected utility on it.

The interaction terms between social signals and the $PostSearch$ dummy variable shows how consumers adjust their weights on the signals on average, after more information about the movie is revealed through search. Interestingly, the results suggest that the popularity information, represented by the number of likes for a movie, becomes more relevant after search comparing to itself when before search (coefficient for $N_{likes} \times PostSearch$ being positive and statistically significant). On the other hand, the friends rental signal does not seem to change its relevance comparing to itself after more information is revealed through search, as coefficient for $N\_FrdRental \times PostSearch$ being positive but statistically insignificant.

The post-search effect of the two channels of social information seems to be moderated by the prices revealed through search. As the "Interacting Price" model shows, the coefficient for $N_{FrdRental} \times Price \times PostSearch$ was positive and statistically significant, but the one for $N_{FrdRental} \times Price \times PostSearch$ was negative and statistically significant. This indicates that when the revealed price for a movie was higher, the friends signal was still more relevant when consumers were making rental decisions, comparing to itself and comparing to the popularity signal, in cases where average prices were revealed.

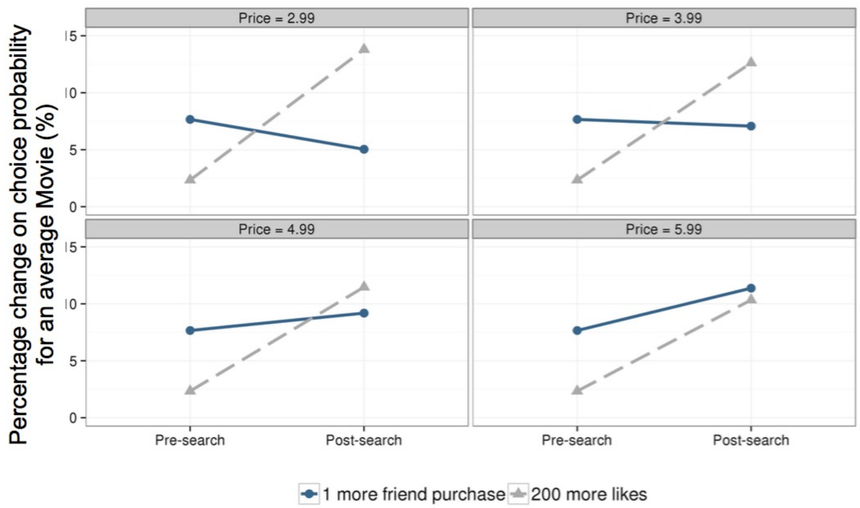
The ratio of these coefficients to the price coefficient allows us to interpret their effects in dollar terms. The average movie price in this study was \$4.6. In our context, one additional like increased from the mean has a monetary value that worth 0.7 cents when deciding which movie to search, but increased to 3.7 cents when making purchase decisions. On the contrary, the average value of a friend rental doesn't change much when moving from search to purchase, \$2.6 when searching, and \$2.9 when purchasing.

Alternatively, we can understand the relative importance between the signals at different stages by comparing how they substitute. We define the marginal rate of substitution of number of likes for friend rental as $MRS_{N\_FrdRental \rightarrow N_{likes}} = -\frac{\delta N\_FrdRental}{\delta N_{likes}}$. Comparing the marginal rates of substitution between signals, a friend rental would worth about 370 additional likes when making search decisions, but this value reduces to 80 when deciding whether to buy.

The effect change of the two channels of social information before and after search as a function of price can be visualized in Figure 3.4. The plots were based on estimation results from the "Interacting Price" model of Table 2. The plots provide examples on four groups of movies with representative prices from low to high, i.e. 1.49\$, 2.99\$, 4.99\$, and 5.99\$. In each quadrant, the y-axis represents the percentage increase on the probability[4] of choosing a movie relative to outside

[4]In the pre-search stage, it corresponds to the probability of searching the item, while in post-search stage it

Figure 3.4: Impact of Social Signals as a Function of Price



option with one more friend rental (solid line) and 200 more likes (dashed line) from the mean, respectively. The x-axis only contains two points which correspond to pre- and post-search stages.

The plot shows that in general the relevance of popularity signal assessed by the crowds increases after search comparing to itself before search was conducted. The relevance of friends rental signals remain on roughly the same level before and after search. However, when evaluating a more expensive movie (e.g., bottom-right) after it was searched, a consumer would increase her reliance on friend rental signals, and likely reduce her weights on popularity signals.

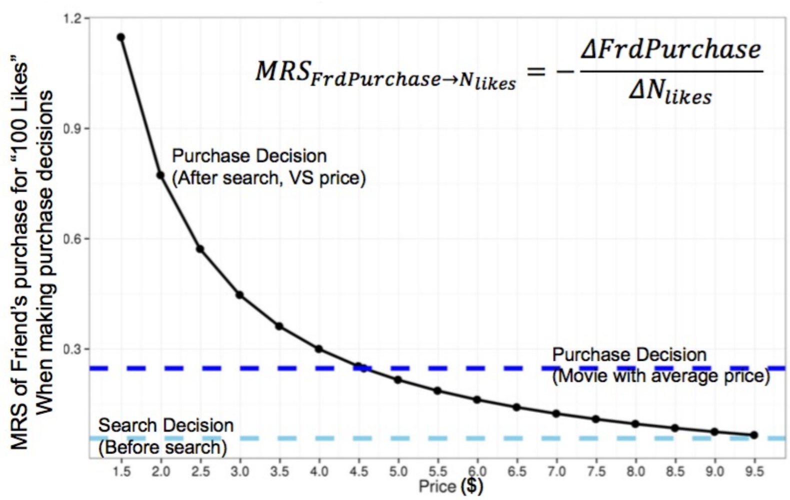
Figure 3.4 shows that when moving from pre-search to post-search stage, the $MRS_{N\_FrdRental \to N_{likes}}$ increases as the crowd signal becomes more salient, while the increase is less obvious with more expensive product. Figure 3.5 shows that how MRS decreases as product price increases.

Search cost

Our estimation results show that the average cost to search for one additional movie was 78 cents. Search cost for a movie reduces with its frequency of appearances within the system. One more display on *VoDMedia* reduces search cost by 46%. Horizontal location of a movie on the catalog page also changes the corresponding search cost for it. One more slot to the right on the TV screen

corresponds to the probability of purchase.

Figure 3.5: MRS as Function of Price



increases search cost by 0.13%.

3.5.2 Valence of Signals from Friends

We also evaluate how the impact of friend rental signals can differ with valence. Since we don't observe the conversation details between friends, we cannot obtain a very accurate measure on the valence of WoM[5]. Nevertheless, we tried to approximate the valence of friends rental signals by leveraging the time dimension of the data. Since we observed whether a friend who rented a movie also watched it to the end, this information can be used as a proxy of whether the friend enjoyed watching the movie. A customer who sat for hours in front of a movie was likely to enjoy the content, therefore if the movie was ever involved in conversations with friends, the WoM was more likely to be positive.

Empirically, we include two additional variables to interact with $N\_FrdRental_{ij}$. We used the friend's cumulative streaming time for the movie divided by the runtime of the movie to obtain a proportional measure of whether she watched it to the end. In our empirical analysis, if $\frac{t_{ij}}{runtime_j} \leq$

[5]Another possible approach was to observe whether the friend who purchased the content also clicked the like button. However, voting a like for a movie involves another costly decision making process. In our data, the support on such events was too sparse for estimation.

50%, we consider the consumer didn't watch the content long enough and she was more likely to not enjoy it that much. $\frac{t_{ij}}{runtime_j} \geq 85\%$ is another criterion that we used to assume a consumer also enjoyed watching the content. In our empirical case, around 83.8% of the purchases resulted in 85% watching in the end.

Table 3.5 shows the estimation results when we distinguish the valence of friends rental signals based on whether the friend also finished steaming the content. Column (1) and (3) correspond to models where we used $\frac{t_{ij}}{runtime_{ij}} \leq 50\%$ to suspect that the consumer didnt enjoy the content. Column (2) and (4) are models where we assume $\frac{t_{ij}}{runtime_{ij}} \geq 85\%$ to articulate a consumer enjoyed the content. These signal valence measures were interacted with friend rental variables.

Our analysis shows that the valence of friend rental signals can have different associations with the focal consumer's expected utility for a movie. The valence differences were likely to affect consumer shopping decisions starting from their search decisions (coefficient for $N\_FrdRental \times Watched < 50\%$ being negative and statistically significant and coefficient for $N\_FrdRental \times Watched > 85\%$ being positive and statistically significant). When a friend rented the movie and enjoyed it, the signal was associated with an increase in the expected utility for the focal user. However, when a friend who rented the movie but didn't enjoy it, the signal can be associated with a decrease in the expected utility. The level of relevances don't seem to change before and after search.

A visualization of the effect comparison between different valences of friends rental signals can be find in Figure 3.6. Again, we give examples on four groups of movies with prices varying among 1.49$, 2.99$, 4.99$, and 5.99$.

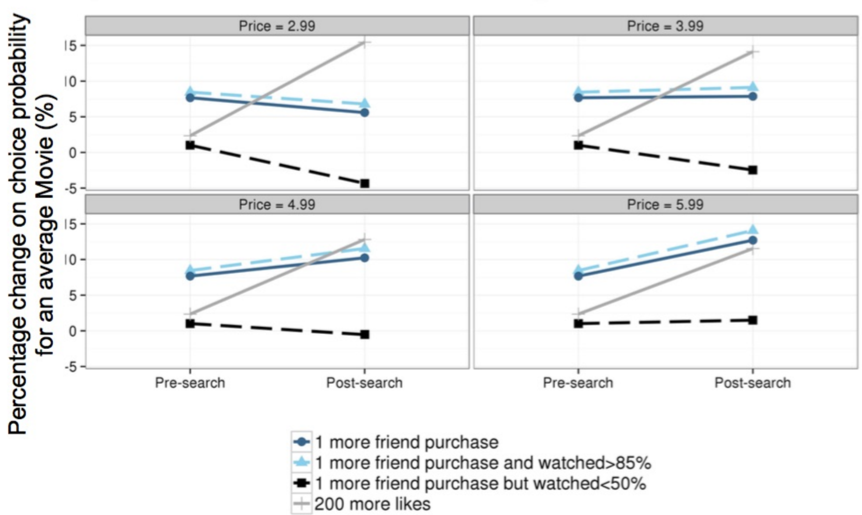
3.6 Discussions and Conclusions

In this paper, we study how consumers combine signals from different social channels as a function of how close they are to the point of purchase. We developed a dynamic structural econometric model that adapts the idea of optimal sequential search and discrete choices to estimate the social deterrents of consideration set formation and conversion decision making. We find that positive signals from the crowds and from the friends both positively predict consumers expected utility

Table 3.5: Estimation results when distinguishing the valence of signals from friends

Utility	Friend purchase and Watched<50%	Friend purchase and Watched>85%	3-way interactions Watched<50%	3-way interactions Watched>85%
Intercept	-1.558***	-1.558***	-1.570***	-1.570***
Price	-0.101***	-0.101***	-0.094***	-0.094***
N likes	0.081***	0.081***	0.025**	0.025**
N Likes × PostSearch			0.103***	0.103***
N Likes × PostSearch × Price			-0.039***	-0.039***
N Friend Purchase	0.101***	0.000	0.076***	-0.000
N Friend Purchase × Proportion Watched<50%	-0.102***		-0.066***	
N Friend Purchase × Proportion Watched>85%		0.103***		0.078***
N Friend Purchase × PostSearch			0.014	-0.000
N Friend Purchase × PostSearch × Proportion Watched<50%			-0.038	
N Friend Purchase × PostSearch × Proportion Watched>85%				0.017
N Friend Purchase× Price× PostSearch			0.069***	0.068***
N Friend Purchase × Price × PostSearch × Proportion Watched>50%			0. 000	
N Friend Purchase × Price × PostSearch × Proportion Watched>85%				0.001
Other Controls (IMDB, prior search, year of release, movie age, etc.)	Yes	Yes	Yes	Yes
(Search Cost)				
Intercept	-1.254***	-1.255***	-1.679***	-1.683***
Appearance frequency	-0.618***	-0.618***	-0.462***	-0.461***
Slot within menu on screen	0.101***	0.101***	0.091***	0.091***
Menu dummies	Yes	Yes	Yes	Yes
(Log Likelihood)	-172168.8	-172164.9	-170440.4	-170438.1

Figure 3.6: Impact of social signals as a function of price when distinguishing the valence of signals from friends



on a product. More interestingly, the weights a consumer put on these social signals seem to change after she reveals further information through search and when she reach the point to make purchase decision. The results show that previous rental information from friends is more relevant than likes from the crowds when consumers are deciding which movies to browse in order to form consideration sets. But when it comes to the point of purchase, the number of likes becomes more relevant.

There can be several explanations on why the relevances of social signals may change at different stages of the shopping process. Most likely, this phenomenon may result from the uncertainty reduction due to the accumulation of information through information search and the difference in the underlying mechanism of how the two sources of signals affect consumer decision makings. Previous researches suggest that information from socially closer others can be more reliable when people make decisions under high uncertainty (Huang et al., 2014; Galaskiewicz and Shatin, 1981; Gu et al., 2014). For example, Huang et al. (2014) showed that consumers demonstrated a tendency to seek support from "friends" and alleviate uncertainties during the early stage of task pursuit; however, the closeness significantly reduced after they reach the advanced stage of task completion and become less uncertain about how to reach the goal. Meanwhile, Gu et al. (2014) find that investors' tendency to seek information from others with similar status or values increases with uncertainty (represented by stock volatility), but decreases with an investor's experience and the amount of information to digest in front of them.

Applying the general conclusions into our contexts, at the beginning of the search process, consumers face higher uncertainty about their shopping goals and little knowledge about product quality. At this moment, information from friends seems to be more diagnostic comparing to one from unfamiliar others. However, when more detailed movie information is revealed through several rounds of information search, uncertainty reduces and consumers start to evaluate the selection of products they have narrowed down into their consideration set. Since the popularity information generated by the crowds can be a valuable source of quality assessment reflecting the "wisdom of the crowds" (Lorenz et al., 2011), consumers may rely more on it to decide which movie to rent in the end.

Understanding the relative importance of popularity information and friends information at different time points of consumer shopping journey has valuable implications for business practitioners. Recent researches show that online vendors often strategically design the search environment to guide consumers' shopping footprints (Dukes and Liu, 2015). This can be done by directing consumers limited attention to a specific information source by making it prominent (Zhu and Dukes, 2017). Foremost, our study helps online vendors to better understand consumer preferences in the lifecycle so that they can more efficiently guide consumers through the massive logs of products. Improving the shopping experiences for consumers may in return enhance the user engagement and monetization value. It seems that a reasonable strategy suggested by our analyses is to highlight friend signals earlier in a consumer's shopping process to guide her through the catalog and narrow down her selections; later when she is comparing the products given her consideration set, highlight the popularity information in order to assist her quality assessments.

Moreover, if the platform aim to promote contents to consumers so that they are more aware of their options, it may highlight the friend information to consumers as it is more relevant. If the goal is to encourage transactions, the platform can make popularity information more salient. Regardless of the stages, since both of the social signals have positive explanatory relationship in the overall purchase propensity, it would be a good idea for vendors to explore more informative popularity information to display, and on the other hand incentivize the creation of social network, encourage conversations between friends, and make the friends generated signals more observable.

The current work has several limitations that may serve as avenues of future work. First, we used observational data to study consumers decisions and the predictive relationship between different sources of social signals. However, using merely observational data may induce identification challenges if we aim to identify social influence from such signals. It would be interesting for future works and marketing practitioners to design randomized field and lab experiments to investigate the causal influences of various social signals at different shopping stages. In particular, in the next chapter we leverage a randomized online lab experiment to address this limitation.

Moreover, when constructing the structural model, we accounted for the sequential arrival of information during different search stages and allow for consumers to update the expected product utility. Nevertheless, our model, like in many other works in the sequential search literature (Kim et al., 2010; Koulayev, 2014; Ghose et al., 2018; Chen and Yao, 2016; Ursu, 2018), assumes that prior to search the general distributions of product features are public knowledge. Assumptions on these beliefs relates to consumers initial assessment on the attractiveness of search. Future work may relax the assumptions by considering consumers dynamically update their beliefs on product characteristics as more information gets in. Finally, signals can differ in their effect in expediting or slowing down a search to purchase transition. It would be interesting to model how consumers devote time in different search stages, especially in a social environment.

Chapter 4

Value of Social Signals in Consumer Search and Purchase: An Randomized Web Experiment

4.1 Introduction

In the last chapter, we studied the explanatory relationship of social signals from friends versus the crowd with consumers' search versus purchase decisions. We show that both popularity information from the crowds, represented by the number of likes, and private information from friends, proxied by the number of friends rentals, had positive explanatory power on consumer information search and product purchase preferences. Interestingly, in the home-based movie context, we find that consumers tend to rely more on the information from friends when deciding which movies to search in order to form a consideration set, and then the popularity information becomes more relevant when getting closer to the point of purchase.

However, using merely observational data may induce identification challenges if we aim to identify social influence from other unobservable mechanisms. It has been well demonstrated in the literature that several underlying mechanisms compete to explain the similar behaviors between socially connected individuals (Manski, 1993; Aral and Walker, 2011b). The most frequently cited explanations are social influence and homophily. When a friend's behavior causes the focal individual to behave in a certain way, which leads to the correlation between the actions, we refer the mechanism to social influence. Homophily, on the other hand, refers to the fact that an individual tends to befriend with another person with similar tastes (Bapna and Umyarov, 2015).

Because of that, it is not surprising that the two individuals are observed with correlated behaviors even when they make independent decisions without any information exchange between each other. Besides, there are also other unobserved stimuli that may confound their behaviors, such as vendors' promotions targeting a cluster of consumers utilizing their knowledge about the social network (Godinho de Matos et al., 2018).

Disentangling social influence from alternative explanations is critical for policy implications (Manski, 1993). If social influence is the driving force of the similar behaviors, companies may utilize their knowledge on the existing social network by targeting the most influential individuals to increase the probability that the induced behaviors propagate through social contagion (Godinho de Matos et al., 2018). At the same time, businesses may incentivize the creation of social ties and encourage interactions between certain nodes (Dewan et al., 2017). However, when homophily is dominant, since individuals make decisions independently instead of through contagion, targeting social influencers may not be as effective. A more reasonable marketing strategy would be to target a well-defined user segments (Bapna and Umyarov, 2015).

In this study, we study the same research questions as in the last chapter with a randomized web experiment to disentangle social influence from alternative explanations. We created an artificial "movie market" by constructing and operating an online Video-on-Demand platform (hereafter called "MoviePlatform") where users can search and rent movies. The shopping experiences on *MoviePlatform* were designed to be very similar to ones of *VoDMedia* users, whom we study in the last chapter with observational data. To increase the validity of participants' choices, we applied an incentive aligned reward strategy so that they can potentially realize their movie choices (Ding, 2007).

On *MoviePlatform*, experiment participants were provided a selection of 48 movies that were displayed in a similar way on the catalog page as that of the *VoDMedia* system. On the catalog page, randomized artificial social information were displayed for each movie. Consistent with last study, the social signals we study in this chapter include number of likes for a movie, which we educated the participants to be cumulative like votes from previous participants, and number of hypothetical rentals, where we introduce the concept of friends through an impression of social network before they enter *MoviePlatform*.

Consumer information search was again defined as the action of clicking through the movie cover page. By doing so participants access the movie landing page, where they read more detailed movie characteristics. On the landing page a user can also vote a “like” for the movie, which would immediately change the number of likes displayed for this movie in the system but within her observation only. Similar to the case of the observational study, price to rent a movie is only shown after participants search the movie. The number of likes and friends rentals were displayed both in the catalog pages and the movie landing pages.

In the experimental setting the value of social signals were both randomized. We also randomize the rental prices and movie display positions on the catalog page to partial out potential biases. We adopted a between-subjects-between-alternatives randomization and applied randomize block design strategy to potentially increase the estimation efficiency (Gerber and Green, 2012).

We recruited 483 experiment participants through Amazon Mechanical Turk in a ten-day period starting from March 9th, 2018. Several post-experiment survey questions were asked to the participants to perform transparency measure, understand the reasons why or why not certain social signal affected their decision making, and collect subjects demographic informationThe footprints of participants on *MoviePlatform* were completely recorded. The clickstream data with timestamps and the concurrent environment data allow us to recover the entire shopping journeys experienced by the participants. The sequential search framework developed in Chapter 3 was applied to estimate the effect of social signals in the search-to-purchase process.

Our analyses show that consistent with the results we find in the observational study, consumers seem to use the information of friends rentals to evaluate the set of movies for which they would like to conduct information search. Popularity information of a movie becomes more relevant to consumers when they are making purchase decisions, conditional on the consideration set formed through information search. In our experimental setting, when deciding which movies to search, one friend rental signal could ”worth” roughly 3600 ”likes”. However, when evaluating whether to rent the movie after it was searched, this number reduced by more than 90%, as the popularity information becomes more important. For *MoviePlatform* users, one additional like increased from the mean has a monetary value that worth 0.03 cents prior to search, but increased to 0.36 cents when making purchase decisions. Consumers were willing to pay on average 68 cents more for a movie with one more friend rental.

The results from the experimental study show very similar trends in terms of the relative significance of the two channels of social information as a function of how close consumers are to the point of purchase. While our experiment is clearly unlike real home-based movie markets in a number of respects, it provides meaningful evidence in addition to our findings in the observational study that the predictive relationship may potentially due to social influence. Our findings suggest that the two channels of social signals can play different roles in consumer shopping stages. With limited consumer attentions, displaying the right information at the right time can enhance consumer online shopping experiences and potentially increase engagement in the long term. The strategy suggested by our results is that online venders could highlight friends-related information in earlier stage to stimulate active information search, and later offer popularity information to assist quality assessments and expedite conversions.

4.2 Related Works

4.2.1 Identification of endogenous social effect

When studying the effect of social information through the social network on people's decision making, one major challenge is to address the reflection problem, namely to separate social influence from other unobservable endogenous factors (Manski, 1993). It has been well demonstrated in the literature that several underlying mechanisms compete to explain the similar behaviors between socially connected individuals (Aral and Walker, 2011b; Bapna and Umyarov, 2015). Chief among them is homophily, which refers to the fact that an individual tends to befriend with another person with similar tastes, such that two individuals are observed with correlated behaviors even when no communications or leaning ever happen.

Researchers have recently come up with several methods to overcome the identification challenge. Aral et al. (2009) and Dewan et al. (2017) applied propensity score matching method to alleviate the interference from unobservables. Godinho de Matos et al. (2014) and Tucker (2008), leveraged instrumental variable methods to partial out homophily from peer influence. The idea is to instrument friends generated signals with the signals from friends of friends that are not directly connected to the ego.

Several works identified social influence through randomized field experiments (Aral and Walker, 2011a; Bapna and Umyarov, 2015). Aral and Walker (2011a) designed a randomized experiment on Facebook with 1.4 million friends of around 10k experimental users to study the peer influence on adoption of a third party application. The paper finds evidence that significant social contagion can be created by embedding viral features into product design. Bapna and Umyarov (2015) conducted a randomized field experiment on Last.fm to study the social influence from online friends on consumers' decisions to upgrade their subscriptions with monetary cost. The authors found that friends adoptions to the premium account had positive impact on the probability that the focal users also pay to upgrade their accounts. The impact reduces when the focal user had more friends within the online social network.

Sometimes field experiments can become practically costly or infeasible. Several researchers leverage online lab experiments to study how decision makings are affected by social signals (Salganik et al., 2006; Salganik and Watts, 2009; Centola, 2010). For example, Salganik et al. (2006) created an online artificial music market where 14,341 participants can listen and download songs to study the effect of observational learning. The participants were randomly assigned into groups where they can either observe or not observe other participants' choices. The study find that true quality of songs can only partially predict consumption when observational learning takes place. Later, in a follow up paper, the authors find that their results were replicable in a pool of participants with rather different demographic features through a online web experiment with 2,930 participants (Salganik and Watts, 2009).

In another study, Centola (2010) designed an online lab experiment to study the diffusion of health behaviors through an artificially constructed online communities. In the three versions of the experiment (N = 98, N = 128, N = 144 respectively), each participant were assigned a few hypothetical friends called "health buddies", who had artificially closer relationship to the focal participant than others. The participants can observe their health buddies' adoption decisions to register for a health-related forum website. The study find evidence of positive social influence and compared the diffusion patterns of two different social network structures.

4.2.2 Incentive Design in Lab Experiment

We follow the literature in experimental economics to design the lab experiment in an incentive compatible fashion. The motivation is to induce realism into hypothetical lab tasks so that truth-telling be the dominant strategy in the Bayesian Nash Equilibrium (Smith, 1976; Smith and Walker, 1993; Ding, 2007; Miller et al., 2011).

Based on the induced value theory (Smith, 1976), three conditions need to be satisfied to solicit incentive-compatible behavior: monotonicity, salience, and dominance. Monotonicity requires the incentives received be monotonic with the hypothetical returns. Dominance requires that the value of incentive provided in the experiments be large enough that exceeds other hidden incentives, such as psychological costs (Smith and Walker, 1993). The most critical condition is salience, which requires that the rewards be directly related to the decisions that a participant makes during the experiment. Consequently, fixed payment to each respondent regardless of their decisions in the experiment may lead to biased and untrustworthy results, because the participant is aware that there is no relationship between his/her behaviors to the reward he or she receives.

To satisfy the above requirements researchers usually need to make the selections consequential by reflecting it in the final rewards they are receiving from the study. In this study, we use a randomizing mechanism similar to the concept of random lottery procedure (a certain percentage of randomly selected participants realize their selections) as the choice realization strategy (Starmer and Sugden, 1991; Ding et al., 2005). The benefits of such design is that it ensures experiment choices consequential in expectation while limits financial costs. The validity of random lottery procedure design has been verified in several empirical studies in the experimental economics literature (Becker et al., 1964; Ding, 2007; Miller et al., 2011; Ding et al., 2005; Brynjolfsson et al., 2017).

4.3 Experiment Design

We create an artificial “movie market” by creating and operating an online Video-on-Demand platform from where experiment participants can search and rent movies. The platform was designed so that its users have very similar experiences to the ones of *VoDMedia* users described in Chapter 3.

4.3.1 *MoviePlatform*

The experiment was conducted through a VoD platform called *MoviePlatform* that we designed, developed and operated exclusively for this study[1]. The platform had very similar features with the VoD system we studied in Chapter 3. The entry screen of *MoviePlatform* contains a catalog page filled with 48 different movies. Figure 4.1 shows an example of the catalog page. Movies are organized into four menus by genres, with headers that clearly identifies the genres of movies it contains, including ”Drama”, ”Action”, ”Comedy” and ”Family”. Twelve movies were included in each menu, which are two horizontal lines of movies with six side-by-side in each line. Different menus are stacked vertically. Users can also use the menu bar displayed on top of the screen to get directed to a menu with one-click. Depending on screen size at least one menu fits the screen at each time and users can scroll up and down to explore different menus. By design, users don’t need to move the cursor left or right to see the movies in each menu, regardless of the browsers they use[2].

Each movie was represented by a rectangle, on top of which shows the movie title and in

[1]The MoviePlatform website is database-backed and created through a Python-based micro web framework called *Flask*. The website was hosted by a WSGI-based web hosting service called *PythonAnywhere* which embed services for *Django*, *Flask*, and *web2py*, etc. The web pages were coded using an open-source front-end web framework called *Bootstrap*. The frameworks utilized in the website ensures stable website access and convenient experiment control and data collections. Code to construct the website available at https://github.com/youngchrisyang/MoviePlatform-Pythonanywhere.

[2]*MoviePlatform* is accessible for desktop/laptop users only and do not support mobile use, which we indicated clearly through out the experiment. By construction the system display features are robust and adaptable to most of the browsers and operating systems.

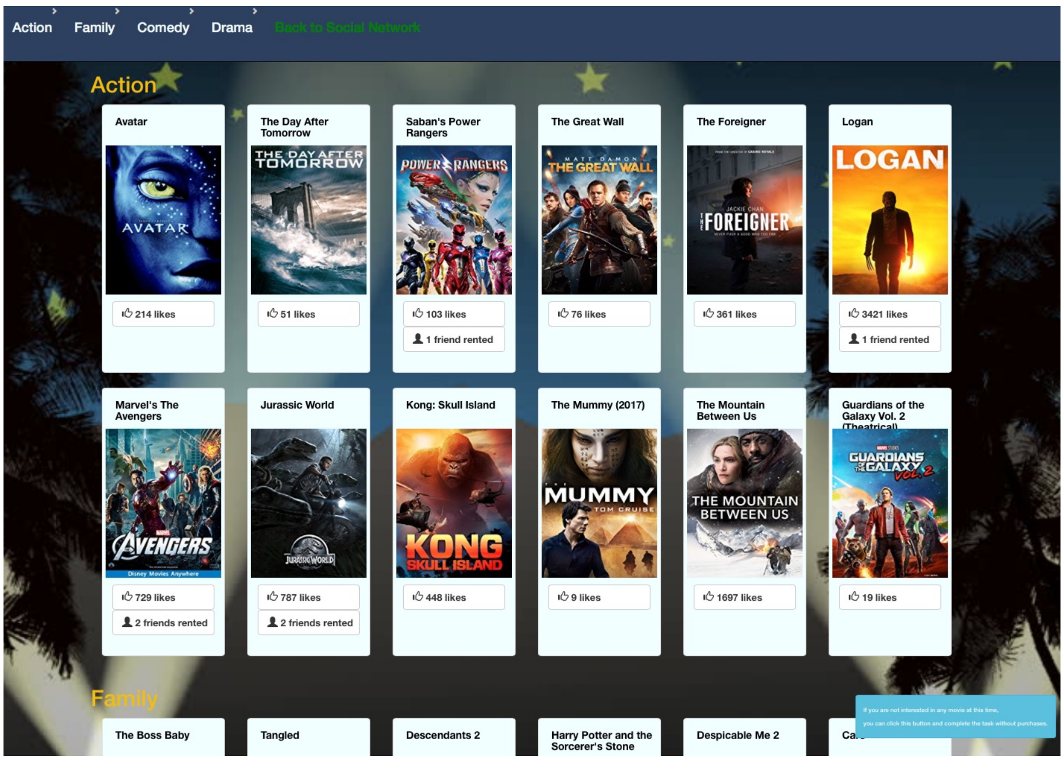


Figure 4.1: Example of catalog page in *MoviePlatform*

the middle the movie cover. Under the cover of a movie, the number of likes of the movie and number of friends who rented the movie were displayed. Clicking on the cover of a movie leads to the movie landing page. From the landing page users can read more detailed movie information, including the cast, directors, the year of release, play length, synopsis, and the social information consistent with what she observed in the catalog page. Consumers can also watch a trailer of the movie within the system by click the trailer link. If a consumer like a movie, she can also click the like button to vote (or cancel the vote) for the movie. The number of likes would update immediately and remain consistent for the rest of her shopping experiences. The numbers observed by other users would not be affected.

Finally, price to "rent" a movie is displayed on the landing page too. An example of the movie landing page can be find in Figure 4.2. Users can explore movies at $MoviePlatform$ back and forth between the catalog page and movie landing pages easily. The shopping journey terminates with either a movie "rental" or abandon of market without "renting" anything.

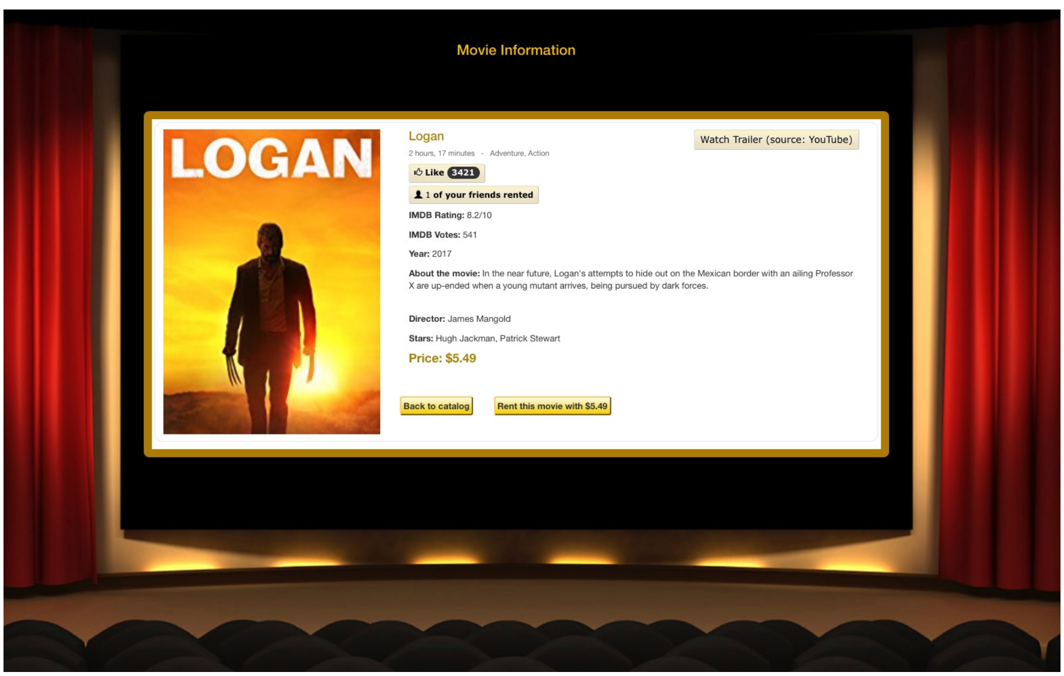


Figure 4.2: Example of movie landing page in *MoviePlatform*.

4.3.2 Randomization

Since our main research question is to understand how different channels of social signals affect consumer decision makings throughout the search and purchase process, we regard the different levels of "numbers of likes" and "numbers of friends rentals" displayed for movies as the treatments. We applied a between subjects between alternatives randomization design. For each experiment participant (*MoviePlatform* user), three dimensions of movie features were randomized for each movie displayed on *MoviePlatform*: (a) levels of social signals, including the number of likes and the number of friends rentals; (b) price to rent the movie on *MoviePlatform*; and (c) position of the movie displayed on the catalog page. Other movie characteristics displayed on *MoviePlatform*, including movie cover image, cast, directors, length of runtime, IMDB ratings, brief story, year of release, and trailer, were not manipulated[3].

To increase the efficiency and power of the experiment, we applied a blocked randomization design. The main idea is to divide observations into homogeneous blocks and then randomly assign the level of treatment within each block, in order to keep the experimental error within each block as well as possible (Gerber and Green, 2012). The randomization procedure is formally defined as follows.

Let $i \in \{1, 2, ..., N\}$ represent users and $j \in \{1, 2, ..., M\}$ represent movies ($M = 48$). Let X_j represent a vector of movie characteristics including IMDB rating, year of release, genres, and rental price charged on Amazon Video. Two movies j and j' with most similar movie characteristics ($d(X_j, X_{j'}) \approx 0$) were matched and form a block $k \in \{1, 2, ..., K\}$. Within each block, the two movies were randomly separated into a High-Likes (HL) group a Low-Likes (LL) group. For a user i, if a movie was assigned to the HL group, the number of likes displayed for that movie was a random draw from a lognormal distribution with mean 6 and standard deviation 1, i.e., $N\_likes_{ij} \sim lognorm(6, 1)$ if $j_i \in HL$. Movies assigned to the LL group has number of likes drawn randomly from a lognormal distribution with mean 4 and standard deviation 1, i.e., $N\_likes_{ij} \sim lognorm(4, 1)$ if $j_i \in LL$.

Independent to the "likes" group assignments, the two movies within a block were randomly

[3] We obtain these movie metadata from Amazon Video, IMDB, and YouTube.

Table 4.1: Distributions for displayed randomized volume of social signals and prices for movie j observed by user i

	Description	Groups	Chance	Distribution
N_likes_{ij}	Number of likes displayed	HL	50%	$Lognormal(6, 1)$
	on Movie j for User i	LL	50%	$Lognormal(4, 1)$
$N\_friend\_rental_{ij}$	Number of friend rentals displayed	HF	50%	Uniform $\{0, 0, 1, 2\}$
	on Movie j for User i	LF	50%	0
P_{ij}	Price charged on MoviePlatform	HP	50%	$\{2.49, 2.99, 3.49, 3.99, 4.49\}$
	on Movie j for User i	LP	50%	$\{4.99, 5.49, 5.99, 6.49, 6.99\}$

assigned into a High-Friends (HF) group and a Low-Friends (LF) group. A user i would observe no friend rentals for a movie assigned to the LF group. Instead if a movie was assigned to the HF group, the number of friends rentals for that movie was drawn uniformly from set $\{0, 0, 1, 2\}$. Therefore among the $M = 48$ movies, in expectation there were 12 movies with at least one friend rental, which is consistent with the number we observed in Table 3.1 in Chapter 3.

Similarly, two movies in the same block were randomly assigned into a High-Price (HP) group and a Low-Price (LP) group, independently from the ”Likes” and ”Friend rental” treatment group assignments. Movies in HP and LP groups had rental prices uniformly randomly drawn from $\{2.49, 2.99, 3.49, 3.99, 4.49\}$ and $\{4.99, 5.49, 5.99, 6.49, 6.99\}$ respectively. The combination of treatment level group assignments for a movie was randomly determined across users. Table 4.1 shows a summary of distributions used for different treatment level groups.

Finally, the vertical position of menus and horizontal position of movies within a menu were fully randomized. The pair of movies matched into the same block appeared in the same menu because menus were identified by movie genres, but their positions within a menu were randomized.

Despite the details in randomized block design, readers can consider the social signals (number of likes and number of friend rentals), movie prices and positions in catalog for movies were all randomized for a user and across users. Table 4.2 shows the correlation between movie characteristics and three movie features that we randomize in the experiment (p-values in brackets). The fact that all correlations are small is an indicator of satisfactory randomization design.

Table 4.2: Correlations between movie characteristics with number of likes, number of friend rentals, and price charged for movies on *MoviePlatform*. The brackets below the correlation estimate show the P-Values the Pearson correlation coefficients

	log(N_likes)	N_frd_rental	Price
log(N_likes)	1	-0.039	0.037
	(NA)	(0.89)	(0.91)
N_frd_rental	-0.009	1	0.023
	(0.89)	(NA)	(0.72)
Price	0.008	0.023	1
	(0.91)	(0.72)	(NA)
IMDB Rating	-0.009	0.053	-0.009
	(0.88)	(0.42)	(0.89)
Year of Release	-0.062	-0.003	-0.018
	(0.34)	(0.97)	(0.79)
Amazon Rating Count	0.003	0.040	0.013
	(0.96)	(0.54)	(0.84)
Genre Drama	-0.025	-0.056	0.003
	(0.70)	(0.39)	(0.96)
Genre Comedy	0.096	0.014	0.022
	(0.14)	(0.84)	(0.73)
Genre Action	0.000	-0.007	-0.014
	(0.99)	(0.91)	(0.83)
Genre Family	0.010	0.020	0.020
	(0.88)	(0.76)	(0.76)

4.3.3 Social Signals Impression

We introduce the social signals appeared on the platform to the experiments participants in the beginning of the study. We tried to keep the instructions concise and abstract so that it would not introduce conceptual biases. The number of likes for a movie was described as the cumulative counts of like votes by previous *MoviePlatform* users. The participants can actually click the like buttons to vote (or cancel the vote) for movies and observe the changes, which hopefully increased the trustfulness of the *likes* signals.

The concept of friends was introduced through a social network impression page. Recall that in the observational study in Chapter 3, we assumed *VoDMedia* users know what their friends had rented before they enter the platform. We understand that it is a limitation that in the online experiment setting we are not able to obtain information from participants' real social networks. Nevertheless, we try to resemble this by introducing an impression of friends by displaying a page that describes the hypothetical social network that a participant embedded in (see an example in Figure 4.3).

On the impression page, each subject was displayed a network of 12 anonymous friends (with "minion" avatars) and herself at the center of the network (with "Gru" avatar). Participants can click on the friends avatar to reveal a list of movies the corresponding friend has rented on *MoviePlatform*. They can explore as many friends' lists as they like and stay as long as they would like to. The lists of friends rented movies were randomized but consistent with the numbers a participant would observe on *MoviePlatform*. Participants can enter *MoviePlatform* from the social network impression page, and come back to the social network impression page from the movie market with one click.

4.3.4 Experiment Implementation

Recruiting platform and task structure

We use *Amazon Mechanical Turk* (in short *MTurk*) to recruit subjects. *MTurk* has been widely used in behavioral economic studies due to its capability to collect responses from geographically and ethnically diverse subjects in a short turnaround time with inexpensive cost[4]. Study participants were workers registered on *MTurk* that were based in United States and each participant can only complete the study once. Throughout the online lab experiments, participants were anonymous and only identified by a unique anonymized ID provided by Amazon.

The core task for participants is to complete a movie shopping journey on *MoviePlatform*. The shopping process starts with accessing the movie catalog page, and terminates after participants

[4]*MTurk* serves as an online crowdsourcing labor market where employees (called *workers*) are recruited by employers (called *requesters*) for the execution of tasks (called *HITs*, acronym for *Human Intelligence Tasks*) in exchange for a wage (called a *reward*).

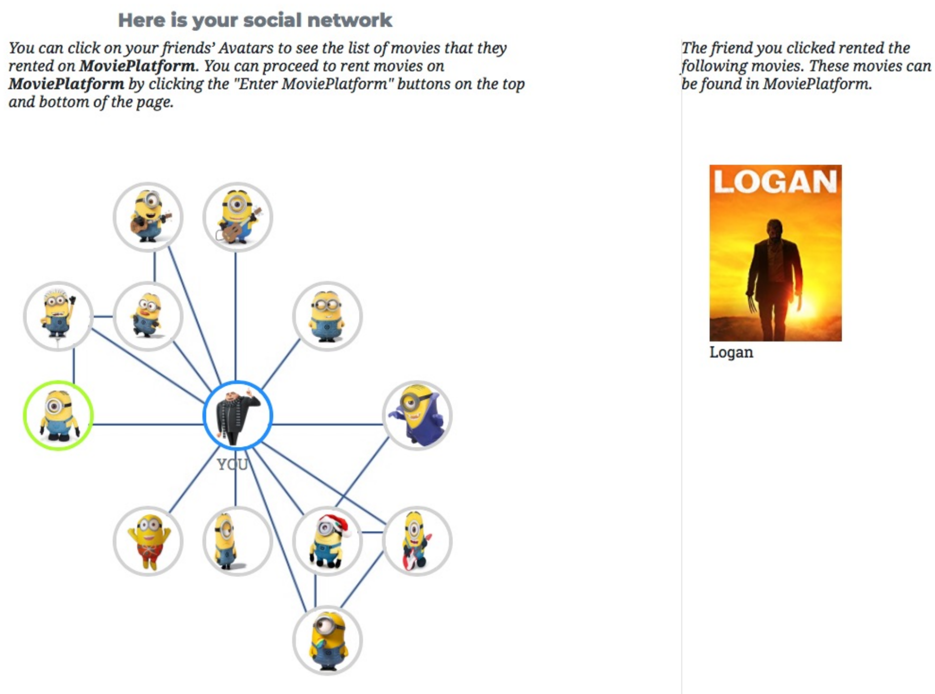


Figure 4.3: Example of social network displayed to participants

choosing to leave *MoviePlatform* after "renting" a movie or abandon the market without renting anything. We use the survey software *Qualtrics* to embed the core experiment task.

After initially accessing the survey, participants were shown the consent form and the general instructions about the study and reward calculations. Then they were shown the social network impression page where they can explore the rental list from 12 hypothetical friends. Then the participants were directed to *MoviePlatform* to complete the core shopping task. The numbers of rentals for movies a participant observe were consistent with what they observe in the social network impression page. Participants were allowed to refer back to the social network page any time during their shopping process. The core shopping task can be completed with a participant confirms "renting" a movie, or chooses to not renting any movies from *MoviePlatform* (opt-out). The "opt-out" option were made available and salient on all pages to participants to make sure they understand the existence of such alternative. A notification page of task completion were displayed to participants after either option above was confirmed.

After completion of the core experiment task, participants need to answer a few exit survey questions in the *Qualtrics* survey from where we measure participants' perceptions about the study and collect demographic information. At the end of the survey, participants may collect the survey completion code and return to the *MTurk* HIT page to redeem the rewards and complete they study. The survey experience from a participant's view point is illustrated in Figure 4.4.

Incentive Design

We follow the literature in experimental economics to design the lab experiment incentive compatible. The motivation is to induce realism into hypothetical lab tasks in order to increase the truthfulness of participants choices (Smith, 1976; Smith and Walker, 1993). In particular, we applied a randomizing mechanism called random lottery procedure similar to Starmer and Sugden (1991) and Ding (2007) as the choice realization strategy. The total reward a participant receives in our study includes two parts: a guaranteed uniform participation fee and an individual-specific lottery reward.

We leverage the lottery reward to realize the movie choices and create incentive-aligned bonus

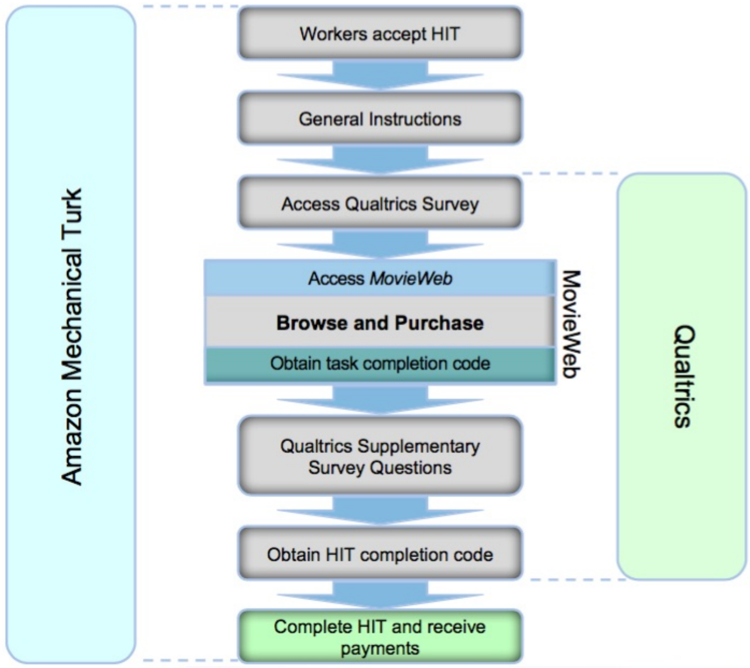


Figure 4.4: A participant's experience in the HIT

to encourage participants to behave as closely as possible to their real-life decision makings. If winning the lottery, the participant can be thought of as being granted a \$10 voucher to spend on renting movies from *MoviePlatform*. She would win an Amazon Video voucher (provided by us) to actually watch the movie she has "rented" during the shopping process in the study, plus keeping the unused portion of \$10 (\$10 minus the price of movie charged on *MoviePlatform*). Consequently, the final reward if winning the lottery is determined by the movie she has chosen and the price charged on *MoviePlatform*[5].

The lottery reward calculation method was introduced prior to the core shopping task, but the lottery results were only revealed at the end of the survey. This is critical to the validity of the random lottery design in order to avoid potential impact of knowing the lottery results on the shopping behaviors. Figure 4.5 shows the instruction on reward calculation we provided in the beginning of the *Qualtrics* survey prior to starting the core shopping task. More details of the

[5]By choosing to abandon the market without any rental, a participant can obtain a lottery reward of \$10 as the corresponding "movie" price was zero.

incentive design were discussed in Section C in Appendix.

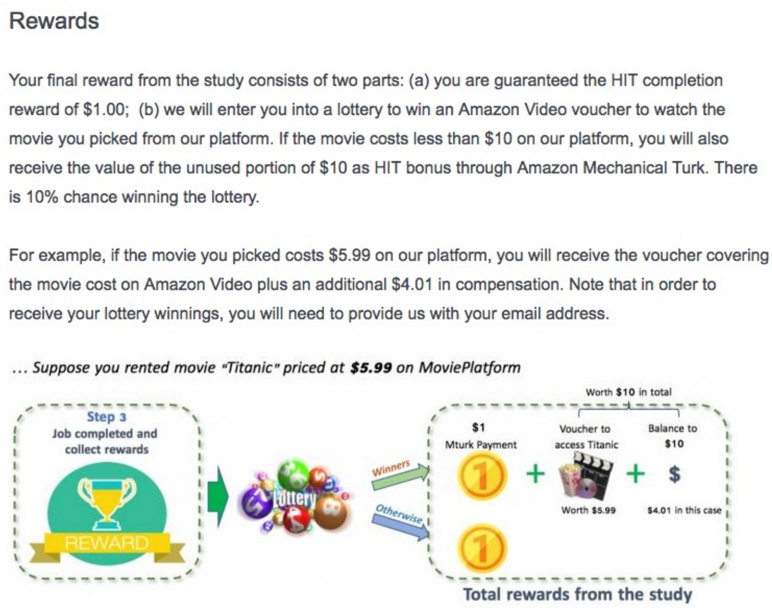


Figure 4.5: Instructions on study reward calculation used in the *Qualtrics* survey

4.4 Data

The experiment subjects were recruited through Amazon Mechanical Turk in ten days starting from March 9th, 2018. Surveys that embedded the core shopping task were distributed through batches scheduled at different time periods of a day (including morning, early afternoon, late afternoon, early night, and late night, in Eastern Standard Time.). Participants need to be currently located in United States to access the study. We also require participants to use laptop/desktop to complete the study to make sure the shopping experiences were comparable to ones of $VoDMedia$ consumers.

In total, 500 participants were recruited for the study and 483 completed the survey. We provide a detailed summary of participants reported sociodemographic and economic information in Section C in Appendix. Among the 483 participants who have completed the survey, 95.9% searched at least one movie by clicking into the movie landing page, and 93.1% ended up "renting" a movie on *MoviePlatform*. On average, they spent roughly 20 seconds on the social network page, explored the movie rental lists from 8 hypothetical friends out of 12 (Table 4.3). Participants spent 65.3 seconds on *MoviePlatform* and browsed for 2.4 different movies on average. This is fairly similar to the number of distinct movies browsed in a search-purchase session by *VoDMedia* users introduced in Section 3.1. It took averagely 8.3 minutes for experiment participants to complete the entire survey. Around 6% participants clicked the *like* button for at least a movie.

Table 4.3: Descriptive statistics for user behaviors on *MoviePlatform*

	N	Mean	Median	Sd	Min	Max
Searched at least one movie (1: Yes; 0: No)	483	0.959	1	0.199	0	1
Number of different movies searched (Given searched at least one movie)	463	2.37	1	2.71	1	21
Number of movies purchased (Given searched at least one movie)	463	0.931	1	0.254	0	1
Time spent to finish the survey (min)	483	8.31	6.98	5.35	1.55	48.52
Time spent on the social network page (sec)	483	21.73	18	20.10	0	59
Time spent on MoviePlatform (sec)	483	65.25	32	87.49	0	628
Number of different hypothetical friends browsed on social network page	483	7.52	8	4.30	0	12

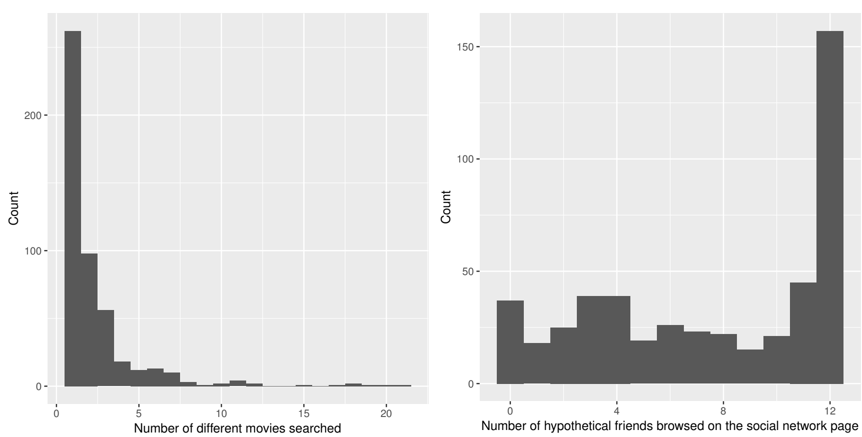


Figure 4.6: Number of different movies and number of hypothetical friends browsed

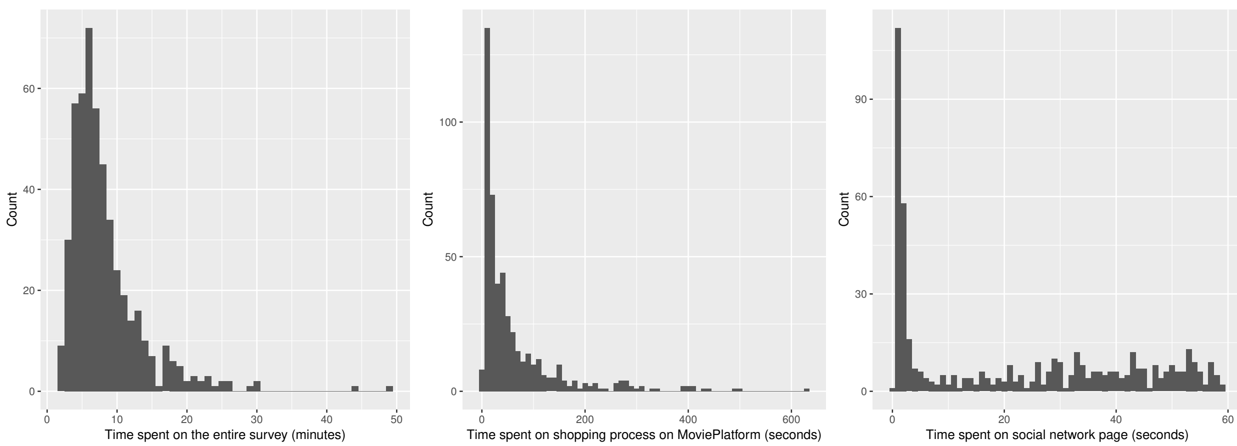


Figure 4.7: Time spent on different stages of the study

In the supplementary survey questions, we recorded several measures of study transparency and clarity (see Table 4.4). We surveyed the subjects on the “overall clarity of the movie renting task”, “transparency of the reward calculation”, and whether they agreed that “truthfully” selecting a movie were of their best interests. The three measures all showed scores higher than 4.6 (out of 5), indicating most of the subjects precept the study instructions and the shopping process transparent and clear.

A quiz was then used to measure whether participants understand the reward calculation was asked, and 85.1% participants gave the correct answer (See Figure C.3 in Appendix).

We also asked whether the participants had ever noticed and combined the social information during the core shopping task. We clearly referred the social information to the ”numbers of likes” and ”numbers of friends rentals” displayed for movies on *MoviePlatform*. More than 54% of them reportedly relied on at least one channel of social information during their shopping experiences on *MoviePlatform*.

Movies provided on *MoviePlatform*

We included 48 movies on *MoviePlatform*. These movies were chosen from the ”Featured Movies” list for ”rent or buy” on Amazon Video platform. The majority of these movies were recently available to the video-on-demand market and selected by the Amazon Video editorial team. We

Table 4.4: Measures of study transparency and clarity

Questions	N	Min	Max	Mean	Median	Sd
Was the movie renting process on MoviePlatform clear to you? (0 - very unclear, 5 very clear)	483	1	5	4.648	5	0.729
Was the reward calculation clear to you?(0 - very unclear, 5 very clear)	483	1	5	4.681	5	0.674
Was it clear to you that it was in your best interest to choose the best movie for you given your interests and price rather than choosing a random movie? (0 - very unclear, 5 very clear)	483	1	5	4.665	5	0.702

selected these movies because they were expected to attract interests from general interests, and more importantly we can realize participants' movie choices through a widely accessible platform (using Amazon Video vouchers). Table 4.5 shows the descriptive statistics for the 48 movies on *MoviePlatform*.

Table 4.5: Characteristics for movies displayed on *MoviePlatform*

Statistic	N	Median	Mean	Min	Max	St. Dev.
Number of Likes	48	130.5	483.688	8	4,941	907.642
Number of Friends Rentals	48	0	0.458	0	2	0.771
IMDB Rating (out of 10)	48	6.600	6.642	5.100	8.200	0.792
Amazon Rating Counts	48	358.5	2,122.000	8	19,783	3,813.799
Price charged by Amazon Video	48	4.990	5.157	3.990	5.990	0.859

4.5 Model and Estimation

We applied the same estimation framework developed in Chapter 3 for data analyses. When a user arrives at *MoviePlatform* with the intention to rent a movie, the website presents a list of movies positioned at different slots on the catalog page. For a given movie, the user is aware of the number of likes and number of friend rentals for the movie. However, she may be uncertain about some other product characteristics, for example, the price to rent the movie.

Consistently we define a search as the user clicking through the movie cover to browse the movie landing page. The click-through resolves the uncertainty about her utility but is meantime costly. To avoid redundancy, I briefly summarize the key expressions of the model and ask the readers to refer to Section 3.4 for the full model development. In section 4.5.1 we first specify the mathematical model of consumer utility and search cost, and in section 4.5.2 we specify the empirical models we used to describe them respectively.

4.5.1 Model Revisit

We represent consumer i's utility function of renting movie j as follows:

$$u_{ij} = X_{ij}\beta + z_j\gamma + \epsilon_{ij} \tag{4.1}$$

where X_{ij} represents a vector of movie characteristics the consumer knows prior to search the movie and z_j represents a vector of movie features that can only be learned after search. The stochastic component ϵ_{ij} represents the unknown idiosyncratic stochastic error to researchers, which we assume has Extreme Type I distribution ($\epsilon_{ij} \sim TypeIEV(0, 1)$).

Prior to search a consumer doesn't observe the true values of z_j but knows its distribution. Therefore their utility estimates were calculated by substituting the expected value of z_j in Eq (4.1). The prior search utility function with superscript 0:

$$u_{ij} = X_{ij}\beta + \mathbb{E}(z_j)\gamma + \xi_{z_j} + \epsilon_{ij} \tag{4.2}$$

where ξ_{z_j} represents consumer's uncertainty about variables z_j.

The goal of search for a consumer is therefore to eliminate the uncertainty prior to search, i.e. to reveal $\xi_{z_j} + \epsilon_{ij}$. We assume the search costs on *MoviePlatform* are known to consumers as private knowledge. Applying the same strategy as Chapter 3 we model the individual-movie-specific search cost has an exponential specification to ensure the switching cost is always non-negative.(Kim et al., 2010):

$$c_{ij} = exp(w_{ij}\eta) \tag{4.3}$$

where w_{ij} represents a vector of variables that affects the search cost of searching movie j

4.5.2 Empirical Specifications

For *MoviePlatform* users, we consider X_{ij} to include the publicly available movie information about movie j, and the randomized social signals (number of likes and number of friends rentals) displayed for movie j to user i. In particular, X_{ij} includes

- N_likes_{ij}. Number of likes for movie j displayed to user i. Normalized.;
- $N\_FrdRental_{ij}$. Number of friends rentals for movie j displayed to user i. Centered;
- $IMDBRating_j$. The concurrent rating out of 10 for movie j on IMDB.com. Normalized;
- $AmazonRatingCnt_j$. The concurrent number of ratings for movie j on Amazon Video. Centered;
- $MovieAge_j$. Age of the movie as of the year of 2018. Centered.
- Genre dummies. Indictor variable of whether the movie belong to a Genre. We used four genre dummies, including Action, Drama, Comedy, and Family

Similar to the *VoDMedia* system, the consumer reveal price information only after searching a movie. In the empirical model, we define z_{ij} as the price charged by *MoviePlatform* for movie j displayed to user i, denoted as $Price_{ij}$.

The empirical model for utility of consumer i on movie j can be represented as follows:

$$\begin{aligned} u_{ij} =& \alpha - \beta Price_{ij} + \gamma_{11} log(N\_likes_{ij}) + \gamma_{21} N\_FrdRental_{ij} \\ & + \gamma_{12} log(N\_likes_{ij}) \times PostSearch_{ij} \\ & + \gamma_{22} N\_FrdPurchase_{ij} \times PostSearch_{ij} \\ & + \mu_1 IMDBRating_j + \mu_2 AmazonRatingCnt_j + \mu_3 MovieAge_j \\ & + \mu_{4:7} \sum_{k=1:4} Genre_k + \epsilon_{ij} \end{aligned} \tag{4.4}$$

We expect γ_{11} and γ_{21} to estimate the average weights a consumer has for one unite change of $log(N\_likes_{ij})$ and $N\_FrdRental_{ij}$ respectively. Following the same development in Chapter 3, we define $PostSearch_{ij}$ as a dummy variable that equals 1 only when entering the utility function after a movie was searched, and 0 otherwise. The coefficients γ_{12} and γ_{22} capture the weight changes of N_likes_{ij} and $N\_FrdPurchase_{ij}$ respectively on the utility level after information revelation through search. μ represents the vector of coefficients corresponding to other movie characteristics controlled in the model.

Since users of *MoviePlatform* need to scroll down to access movies located in deeper menus, search cost is expected to increase with the vertical order of menu it belongs to in the version displayed to a consumer (menu order increases from top to bottom). In the meantime, consumers do not have to move the screen horizontally to access movies. We model the search cost as the baseline search cost plus the additional cost associated with scrolling down the screen to access the movie.

$$c_{ij} = exp(\eta_0 + \eta_1 Menu\_Order_{ij}) \tag{4.5}$$

The estimation of η_0 was used to measure the average baseline cost to search an additional movie. We expect η_1 to capture the additional search cost associated with scrolling down to access one additional menu vertically. The ratio between estimated c_{ij} and the price coefficient in the utility function β converts the search cost into monetary value.

4.6 Results

We estimate the structural model described in the section 3.4 on the search-purchase sessions from 463 experiment participants who searched at least one movie on *MoviePlatform*. We used the same maximum likelihood estimator developed for Chapter 3 and estimated the empirical models using the Maximum Simulated Likelihood method. For detailed estimation strategy please refer to section 3.4.2.

Table 4.6: Main Estimation Results Following the Sequential Search Framework

	Model (1)	Model (2)
Utility		
Intercept	-0.077* (0.044)	0.001 (0.137)
Price	-0.070*** (0.007)	-0.074*** (0.007)
N_likes	0.001 (0.010)	0.003 (0.010)
N_frd_rental	0.069*** (0.013)	0.070*** (0.013)
IMDB Rating	0.013ˆ (0.008)	0.023* (0.011)
N_likes × PostSearch	0.034* (0.015)	0.036** (0.015)
N_frd_rental × PostSearch	-0.023ˆ (0.018)	-0.021 (0.018)
Amazon Rating Count		0.091ˆ (0.061)
Movie Age		-0.001 (0.002)
Genre Dummies	No	Yes
Search Cost		
Intercept	-1.549*** (0.062)	-1.644*** (0.068)
Menu Order	0.028*** (0.007)	0.034*** (0.007)
Number of participants	463	463
Log Likelihood	-4417.19	-4382.65
ˆ<0.10; * <0.05; ** <0.01; *** <0.001		

4.6.1 Estimation Results

Table 4.6 shows the estimation results for both the utility model and search cost model. We showed two models in the table: Model (1) serves as the baseline model while Model (2) adds additional movie characteristics controls. Both models used the same empirical expression for search cost as Eq. 4.5.

The price coefficients were negative and statistically significant, indicating utility reduces with price. Prior to a movie being searched, the number of friends rentals seems to positively impact a consumer's expected utility for the movie (γ_{21} positive and statistically significant). The number of likes displayed for a movie didn't seem to change consumer's expected utility for the movie, as γ_{21} being positive but statistically insignificant.

After a movie was searched and more information about a movie was revealed, consumers seem to adjust their weights on the two social signals (number of likes and number of friends rentals) to form the expected utility for a movie. γ_{12} being positive and statistically significant indicates that on average a consumer increased reliance on the popularity information (number of likes). The results show that the popularity information still plays an important role (represented by $\gamma_{11} + \gamma_{12}$) in a consumer's rental decision since it only depends on post-search utilities (and the utility of the outside option).

γ_{22} is negative but statistically insignificant. This indicates that consumers seemed to slightly reduce their reliance on friend signals while making rental decisions, but the statistical support was weak. Nevertheless, the overall impact of friend signals, represented by $\gamma_{21} + \gamma_{22}$ were still positive and statistically significant.

Another method to understand consumers' adjustment of relative reliances on the two channels of social signals before and after further information was collected through search is to calculate the marginal rates of substitution between them. The results in Table 4.6 indicate that when forming the expected utility for a movie before searching it, one friend rental signal could worth roughly 3.6 thousands "likes". However, after the movie was searched the popularity information became more important, the number of likes needed to substitute one friend rental signal reduced to roughly two hundreds.

We can also obtain the money value of social signals by calculating the ratio between the variable coefficients and the price coefficients. For *MoviePlatform* users, one additional like increased from the mean has a monetary value that worth 0.03 cents when setting the expected utility level prior to search, but increased to 0.36 cents when making purchase decisions. Consumers were willing to pay 68 cents more for a movie with one more friends rental.

The average cost for searching one additional movie located in the top menu[6] was around \$2.6[7]. If the movie locates in one more menu vertically deeper on the catalog page, the search cost for that movie increased by roughly 3%.

4.6.2 Comparing With The Observational Study

When designing the online lab experiment, we aim to create movie shopping experiences similar to ones of *VoDMedia* users so that results from both parts are comparable. Foremost, the procedures to complete a shopping journey are the same for users in both platforms. Consumers started the shopping process by accessing the catalog page, proceeded with rounds of active information search by clicking into movie landing pages, and complete the journey with either a rental or not.

Table 4.7: Comparing the observational study and experiment study: Marginal rates of substitution between signals and signal Dollar values

	Observational Study		Experimental Study	
	Before Search	After Search	Before Search	After Search
Number of likes a friend rental signal worths	∼340	∼80	∼3400	∼200
Dollar value of a friend rental signal	∼2.6 dollars	∼2.9 dollars	∼0.95 dollars	∼0.68 dollars
Dollar value of one additional like	∼0.7 cents	∼3.7 cents	∼0.03 cents	∼0.36 cents

The information revelation processes in both platforms were mostly identical. We assume consumers had knowledge on only a subset of movie characteristics prior to clicking through the movie landing page. Search, defined as the click through behavior, reveals further information

[6]The orders of the four menus take value from $\{0, 1, 2, 3\}$

[7]The estimated baseline search cost in the experimental setting was find larger than the estimated baseline search cost in the observational setting introduced in last chapter (78 cents). This is potentially because users were more time sensitive in the experimental setting.

about a movie and resolves a consumer's uncertainty about the movie. In particular, users in both platform observed the number of likes and number of friends rentals before they click through the movie landing page, but only gains knowledge on how much the movie costs after searching.

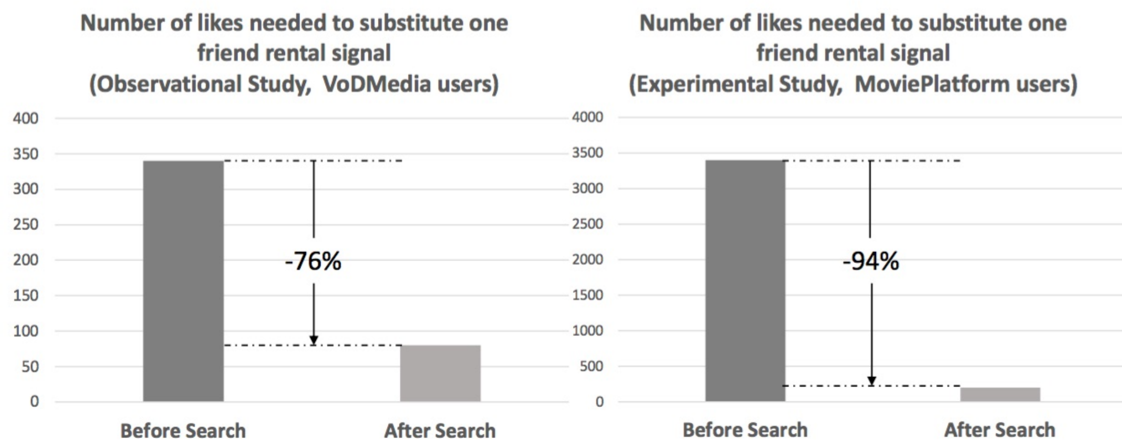


Figure 4.8: Marginal rates of substitutions between social signals in the observational study vs experimental study

The most substantial difference is that in the experiment study, both the social information and prices were randomized. Randomization provides better identification support by partial out the influence of signals from other underlying explanations.

Nevertheless, the estimation results applying the sequential search framework show very similar trends in terms of the relative significance of the two channels of social information throughout a consumer's shopping journey. We find that in both contexts consumers seem to use the information on friends rentals to evaluate whether to spend time to know more about a movie. Popularity information of a movie, represented by the number of likes in our contexts, becomes more relevant to consumers only when they are making purchase decisions after they have collected information through search.

Table 4.7 shows a comparison between the two contexts in measures of marginal rates of substitution and dollar values. In the experiment context, the number of likes a friend rental signal substitutes reduced by 94% (from ~ 3400 to ~ 200) when moving from the search stage to the purchase stage, comparing to a 76% decrease (from ~ 340 to ~ 80) as we estimated in the observational study context (See Figure 4.8). This similarity in trend can also be find in the monetary value of signals (See Figure 4.9). While the value of friend rental signals changes only slightly when moving from making search decisions to making purchase decisions in both contexts, the dollar value of likes increased by 5x and 12x in the observational study and experimental study, respectively.

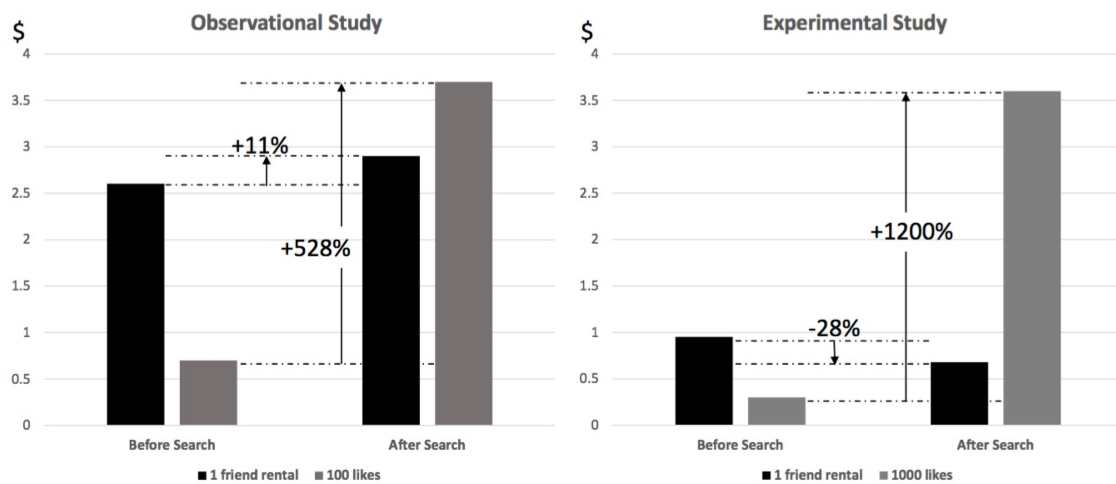


Figure 4.9: Dollar values of social signals in the observational study vs experimental study

4.7 Discussion and Conclusion

In this study, we created an artificial movie market where users can search and rent movies. Leveraging an incentive-aligned randomized web experiment we extensively studied the social influence of two sources of social signals on consumer product information search and purchase decisions. The experiment was designed to further study and identify the relationships we observed in Chapter 3. We find that consumers expressed the tendency to rely more on the friends signals when deciding which movies to search, and the popularity signals becomes more relevant for consumer purchase decisions.

The results in the experimental context are consistent with the ones from the observational study in terms of the relative importance of social signals as a function of how close consumers are to the point of conversion. This further suggests that the relationship we observed earlier may origin from social influence rather than homophily.

We eager to understand why consumers care about these social signals in the first place, and in the movie market why the two channels of signals can potentially play different roles in the conversion funnel. As part of the post-experiment survey questions, we asked the participants why or why not they choose not to rely on the two sources of social information (the survey questions can be find in Section C in the Appendix).

Around 41% participants reported that they have relied on the numbers of likes when shopping on *MoviePlatform*. The most prevalent reported reason is that the cumulated like votes are *"good quality indicators for movies"*. High number of likes *"shows the how well accepted"* the content is, or at least shows it's *"not terrible"*. A few others reported that they *"regard them as ratings"*, and they *"usually agree on the overall ratings and reviews"*. These answers show that if consumers care about the popularity signals, they would most likely use them for product quality assessment.

When we surveyed why consumers relied on the friends rental signals (slightly more than 30% participants reported that they did combine these signals when choosing movies on *MoviePlatform*), the most popular answer was that they have similar *"interests and preferences"* with friends and *"if they (their friends) enjoyed a movie it is likely I (they) will too"*. Meanwhile, more than a few participants pointed out that sometimes if they don't know what movies to watch, they would *"ask friends for recommendations"*. This is because they *"trust (their) friends on movie selections"*. Interestingly, some participants also reported that they ask their friends to help pick movies, in order to *"narrow down the selections"*. This may explain why friends signals play a role when consumers select which movies to search. A few others indicated that they find these signals useful because these related movies add value to their conversations (e.g. *"I see my friends and I can than have something to chat about"*). Several participants reported that they began to trust these signals when they observe movies they like also rented by their hypothetical friends.

There were also 46% participants indicated that they didn't rely on any of the these signals while shopping for movies on our platform. When asked about reasons, more than 90% of them reported that they would *"only trust (their) own tastes"* when it came to movies such that others' opinions were irrelevant. A few participants even pointed out that they often have opposite tastes than others, even friends. Some participants indicated that they know enough about movies from various information sources so that when making selections they would simply ignore these signals. The reasons seem to converge to the explanation that movies are hedonic goods and consumers may have more outside information about the true quality of the products, as pointed out by the authors in Godinho de Matos et al. (2016).

These answers suggest that the positive influences from these signals can due to a hybrid set of reasons. For people who care about these signals, the number of likes are more relevant to product quality assessment, just like other aggregated popularity signals such as ratings and reviews. On the other hand, friends signals influence consumer decision makings in a more complex mechanism. They may serve as recommendation signals from trustworthy sources, indicate the product can have social value, or filter out more relevant alternatives for consumers to reduce their search costs. Our discussion on why the relative importance of signals change at different shopping stages in the last chapter, therefore, is still valid and further supported here. We conjecture that when facing more uncertainty about the choices, consumers trust more on the information from credible connections. But when measuring product quality, popularity signals are still relevant because they reflect the wisdom from crowds.

Our findings have important managerial implications for online marketplace of digital experience goods such as online videos, books, and music where both popularity and friend influence might be at play. First, like what we suggested in the last chapter, a reasonable strategy to enhance consumers shopping experiences would be to highlight customized friends information from consumers' own social network earlier in their shopping process but later highlight the popularity information to assist product quality assessments. Moreover, since the relationships we find between the social signals and consumer decision makings were likely due to social influence instead of other underlying mechanisms such as homophily, business practitioners should encourage the creation of consumer social networks and promote communications between socially connected consumers. For example, the platform can create features that allow consumers to recommend or share their comments to friends directly. Finally, although the two sources of social information exhibit different influences along the conversion funnel, both exert positive impact on the overall purchase intentions. Marketers shouldn't be impeded to explore more valuable social signals leveraging knowledge on both the entire population and the social graphs among consumers.

Turning to limitations, our experiment is clearly unlike real movie market in a number of respects. For example, participants may not have the exact shopping intentions when the platform first appear in front of them than real consumers sitting on the couch with zapper in hand. Although only a small proportion of them reported in the survey, some experiment participants may have questioned the authenticity of the social information displayed during their shopping journeys. Although these differences limit the immediate relevance of our experiment to real market, our findings nevertheless suggest that social influences from the two different sources exerts important but different effects on search versus purchase decisions, especially when the trend of importance changes were consistent with what we observed in a different real market. Overall, combining with the observational analyses in the previous chapter, this study provides robust empirical regularities to the discussion of social influence on consumer shopping behaviors and practical managerial strategies.

Chapter 5

Conclusions

Telecommunication service providers can leverage various types of managerial strategies to remain profitable when facing competitions from innovative business models and technologies in the new digital age. This thesis examines the impacts of some of the firm initiated managerial strategies on consumers' switching, searching, and purchase decision makings.

In Chapter 2, we study consumers' subscription relationship with service providers by focusing on the role of lock in contract, a "loyalty program" created by the providers to keep consumers subscribed for a long enough period. It has been fiercely discussed by policy makers whether the status quo contracts (24-months) are outdated and need to be shortened to extensively protect consumer well-being. Using data from a large telecommunication service provider, we estimated a discrete choice model and conducted several policy simulations to study what may happen if the lock in length was reduced from what they were nowadays. We find that the shortening of lock in does make consumers more flexible in their choices. However, from a static point of view, the welfare consumers gain from such policies may not overcome the profit loss for the providers side, which to a great extent was caused by the increased acquisition costs on the providers side. This fact may jeopardize the overall success of the regulations because firms may collaboratively increase service prices and make consumers worse off.

The analyses show that merely reducing lock in periods may be insufficient to improve consumer well-being. Regulators need to account for the additional costs that firms incur to set up the consumers that churn. Although consumers can benefit from such policies to obtain more flexibility, firms are likely to increase prices to cover such costs, which are likely to pass on to consumers.

This may end up hurting consumers more than benefiting them. Therefore, regulators need to consider policies in tandem to achieve the desired objectives.

Given the several limitations our study experiences, we hope future studies can extend ours to provide further evidences to policy makers. We studied the impact of lock in shortening policies from a rather static perspective. Researchers can develop dynamic analytical models to study the market effect in the short and long term, and empirically instantiate the models. Moreover, in our study, although we obtained information about the local competition and best offers, we have to make several assumptions about where consumers switched to. Future research works can consider using data from multiple sources. For example, albeit difficult to realize, it can be extremely helpful if researchers can simultaneously obtain household data from several major service providers in a market, in order to understand where consumers actuarially switch to and why do they switch in the first place.

In the later part of the thesis, we investigate the search and purchase decision makings of consumers in a social environment by studying a specific service market where consumers rent movies to watch at home. Our research questions focus on the understanding of whether and how consumers combine social information from the crowds versus the friends at different stages of the conversion funnel. We proposed a structural econometric model to describe consumers' search and purchase behaviors as well as to quantify consumer search costs. The model takes into account the dynamic updating of information and consumer decision making under uncertainty. It combines an optimal stopping framework with a discrete choice model to jointly measure the paths from search to purchase.

To address the research questions, we applied the model to two connected yet distinct empirical contexts. We first studied a VoD market created by a large telecommunication service provider using large-scale household level clickstream and transactional data. Later, we created an artificial VoD market and leveraged a randomized web experiment to further understand the mechanism of social influences. We find that, in both contexts, consumers seem to rely on information from their socially connected friends to determine which movies to conduct information search. But later, when consumers are deciding which movies to rent among the movie they have searched, the popularity information becomes more relevant to consumers when they are determining whether to purchase a content. We also find evidence that both signals contribute positively to the overall chance that a movie is chosen.

Our findings implies that online platforms, where both sources of information may be at play, can leverage their knowledge on consumer social connections to strategically display social information to affect consumer decision makings. For example, understanding that friends information plays a more important role in the search stage, platforms can highlight such information to help consumers navigate through the complex catalog and narrow down selections quickly. This may greatly improve consumers shopping experiences and prevent early termination of the shopping process due to tedious information search. Meanwhile, consumers may value signals that are helpful for quality assessment in the later part of the shopping journey. The vendors, therefore, can provide various forms of popularity information, such as the number of likes in our example, or crowds ratings and reviews obtained from credible sources to help consumers reduce uncertainty. Potentially, information from friends may still be quite valuable at this moment if it can be enriched so that consumers can infer product quality from it. For instance, consumers may not be able to assess the movie quality by just observe their friends purchased it, but would be if they know whether they enjoyed watching it, or whether their friends had purchased similar contents before like them.

For future works, we suggest empirical works to investigate different formats or realizations of social information from both the crowds and the social network. Previous research works showed that the valence, volume and dispersion of social signals may have different influence on product sales (Liu, 2006; Duan et al., 2008). We are more than eager to understand whether social signals from the crowds and the friends, if appear in different formats, can have different impact on consumer search and purchase decisions than what we have identified in this study. This will further help us to extensively understand the underlying mechanisms of social influences on consumer information search and product purchase. Moreover, research works can be devoted to developing alternative structural models on consumer information search that relax several assumptions exerts in the traditional optimal search models. For example, consumers can be thought of with only bounded rationality (Simon, 1957), such that when they make search decisions, their decisions are shaped by the cognitive limitations of minds, time, and information, which is especially likely to happen when unlimited information competes for limited consumer attentions. Finally, as shopping platforms are continuously transiting to the mobile ends, empirical works can investigate how the two channels of social information can play different roles when information competes attentions more fiercely on smaller screens.

Bibliography

Aral, S., Muchnik, L., and Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549. 4.2.1

Aral, S. and Walker, D. (2011a). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management science*, 57(9):1623–1639. 4.2.1

Aral, S. and Walker, D. (2011b). Identifying social influence in networks using randomized experiments. *IEEE Intelligent Systems*, 26(5):91–96. 1, 4.1, 4.2.1

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817. 3.2.1

Bapna, R. and Umyarov, A. (2015). Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. *Management Science*, 61(8):1902–1920. 1, 4.1, 4.2.1

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512. 3.2.1

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science*, 9(3):226–232. 4.2.2, C.2.2

Bijwaard, G. E., Janssen, M. C., and Maasland, E. (2008). Early mover advantages: An empirical analysis of european mobile phone markets. *Telecommunications Policy*, 32(3):246–261. 2.2

Borenstein, S. (1991). Selling costs and switching costs: explaining retail gasoline margins. *The RAND Journal of Economics*, pages 354–369. 2.2

Branco, F., Sun, M., and Villas-Boas, J. M. (2012). Optimal search for product information. *Management Science*, 58(11):2037–2056. 1, 3.1, 3.2.4, 3.4.2

Bronnenberg, B. J., Kim, J. B., and Mela, C. F. (2016). Zooming in on choice: How do consumers search for cameras online? *Marketing Science*, 35(5):693–712. 3.2.4

Brynjolfsson, E., Eggers, F., and Gannamaneni, A. (2017). Using massive online choice experiments to measure changes in well-being. 4.2.2

Cabral, L. (2009). Small switching costs lead to lower prices. *Journal of Marketing Research*, 46(4):449–451. 2.1, 2.2

Calvo-Armengol, A. and Jackson, M. O. (2004). The effects of social networks on employment and inequality. *The American Economic Review*, 94(3):426–454. 1

Capgemini (2009). Quest for margins: Operational cost strategies for mobile operators in europe. *Telecom & Media Insights*. 2.1

Centola, D. (2010). The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197. 4.2.1

Chen, Y., Wang, Q., and Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of marketing research*, 48(2):238–254. 3.1, 3.2.1

Chen, Y. and Yao, S. (2016). Sequential search with refinement: Model and application with click-stream data. *Management Science*. 3.1, 3.2.4, 3.4.2, 3.6, B.2

Cheng, S. and Long, J. S. (2007). Testing for iia in the multinomial logit model. *Sociological methods & research*, 35(4):583–600. 2.5

Cheung, C. M. and Lee, M. K. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision support systems*, 53(1):218–225. 3.1

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354. 3.2.1, 3.2.3

Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *n engl j med*, 2007(357):370–379. 1

Confraria, J., Ribeiro, T., and Vasconcelos, H. (2017). Analysis of consumer preferences for mobile telecom plans using a discrete choice experiment. *Telecommunications Policy*, 41(3):157–169. 2.2

Congress (2013). Unlocking consumer choice and wireless competition act. 2.1

CRTC (2012). Proceeding to establish a mandatory code for mobile wireless services: Telecom notice of consultation crtc. 2.1

CRTC (2013). The wireless code. 2.1

CTIA (2014). Consumer code for wireless service. *The Wireless Association*. 1

Damodaran, A. (2015). Cost of capital data by sectors. 2.6.1

De los Santos, B., Hortaçsu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review*, 102(6):2955–2980. 3.2.4

Dellarocas, C., Zhang, X. M., and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, 21(4):23–45. 3.2.3

Dewan, S., Ho, Y., and Ramaprasad, J. (2017). Popularity or proximity: Characterizing the nature of social influence in an online music community. *Information Systems Research*, Forthcoming. 3.1, 3.2.1, 3.2.2, 4.1, 4.2.1

Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research*, 44(2):214–223. 4.1, 4.2.2, 4.3.4, C.2.2

Ding, M., Grewal, R., and Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of marketing research*, 42(1):67–82. 4.2.2

Doganoglu, T. (2010). Switching costs, experience goods and dynamic price competition. *QME*, 8(2):167–205. 2.2

Duan, W., Gu, B., and Whinston, A. B. (2008). Do online reviews matter?an empirical investigation of panel data. *Decision support systems*, 45(4):1007–1016. 3.2.3, 5

Duan, W., Gu, B., and Whinston, A. B. (2009). Informational cascades and software adoption on the internet: an empirical investigation. *Mis Quarterly*, pages 23–48. 3.2.1

Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2009). Do switching costs make markets less competitive? *Journal of Marketing research*, 46(4):435–445. 1, 2.1, 2.2

Dukes, A. and Liu, L. (2015). Online shopping intermediaries: the strategic design of search environments. *Management Science*, 62(4):1064–1077. 3.1, 3.6

Edelman, D, M. S. (2015). The new consumer decision journey. *McKinsey Quarterly*, pages 23–48. 1, 3.1, 3.2.4

Eliashberg, J. and Shugan, S. M. (1997). Film critics: Influencers or predictors? *The Journal of Marketing*, pages 68–78. 3.2.3

Epling, N. (2002). Price discrimination amid heterogeneous switching costs: A competitive tactic of the telephony resale fringe. *Yale University, unpublished*. 2.2

Erdem, T. and Keane, M. P. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, 15(1):1–20. 5

European-Union (2009). Regulatory framework for electronic communications in the european union. 1, 2.1

Farrell, J. and Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, 3:1967–2072. 2.1, 2.2, 2.4.2

Feinberg, F. M. and Huber, J. (1996). A theory of cutoff formation under imperfect information. *Management Science*, 42(1):65–84. 1, 3.1

Galaskiewicz, J. and Shatin, D. (1981). Leadership and networking among neighborhood human service organizations. *Administrative Science Quarterly*, pages 434–448. 3.2.2, 3.6

Gans, J. S. (2001). Regulating private infrastructure investment: optimal pricing for access to essential facilities. *Journal of Regulatory Economics*, 20(2):167–189. 1, 2.2

Genakos, C. and Valletti, T. (2011). Testing the waterbed effect in mobile telephony. *Journal of the European Economic Association*, 9(6):1114–1142. 2.6.2

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton. 4.1, 4.3.2

Ghose, A., Ipeirotis, P. G., and Li, B. (2018). Surviving social media overload: Predicting consumer footprints on product search engines. Technical report, Management Science, Forthcoming. 3.2.4, 3.4.1, 3.4.1, 3.6, B.2, B.3

Godinho de Matos, M., Ferreira, P., and Belo, R. (2018). Target the ego or target the group: Evidence from a randomized experiment in proactive churn management. *Marketing Science*, Forthcoming. 4.1

Godinho de Matos, M., Ferreira, P., and Krackhardt, D. (2014). Peer influence in the diffusion of iphone 3g over a large social network. *MIS Quarterly*, 38(4):1103–1134. 4.2.1

Godinho de Matos, M., Ferreira, P., Smith, M. D., and Telang, R. (2016). Culling the herd: Using real-world randomized experiments to measure social bias with known costly goods. *Management Science*, 62(9):2563–2580. 3.2.1, 4.7

Grzybowski, L. (2008). Estimating switching costs in mobile telephony in the uk. *Journal of Industry, Competition and Trade*, 8(2):113–132. 2.2

Gu, B., Konana, P., Raghunathan, R., and Chen, H. M. (2014). Research notethe allure of homophily in social media: Evidence from investor responses on virtual communities. *Information Systems Research*, 25(3):604–617. 3.6

Honka, E. and Chintagunta, P. (2016). Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science*, 36(1):21–42. 3.2.4

Honka, E., Hortaçsu, A., and Vitorino, M. A. (2016). Advertising, consumer awareness, and choice: Evidence from the us banking industry. B.3

Huang, S.-c., Broniarczyk, S. M., Zhang, Y., and Beruchashvili, M. (2014). From close to distant: The dynamics of interpersonal relationships in shared goal pursuit. *Journal of Consumer Research*, 41(5):1252–1266. 3.2.2, 3.6

Iyengar, S. S. and Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995. 3.1

Janis, I. L. (2008). Groupthink. *IEEE Engineering Management Review*, 36(1):36. 3.2.1

Jansen, B. J. and Schuster, S. (2011). Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, 12(1):1. 3.2.4

Jussim, L. and Osgood, D. W. (1989). Influence and similarity among friends: An integrative model applied to incarcerated adolescents. *Social Psychology Quarterly*, pages 98–112. 3.2.2

Ke, T. T., Shen, Z.-J. M., and Villas-Boas, J. M. (2016). Search for information on multiple products. *Management Science*, 62(12):3576–3603. 1, 3.1, 3.2.4

Keane, M. P. (1992). A note on identification in the multinomial probit model. *Journal of Business & Economic Statistics*, 10(2):193–200. 2.5

Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2010). Online demand under limited consumer search. *Marketing science*, 29(6):1001–1023. 3.1, 3.2.4, 3.4.1, 3.4.1, 3.6, 4.5.1, B.2, B.2

Klemperer, P. (1987). Markets with consumer switching costs. *The quarterly journal of economics*, pages 375–394. 2.1, 2.2

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *The review of economic studies*, 62(4):515–539. 2.1, 2.2

Knittel, C. R. (1997). Interstate long distance rates: search costs, switching costs, and market power. *Review of Industrial Organization*, 12(4):519–536. 2.2

Kotler, P. and Armstrong, G. (2010). *Principles of marketing*. Pearson education. 1, 3.2.4

Koulayev, S. (2014). Search for differentiated products: identification and estimation. *The RAND Journal of Economics*, 45(3):553–575. 3.2.4, 3.4.1, 3.4.1, 3.6

Lee, J. Y. (2014). Three essays on social learning. *Publicly Accessible Penn Dissertations*. 3.2.2

Lee, Y.-J., Hosanagar, K., and Tan, Y. (2015). Do i follow my friends or the crowd? information cascades in online movie ratings. *Management Science*, 61(9):2241–2258. 3.1, 3.2.3

Lewis, A. S. and Overton, M. L. (2013). Nonsmooth optimization via quasi-newton methods. *Mathematical Programming*, pages 1–29. B.3

Li, X. and Wu, L. (2013). Measuring effects of observational learning and social-network word-of-mouth (wom) on the sales of daily-deal vouchers. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pages 2908–2917. IEEE. 3.2.1, 3.2.2

Lieberman, M. B. and Montgomery, D. B. (1998). First-mover (dis) advantages: Retrospective and link with the resource-based view. *Strategic management journal*, pages 1111–1125. 2.2

Litman, B. R. (1983). Predicting success of theatrical movies: An empirical study. *The Journal of Popular Culture*, 16(4):159–175. 3.2.3

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 70(3):74–89. 3.2.3, 5

Lorenz, J., Rauhut, H., Schweitzer, F., and Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22):9020–9025. 3.2.1, 3.6

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542. 1, 4.1, 4.2.1

McKinnon, K. I. (1998). Convergence of the nelder–mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, 9(1):148–158. B.3

McKinsey (2015). Global media report 2015. *McKinsey Quarterly*. 1

Mehta, N., Rajiv, S., and Srinivasan, K. (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science*, 22(1):58–84. 1, 3.1, 3.2.4

Miller, K. M., Hofstetter, R., Krohmer, H., and Zhang, Z. J. (2011). How should consumers' willingness to pay be measured? an empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, 48(1):172–184. 4.2.2, C.2.2

Mortensen, D. T. (1970). Job search, the duration of unemployment, and the phillips curve. *The American Economic Review*, 60(5):847–862. 3.2.4

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341(6146):647–651. 3.2.1

Netzer, O., Lattin, J. M., and Srinivasan, V. (2008). A hidden markov model of customer relationship dynamics. *Marketing science*, 27(2):185–204. 1

OVUM (2015). Telecoms, media and entertainment outlook 2015. 1, 2.1

Prince, J. and Greenstein, S. (2014). Does service bundling reduce churn? *Journal of Economics & Management Strategy*, 23(4):839–875. 1

Reinganum, J. F. (1982). Strategic search theory. *International Economic Review*, pages 1–17. 3.2.4

Reinstein, D. A. and Snyder, C. M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The journal of industrial economics*, 53(1):27–51. 3.2.3

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856. 3.1, 3.2.1, 4.2.1

Salganik, M. J. and Watts, D. J. (2009). Web-based experiments for the study of collective social dynamics in cultural markets. *Topics in Cognitive Science*, 1(3):439–468. 4.2.1

Sharpe, S. A. (1997). The effect of consumer switching costs on prices: a theory and its application to the bank deposit market. *Review of Industrial Organization*, 12(1):79–94. 2.2

Shcherbakov, O. (2016). Measuring consumer switching costs in the television industry. *The RAND Journal of Economics*, 47(2):366–393. 2.2

Shin, J. and Sudhir, K. (2008). Switching costs and market competitiveness: De-constructing the relationship. Technical report. 2.2

Shy, O. (2002). A quick-and-easy method for estimating switching costs. *International Journal of Industrial Organization*, 20(1):71–87. 2.2

Simon, H. A. (1957). Models of man; social and rational. 5

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2):274–279. 4.2.2, 4.3.4, C.2.2

Smith, V. L. and Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, 31(2):245–261. 4.2.2, 4.3.4, C.2.2

Stango, V. (2002). Pricing with consumer switching costs: Evidence from the credit card market. *The Journal of Industrial Economics*, 50(4):475–492. 2.2

Starmer, C. and Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? an experimental investigation. *The American Economic Review*, 81(4):971–978. 4.2.2, 4.3.4, C.2.2

Stigler, G. J. (1961). The economics of information. *Journal of political economy*, 69(3):213–225. 1, 3.1, 3.2.4

Sun, B. (2005). Promotion effect on endogenous consumption. *Marketing science*, 24(3):430–443. 5

Tefficient (2013). How six months contracts changed mobile in belgium and denmark. 2

Train, K. E. (2009). *Discrete choice methods with simulation*, volume 8. Cambridge University Press Cambridge. 2.4.2

Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities. *Management Science*, 54(12):2024–2038. 4.2.1

Tucker, C. and Zhang, J. (2011). How does popularity information affect choices? a field experiment. *Management Science*, 57(5):828–842. 3.1, 3.2.1

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, Forthcoming. 3.2.4, 3.6

Viard, V. B. (2007). Do switching costs make markets more or less competitive? the case of 800-number portability. *The RAND Journal of Economics*, 38(1):146–163. 2.2

Villas-Boas, J. M. (2015). A short survey on switching costs and dynamic competition. *International Journal of Research in Marketing*, 32(2):219–222. 2.1

Wang, R. and Sahin, O. (2017). The impact of consumer search cost on assortment planning and pricing. *Management Science*. 3.1

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, pages 641–654. 1, 3.1, 3.2.4, 3.4.1, 3.4.1

World-Bank (2013). Real interest data from the world bank data. 5

Yao, S., Mela, C. F., Chiang, J., and Chen, Y. (2012). Determining consumers' discount rates with field studies. *Journal of Marketing Research*, 49(6):822–841. 5, 2.6.3

Zhang, J. (2010). The sound of silence: Observational learning in the us kidney market. *Marketing Science*, 29(2):315–335. 3.2.1

Zhang, X. and Dellarocas, C. (2006). The lord of the ratings: is a movie's fate is influenced by reviews? *ICIS 2006 proceedings*, page 117. 3.2.3

Zhu, Y. and Dukes, A. (2017). Prominent attributes under limited attention. *Marketing Science*, 36(5):683–698. 3.1, 3.6

Chapter A

Appendix for Chapter 2

A.1 Simulation Details

A.1.1 Churn Rate Estimation

To determine how exogenous changes of the lockin period change the market outcomes in terms of consumer surplus and firm profit we use the model presented in the previous section to determine how consumer switching behavior changes as lockin changes.

Figure A.1 shows the simulated churn probabilities over time. These probabilities are obtained by computing the market share of the churn alternative in our multinomial choice model. The plot shows that the churn rates increase over time when the lock-in period is active. This is consistent with the fact that the penalty that customers need to pay to churn reduces as the lockin period ends. There is large discontinuous increase in the probability of churn near the end of the lockin period after which the likelihood of churn decreases smoothly over time because customers who choose to stay with the carrier become increasingly loyal over time.

The color gradient of the different lines depicted highlights the probability of churn for different initial lengths of the lock-in period, which we obtain using simulation over the models estimated in column (1) of Table 2.3.

For shorter contracts, the monthly probability of churn is higher because switching costs decline. However, after the contract experies, shorter contracts are predicted to have less churn then longer ones. As possible explanation is that the likely churners drop out earlier, thus the clients that remain are less prone to churn, and another possibility is that consumers that were locked into longer contracts perceive differently the limitations associated to being in these contracts and become more willing to churn. In our empirical case, the survival probability converges after around five years.

The figure also shows how the probability of churn vary with different intensities market competition intensity. In markets with higher number of firms (dashed lines), the baseline probability of churn is higher because consumers experience smaller switching costs (γ_{10} and γ_{11} are positive and statistically significant in Table 2.3).

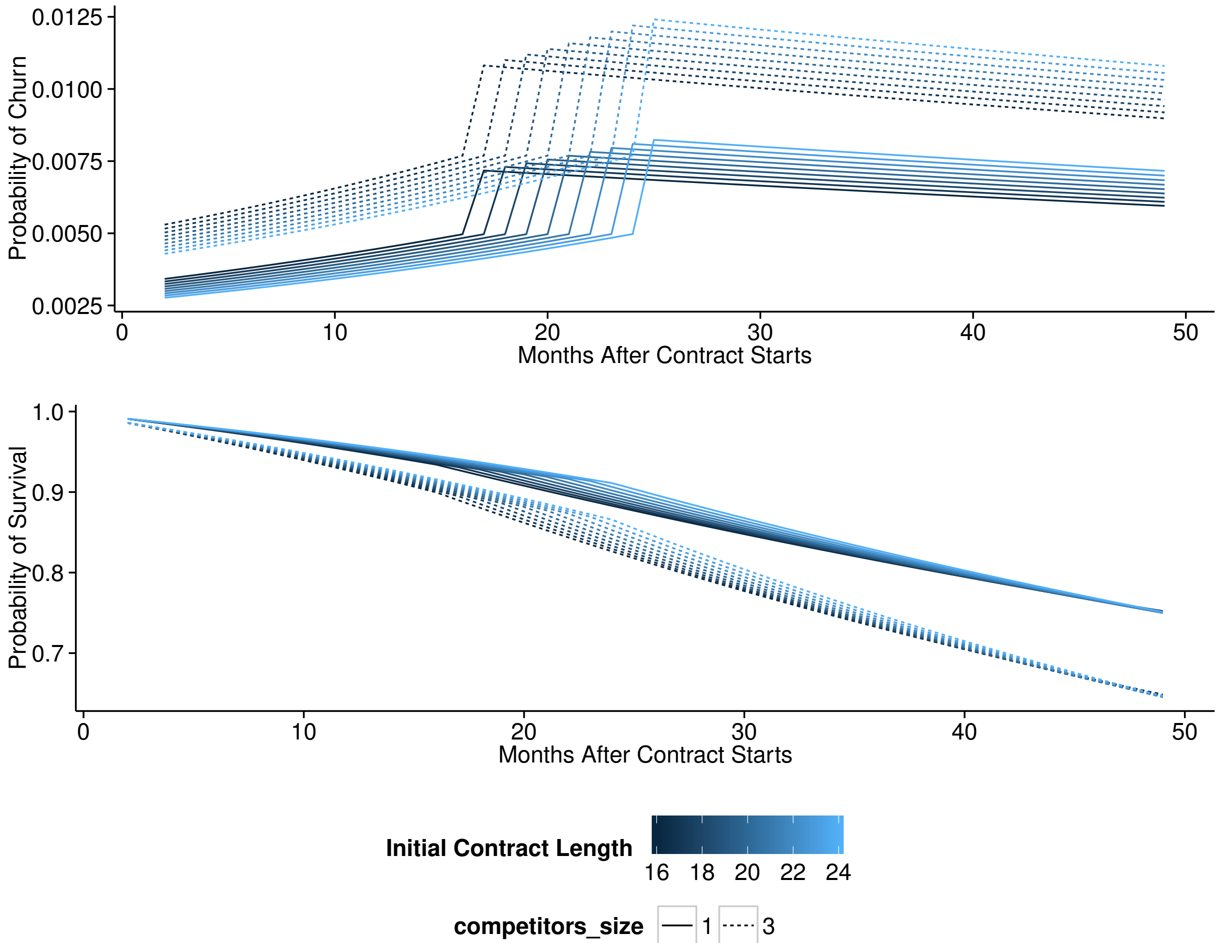


Figure A.1: Probability of churn (Top) and probability of survival ($Bottom$) when the length of the lock-in period varies according to the estimation results in column (1) of Table 2.3. Two market local competition scenarios were represented: solid curves correspond to market with one competitors (two service providers market) and dashed curves correspond to one with three competitors (four service providers market).

A.2 Complementary Policy Simulations

A.2.1 Price changes to maintain Consumer Surplus

We also study what increase in price would render consumers indifferent. Consider the case when the regulation allows firms to increase price in order to rescue part of their profit loss, but only to some extent so that consumer welfare is not reduced from the original level. Figure A.2 shows how the firm increases price to counter a reduction in the lock-in period to rescue part of profit loss while keeping consumer surplus constant and how firm profit is affected. Since the changes in consumer surplus substantially depends on their discount factors, the change in price needed also varies accordingly. When consumers only slightly discount the future (curves with ”round” dots), the firm can increase price by roughly 1% to counter a reduction in the lock-in period of 8 months in a three-competitor market. This price increase is more than half of the price increase when firms react to maintain profit if without price regulation. For example, if the lock-in period is reduced to 16 month and consumer discount rate is low (curves with ”round” dots), the firm would lose $13, 0.4% of its original profit from a consumer's lifetime if the lock-in period is 24 months. This would still save the firm more than half of the profit loss comparing to a case where no price increase is allowed (Figure 2.3). When consumer discount rate is high (curves with ”square” dots), however, firms are only allowed to slightly increase price to make consumers indifferent (0.25%), which limits firm's ability to recover profit. Policy interventions are less needed in the first place if consumers substantially discount future because their benefit from reduced switching cost is also limited. These results are also qualitatively similar across the rates of return that we simulate. Our analyses show how price regulation can pair with the lock-in period reduction to protect consumer welfare while to some extent protect the firm from losing much profit. The effectiveness of this policy may be sensitive to consumer discount rates.

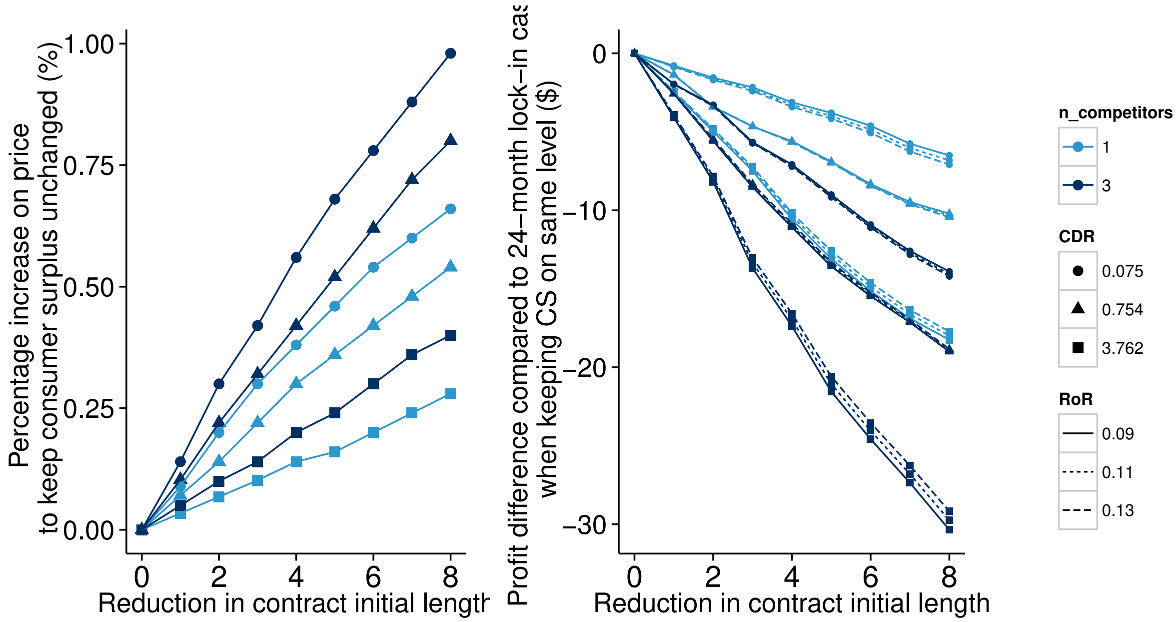


Figure A.2: Percentage change in price ($left$) and percentage change in firm profit ($right$) when the lock-in period reduces and the firm reacts to increase price up to making consumers indifferent to the status quo of 24 months lock-in period. Two market local competition scenarios were represented: red curves correspond to market with one competitors (two service providers market) and blue curves correspond to one with three competitors (four service providers market). Plots are based on simulations using estimation results in column (1) of Table2.3.

A.3 Additional Statistics

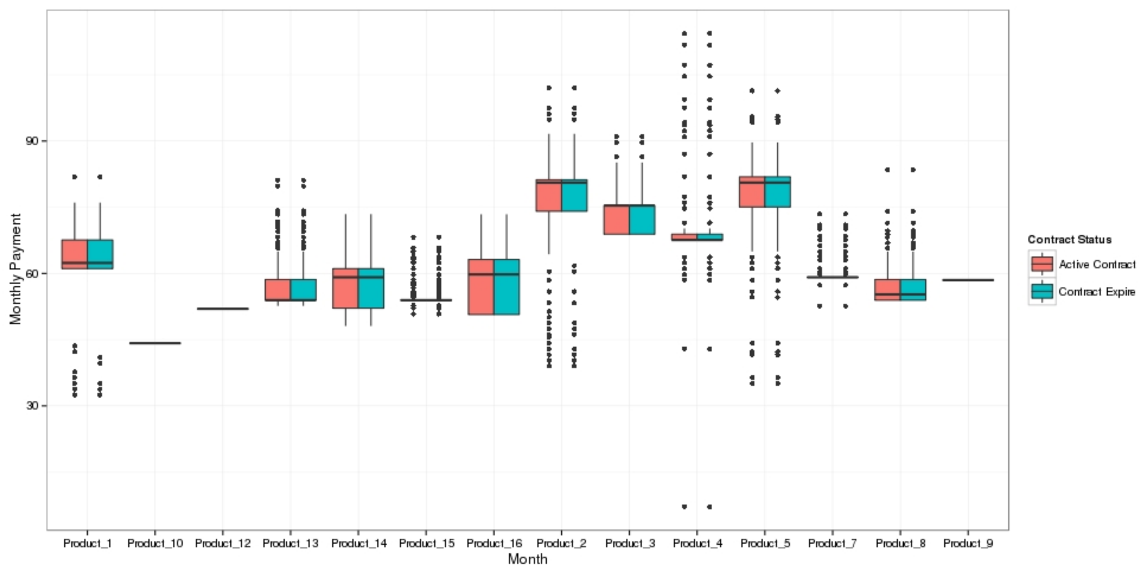


Figure A.3: Boxplots of prices of same products for consumers before and after contract expiry. The plots were broken down to each service products and the boxplots show the 5%, 50% and 95% quantiles.

Chapter B

Appendix for Chapter 3

B.1 Distribution of Uncertainty

In this appendix we discuss the distribution of uncertainty prior to search, followed by how to calculate reservation utilities theoretically and empirically.

Assume in Equation 3.1 z_j is of dimension k, and $z_j^{(n)}$ follows a distribution with $p.d.f.\ f^{(n)}(\cdot)$ and $c.d.f\ F^{(n)}(\cdot), 1 \leq n \leq k$. For calculation simplicity and taking advantage of the additive property of normal distribution, we assume that $z_j^{(n)} \sim N(0,1), 1 \leq n \leq k$. Empirically this can be achieved by normalizing z_j. Let Σ_z be the covariance matrix consisting of z_js. Then

$$z_j\gamma \sim N(0, \sigma_z^2)$$

, where

$$\sigma_z^2 = \sum_{n=1}^{k}(\gamma_{(n)}^2 \Sigma_{nn}) + 2\sum_{m=2}^{k}\sum_{n=1}^{m-1}\gamma_{(m)}\gamma_{(n)}\Sigma_{mn}$$

Specifically when z_j's are independently normally distributed, the variance of the search revealed term can be simplified as $(diag(\sigma\sigma^T)^T diag(\gamma\gamma^T))^2$.

B.2 Calculating Reservation Utility

The highest computational complexity is from the calculation of reservation utility especially when the dimension of products is big. We use an interpolation based approach by solving them outside the iterations similar with previous works in the literature (Kim et al., 2010; Chen and Yao, 2016; Ghose et al., 2018). From Equation 3.4, given the distribution of utility u_{ij} and searching cost c_{ij}, the corresponding reservation utility can be calculated by numerically solving the integral. However, reservation utility needs to be calculated in each iteration of parameter value updates during the optimization. It would be computational inefficient to conduct the calculations within iterations.

We use a interpolation based approach by solving them outside the iterations in a flavor similar to (Kim et al., 2010; Chen and Yao, 2016) but slightly different to account for information updates observed by researchers. In particular, following Equation 3.4 and let $V_{ij} = X_{ij}\beta + \mathbb{E}(z_j)\gamma$ to represent the determined part of utility prior to search and $f(\cdot)$ being the $c.d.f$ of uncertain part of utility:

$$c_{ij} = \int_{R_{ij}}^{\infty} (V_{ij} - R_{ij}) f(V_{ij}) d(V_{ij}) \tag{B.1}$$

From the above equation,

$$c_{ij} = (1 - F(R_{ij})) \int_{R_{ij}}^{\infty} (V_{ij} - R_{ij}) \frac{f(V_{ij})}{1 - F(R_{ij})} d(V_{ij}) \tag{B.2}$$

Specifically, when $f(\cdot)$ is normal, following the property of expectations of truncated normal distribution, we obtain

$$c_{ij} = (1 - \Phi(\eta_{ij}))(V_{ij} - R_{ij} + \frac{\phi(\eta_{ij})}{1 - \Phi(R_{ij})}) \tag{B.3}$$

where $\eta_{ij} = (R_{ij} - V_{ij})/\sigma_{ij}$ follows a standard normal distribution (Kim et al., 2010).

Therefore, for a given pair $\{c_{ij}, V_{ij}\}$, we can calculate the corresponding reservation utility R_{ij} by solving Equation B.3. We can simplify create this table outside of the estimation loop and then construct a look-up grid of the triple $\{c_{ij}, V_{ij}, R_{ij}\}$ to a substantially accurate level. When given $\{c_{ij}, V_{ij}\}$ inside an iteration, we use method of polynomial interpolation to approximate the reservation utilities.

B.3 Estimation Algorithm

We coded our own estimator which uses maximum simulated likelihood to estimate the utility and search cost parameters. Given a set of parameter estimates, the overall log likelihood for data can be calculated according to Equation 3.13. The algorithm optimizes the set of parameters iteratively through heuristic search. The maximization of the particular likelihood function, however, is a non-smooth optimization problem (Honka et al., 2016). Traditional newton-method based optimizer can result in local optimum and non-convergence. To solve similar problems, researchers either used downhill simplex method such as *Nelder-Mead* (Ghose et al., 2018), or applied a kernel smoother to approximate the original problem and solve it using Newton-based approach (Honka et al., 2016). However, the Nelder-Mead technique is a heuristic search method that can converge to non-stationary points (McKinnon, 1998). The kernel smoothing based optimizer requires finding proper smoothing parameters which may be computational costly and limit the generalizability.

We approach the optimization by comparing the choices of algorithms on various support of starting values. In our particular case, the best-performing algorithm is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. BFGS is a quasi-Newton method that has proven to have decent performance even for non-smooth optimizations (Lewis and Overton, 2013). With our data BFGS converges faster to the global optimal and is more robust to different reasonable starting values. We also tested BFGS with L-1 regularization as robustness check to avoid potential overfitting problem by minimizing

$$LL(\Theta) + \lambda|\Theta|_1 \tag{B.4}$$

We vary λ from 1 to 100 and didn't observe significant change to the estimates.

B.4 Model Performance and Alternative Models

To understand how the econometric model we have developed performs in describing consumer decision making, we compare the estimation results to those from other alternative benchmark models. In particular, since search decisions (which movies to search) and purchase decisions (which movie to purchase given the consideration set) were discrete choices, we compare the sequential search model with a set of logistic regression models. The logit model assumes binary choices of whether to search or purchase a movie.

The prediction of search and purchase probabilities in the developed sequential search econometric model can be achieved by substituting the model-estimated coefficients into Eq 3.9 and calculate $L_{i\lambda}^{search}$ and $L_{i\lambda}^{purchase}$.

We use four metrics to measure the model predictive performances on the test samples: ”Rooted Mean Squared Error” (RMSE), ”Mean Absolute Error” (MAE), ”ROC curve”, and ”the Area under ROC curve” (AUC). We randomly partition our dataset into two subsets: one with observations from 70% of the total search-purchase sessions as the training sample and the other with 30% of the search-purchase sessions as the test (holdout) sample. We perform a 10-fold cross validation to avoid potential biases from the partitions.

Table B.1 shows the estimation results using the benchmark logistic regression models. The logit modes used exactly same predictors and transformations as used in the sequential search estimation in Table 3.4. From left to right, the models correspond to the estimations for search decisions among all movies, purchase decisions among all movies, and purchase decisions given the consideration set, respectively.

The predictive performance comparisons are summarized in Table B.2. Overall, we find that our sequential search model prediction has higher predictive power than the alternative logistic regression model. The sequential search model outperforms the baseline models in out-of-sample predictive performances with respect to RMSE, MAE, and AUC. For example, when predicting which movies to search, the sequential search model represents a 20.5% decrease in RMSE, and 10.5% increase in AUC. When predicting which movies to purchase conditional on the consideration set, the sequential search model improves the predictive performance by reducing the RMSE by 8.4% and increasing the AUC by 23.2%. Figure B.1 shows the corresponding ROC curves.

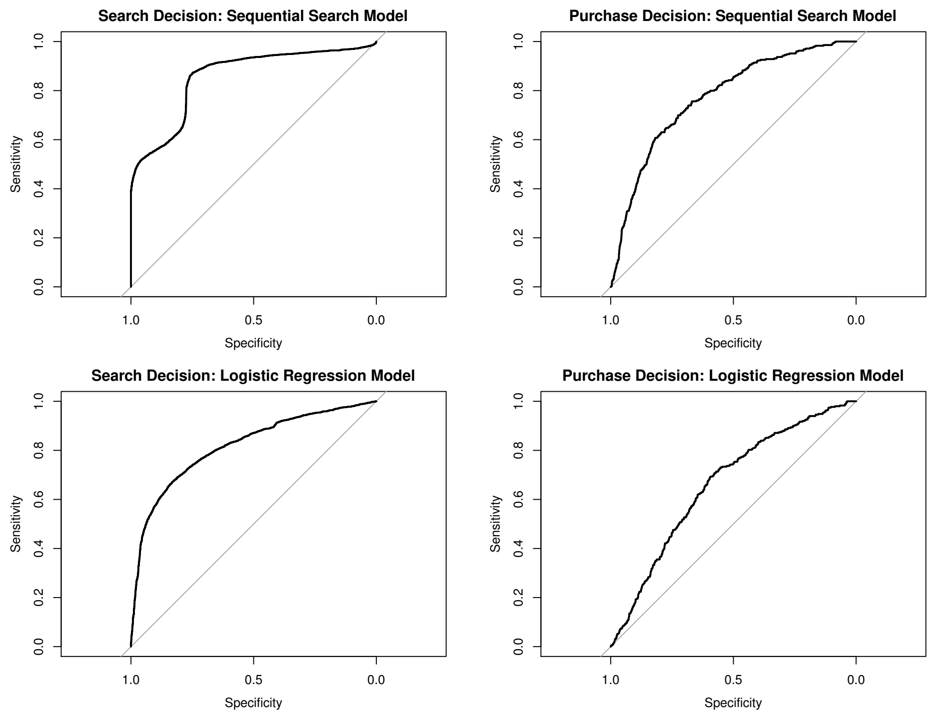


Figure B.1: Comparison of the ROC curves

Table B.1: Alternative Model: Logit

	Dependent variable:		
	Search	Purchase	Purchase (conditional)
price		-0.500^{***}	-0.435^{***}
		(0.045)	(0.066)
N_likes	0.225^{***}	0.827^{***}	0.332^{***}
	(0.010)	(0.035)	(0.034)
N_FrdRental	0.489^{***}	0.642^{***}	0.321^{***}
	(0.034)	(0.067)	(0.097)
IMDB Rating	0.018	0.260^{***}	0.149^{***}
	(0.011)	(0.044)	(0.043)
IMDB Votes	0.016	-0.207^{***}	-0.097^{**}
	(0.011)	(0.045)	(0.047)
Movie Age	-0.144^{***}	-1.105^{***}	-0.074
	(0.014)	(0.092)	(0.057)
N_CumSearch	0.647^{***}	0.128^{***}	-0.265^{***}
	(0.016)	(0.015)	(0.077)
Available Window on VoD	-0.252^{***}	-0.697^{***}	-0.188^{***}
	(0.008)	(0.033)	(0.030)
Days Since Available VoD	-0.210^{***}	-0.745^{***}	-0.056
	(0.023)	(0.113)	(0.068)
Appearance Frequency	0.255^{***}		
	(0.004)		
Slot within menu	-0.953^{***}		
	(0.025)		
Constant	-6.076^{***}	-5.504^{***}	-1.473^{***}
	(0.065)	(0.243)	(0.207)
Menu Dummies	Yes	Yes	Yes
No. of Sessions	5,160	5,160	5,160
Log Likelihood	−90,918.000	−10,461.000	−3,890.000
Akaike Inf. Crit.	181,860.000	20,942.000	7,800.000
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.2: Comparison in predictive metrics: sequential search model versus logistic regression model

metrics	Sequential Search Model (Search Decision)	Logistic Regression Model (Search Decision)	Sequential Search Model (Purchase Decision)	Logistic Regression Model (Search Decision)
RMSE	0.031	0.039	0.296	0.323
MAE	0.001	0.003	0.088	0.105
AUC	0.855	0.774	0.816	0.662

B.5 Additional Descriptive Statistics

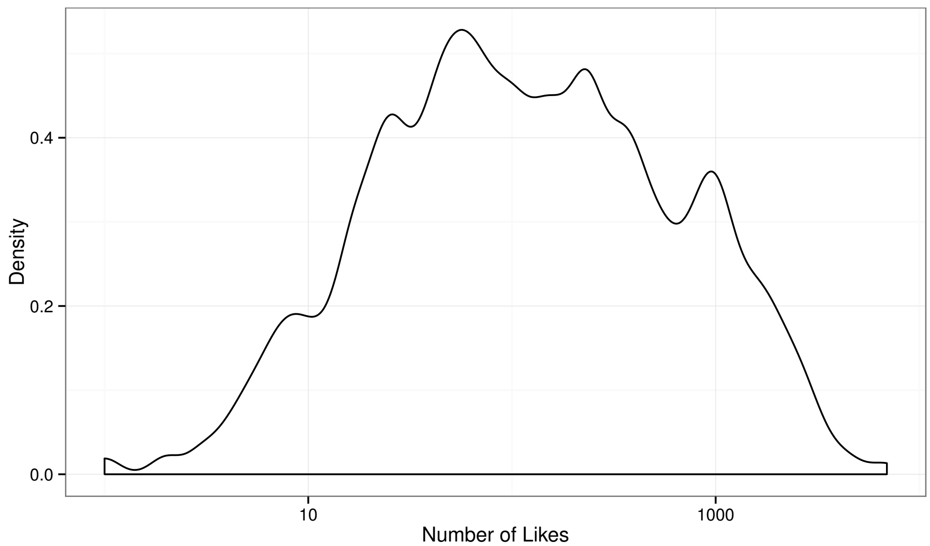


Figure B.2: Distribution of ”Number of Likes” for movies when in front of consumers. The x-axis is in Log-scale.

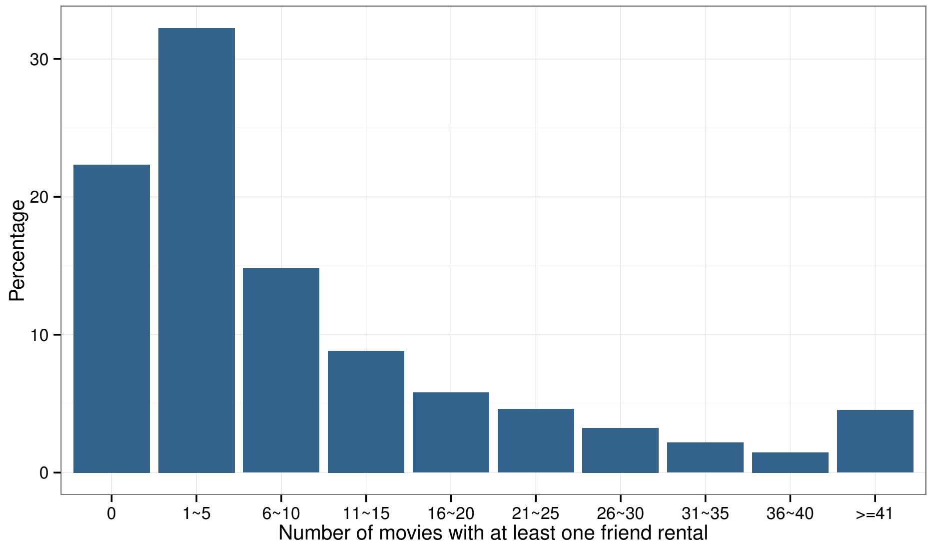


Figure B.3: Distribution of number of movies (among the 1771 available movies) that had at least one friend rentals when search session begins for all the users studied.

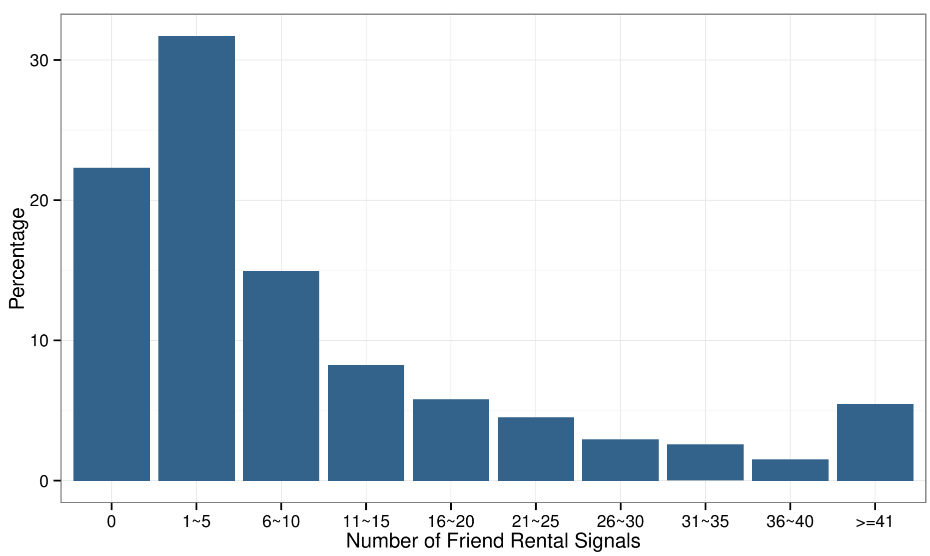


Figure B.4: Distribution of number of "Friends Rental Signals" that consumers could observe at the beginning of the search session. The histogram is quite similar to Figure B.3 as the friend rental signals were dispersed among movies.

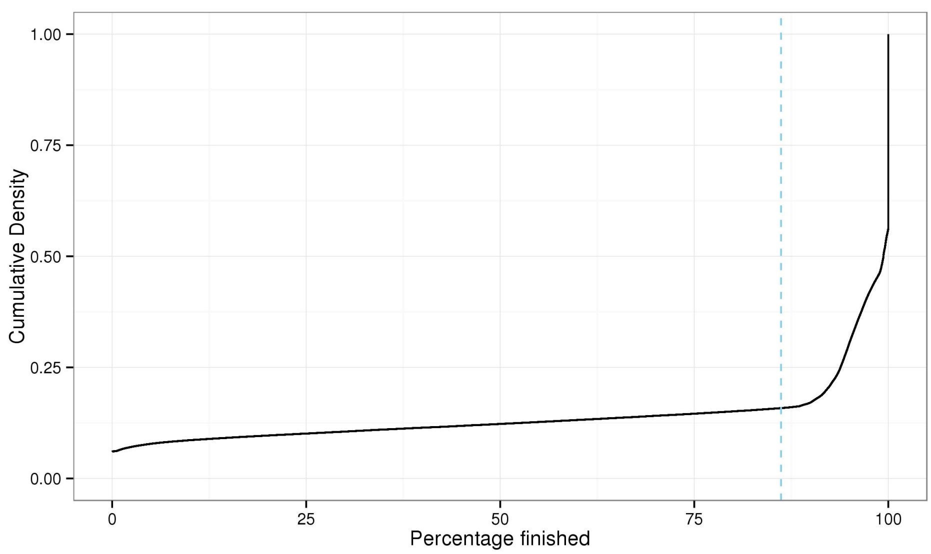


Figure B.5: Cumulative Density Plot for the proportion of movie streamed by consumers who rented. The dashed line corresponds to 85% of the movie runtime being streamed (cumulatively after rentals).

Chapter C

Appendix for Chapter 4

C.1 Ethics Statement

The study introduced in this chapter was approved by the Carnegie Mellon University Institutional Review Board (IRB) as Exempt in accordance with 45 CFR 46.101(b)(2) under IRB ID "STUDY2018_00000122: Modeling consumer digital shopping journey".

Prior to participating, all subjects were acknowledged the following Informed Consent Statement. The consent statement appears as the first page in the Qualtrics survey. Subjects can proceed to the survey only when they read and agreed with the terms and conditions. The consent statement reads as the following:

INTRODUCTION

Consumer online shopping journeys are likely guided by different sources of information. This study examines consumer movie shopping behaviors in an online video-on-demand platform. The study is led by Baojiang Yang from Department of Engineering and Public Policy at Carnegie Mellon University. You are recruited as a participant via Amazons Mechanical Turk Service (www.mturk.com).

PURPOSE

The purpose of the study is to examine consumer behaviors when they are shopping for movies in an online video-on-demand platform.

COMPENSATION

For your complete participation, you will be paid with a task completion reward ($1.00). Additionally, we will automatically enter you into a lottery to actually obtain the movie you selected on our platform. If the movie costs less than $10 on our platform, we will pay you for the unused portion of $10 through HIT bonus. Note that in order to receive your lottery winnings (if you win), you will need to provide us with your email address.

PROCEDURE

If you agree to participate, you will first read an example telling you what you will do in this study. In this study, you will be guided to a secured website and choose a movie you would like to rent. This is a safe website that we constructed specifically for this research project. You will need to answer a few survey questions after your movie choice. People typically take approximately 5 to 10 minutes to complete the session. Once you finish the survey, you will be given a code to confirm your participation on Amazons Mechanical Turk.

VOLUNTARY NATURE

Your participation in our study is completely voluntary. If at any point during the experiment, you do not wish to continue, you are free to leave without any penalty. Deciding to participate or not will not affect your standing or relationship with Carnegie Mellon University.

ANONYMITY

All of your responses will be anonymous. Only the researchers for this study will have access to your responses. However, please be aware that (a) do not leave survey open if using a public computer or a computer others may have access to, (b) it is best to clear the browser cache and page history (see your browser instructions) after you complete the study.

BENEFITS

There may be not personal benefit from your participation in the study but the knowledge received may be of value to humanity.

RISKS

The risk of participation is no greater than that experienced in everyday life. This means that you won't be taking any risks by choosing to participate in this study.

CONTACTS

If you have questions about the study, you may contact us at MovieStudyCMU@gmail.com. You should direct any questions or concerns about your rights as a research participant to the IRB of Carnegie Mellon University by email at irb-review@andrew.cmu.edu.

Please go to the next page ONLY if you consent to the procedures described above. In other words, if you complete the session by going to the following pages, that means you consent to the procedures described above.

C.2 Experiment Implementation

C.2.1 Survey Structure

The randomized experiments are embedded in a *Qualtrics* survey. The survey consists two parts: *the core movie shopping task* and *post-task survey questions*. The survey instructions were given in the very beginning of the survey. It will tell participants the survey structure and reward calculation rules.

When accessing the core movie shopping task, a link from *Qualtrics* directed the subjects to the *social network impression page*, from where subjects click into *MoviePlatform*. After task completion, the subjects were guided back to the *Qualtrics* survey to answer a few exit survey questions. Successfully finishing the supplementary questions will lead the participant to the end of the survey. A HIT completion page will show up, where a personalized HIT completion code will be displayed. A participant would need to use the code in *MTurk* to be counted as accomplished the HIT.

C.2.2 Reward calculation

In the study, we follow the experimental economics literature to design the reward mechanism so that it is incentive compatible. (Smith, 1976; Smith and Walker, 1993; Ding, 2007; Miller et al., 2011). Particularly, we use a randomizing mechanism to determine which of the participants choices were fulfilled. This process is called random lottery procedure and is used widely in experimental economics (Starmer and Sugden, 1991; Becker et al., 1964; Ding, 2007).

In this section, we formally define the reward calculation method. The total reward a participant receives consists two parts: a guaranteed participation fee and a lottery reward. We leverage the latter to create incentives for subjects to behave close to real life.

Mathematically, let q be the final rewards, w be the participation fee, and M be the "initial" credit we give to the participants if winning the lottery. p stands for the price associated with the movie he/she rented on $MoviePlatform$. ρ being the probability winning the lottery. u represents the monetary valuation the participant puts on the movie that he/she rented through $MoviePlatform$ (can be thought of as her willingness-to-pay for the movie). The expected total reward a participant i gets is the following (suppose he rented j on MoviePlatform):

$$\mathbb{E}(q_{ij}) = w + \rho(M - p_{ij} + u_{ij}) \tag{C.1}$$

or

$$q_{ij} = \begin{cases} w + M - p_{ij} + u_{ij}, & \text{if won lottery and rented movie with price } p \\ w + M, & \text{if won lottery but didn't purchase any movie} \\ w, & \text{otherwise} \end{cases} \tag{C.2}$$

Lottery winners receive initial movie purchase funding/credits M, which can be considered obtained by the subjects as "movie rental budget". Since the participants know whether they were selected in the lottery only after they made their rental choices on $MoviePlatform$, we expect they are aware that their choices are consequential throughout the shopping process.

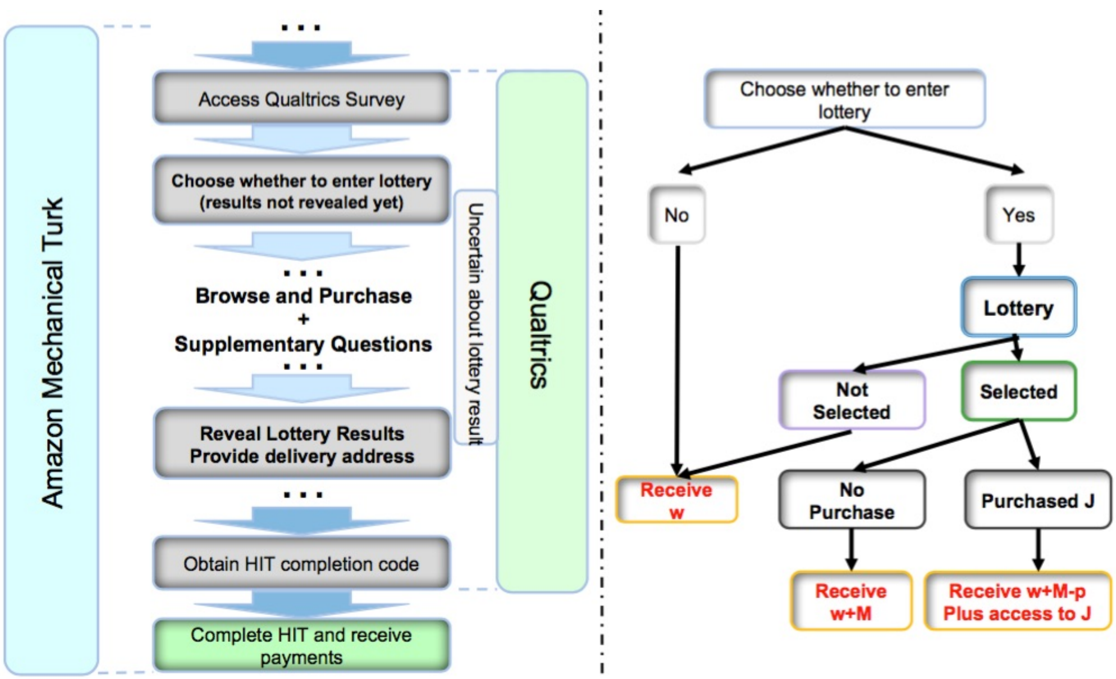


Figure C.1: Left shows how lottery is embedded in Qualtrics survey. Right shows how final payment is calculated.

C.3 Alternative Models: Logistic regression

Table C.1: Alternative Model: Logistic Regression

	Dependent variable:		
	search	purchase	purchase (conditional)
Price		-0.099^{***}	-0.294^{***}
		(0.035)	(0.052)
N_likes	0.029	0.138^{**}	0.185^{**}
	(0.036)	(0.054)	(0.077)
N_FrdRental	0.294^{***}	0.358^{***}	0.089
	(0.044)	(0.061)	(0.086)
IMDB Rating	0.028	0.138^{*}	0.222^{**}
	(0.054)	(0.077)	(0.101)
Genre Drama	0.667^{***}	0.856^{***}	0.481^{**}
	(0.101)	(0.147)	(0.204)
Genre Comedy	-0.307^{***}	-0.142	-0.049
	(0.100)	(0.141)	(0.204)
Genre Action	-0.030	-0.150	-0.272
	(0.119)	(0.168)	(0.249)
Genre Family	-0.687^{***}	-0.921^{***}	-0.253
	(0.153)	(0.212)	(0.319)
ratingAmazonCNT	1.344^{***}	2.421^{***}	2.354^{***}
	(0.282)	(0.415)	(0.642)
Movie Age	0.022^{***}	0.002	-0.027^{*}
	(0.008)	(0.012)	(0.015)
Menu_Order	-0.286^{***}		
	(0.035)		
Constant	-0.082	0.985	4.863^{***}
	(0.602)	(0.880)	(1.365)
No. of Sessions	483	483	483
Log Likelihood	$-3,383.000$	$-1,896.000$	-550.200
Akaike Inf. Crit.	6,787.000	3,815.000	1,122.000
Note:			*p<0.1; **p<0.05; ***p<0.01

C.4 Supplementary survey questions

We complement the randomized lab experiment with a series of exit survey questions,. The questions were designed to measure the transparency of experiment requirements and validity of incentive design, understand mechanisms why people have combined or not combined social signals during the shopping task, and collect demographic information. In addition to Fig. **??** we introduced earlier to measure the study transparency, the following graphs in this section display two other questions that we asked in the survey and the corresponding distribution of the participants' answers.

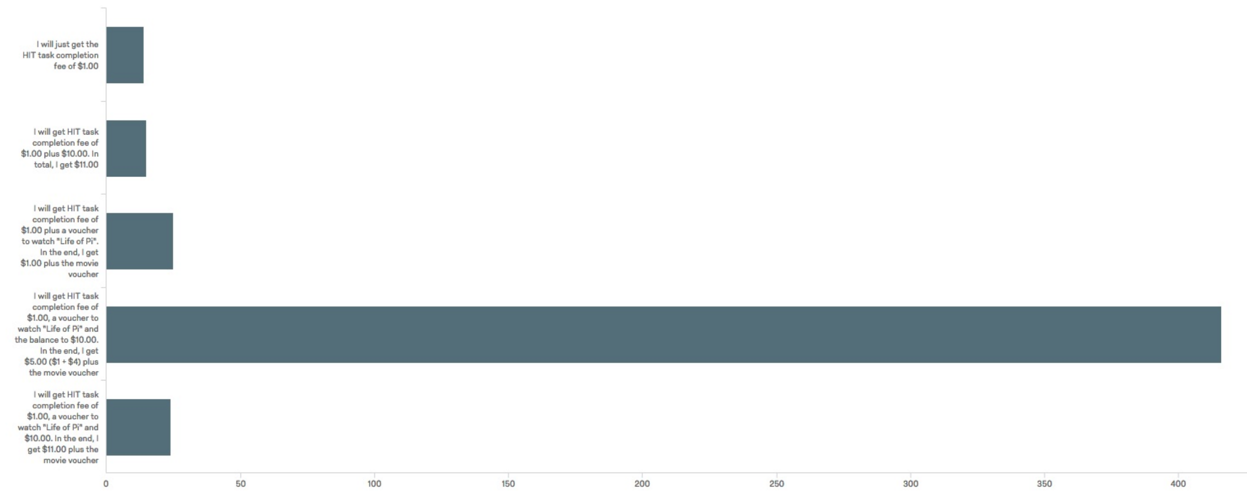


Figure C.2: Quiz to evaluate whether participants understood the reward calculations. The participants were asked ”Suppose you rent movie ”Life of Pi” priced at $6.00 at MoviePlatform, you win the lottery and you complete your task. What would you get as a reward for your participation in the study?”. Five options were provided in random order for participants to choose from (respective to the five options listed vertically in the graph). (1) I will just get the HIT task completion fee of $0.50; (2) I will get HIT task completion fee of $0.50 plus $10.00. In total, I get $10.50; (3) I will get HIT task completion fee of $0.50 plus a voucher to watch ”Life of Pi”. In the end, I get $0.50 plus the movie voucher; (4) I will get HIT task completion fee of $0.50, a voucher to watch ”Life of Pi” and the balance to $10.00. In the end, I get $4.50 ($0.50 + $4) plus the movie voucher; (5) I will get HIT task completion fee of $0.50, a voucher to watch ”Life of Pi” and $10.00. In the end, I get $10.50 plus the movie voucher. Each participant can choose a single answer. Here Option (4) here is the right answer. When a participant missed the right answer, we reaffirm the calculation by giving the explanation of the right answer.

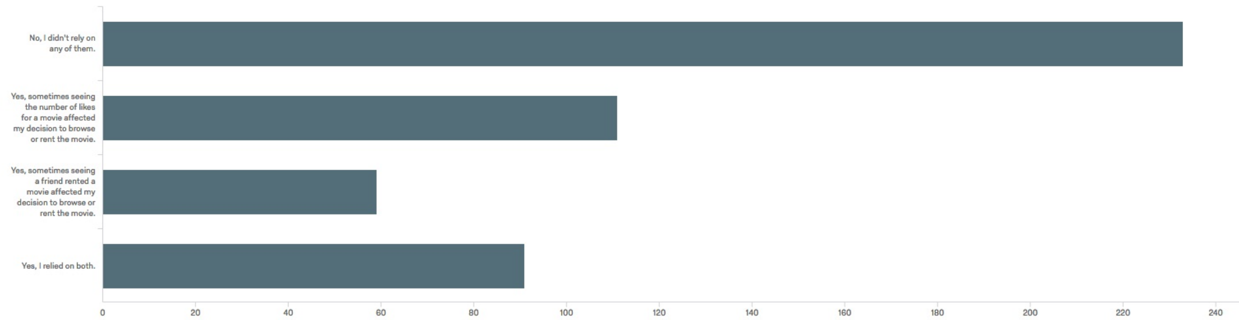


Figure C.3: In this survey question we asked whether the participants had relied on any of the social information displayed for a movie, including the number of likes and number of friends rentals. We showed an example of where these signals appear on *MoviePlatform* to remind the participants their shopping experiences and in particular direct them to the signals appeared specifically on the platform. Then we asked ”While browsing movies at MoviePlatform, did you rely on information about how many people liked a movie and how many of your friends have rented a movie?” Four options were provided in randomized orders to each participant: (1) No, I didn’t rely on any of them; (2) Yes, sometimes seeing the number of likes for a movie affected my decision to browse or rent the movie; (3) Yes, sometimes seeing ¡strong¿a friend rented ¡/strong¿a movie affected my decision to browse or rent the movie, and (4) Yes, I relied on both. The options are exclusive. The plot shows the number of selections to each options respectively.

C.5 Demographic Characteristics of Experiment Participants

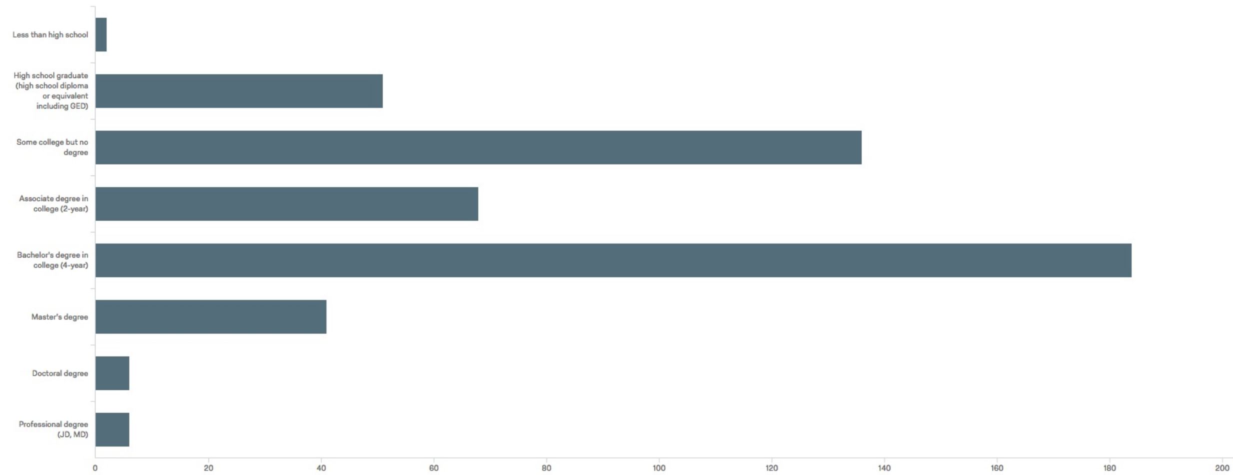


Figure C.4: Distribution of participants' educational background

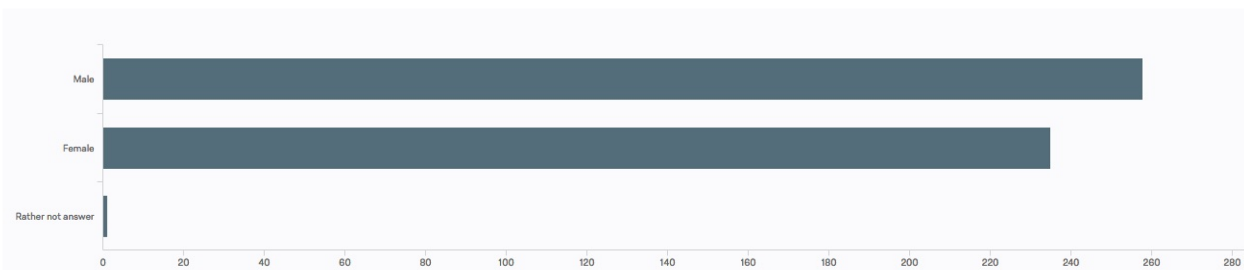


Figure C.5: Distribution of participants genders

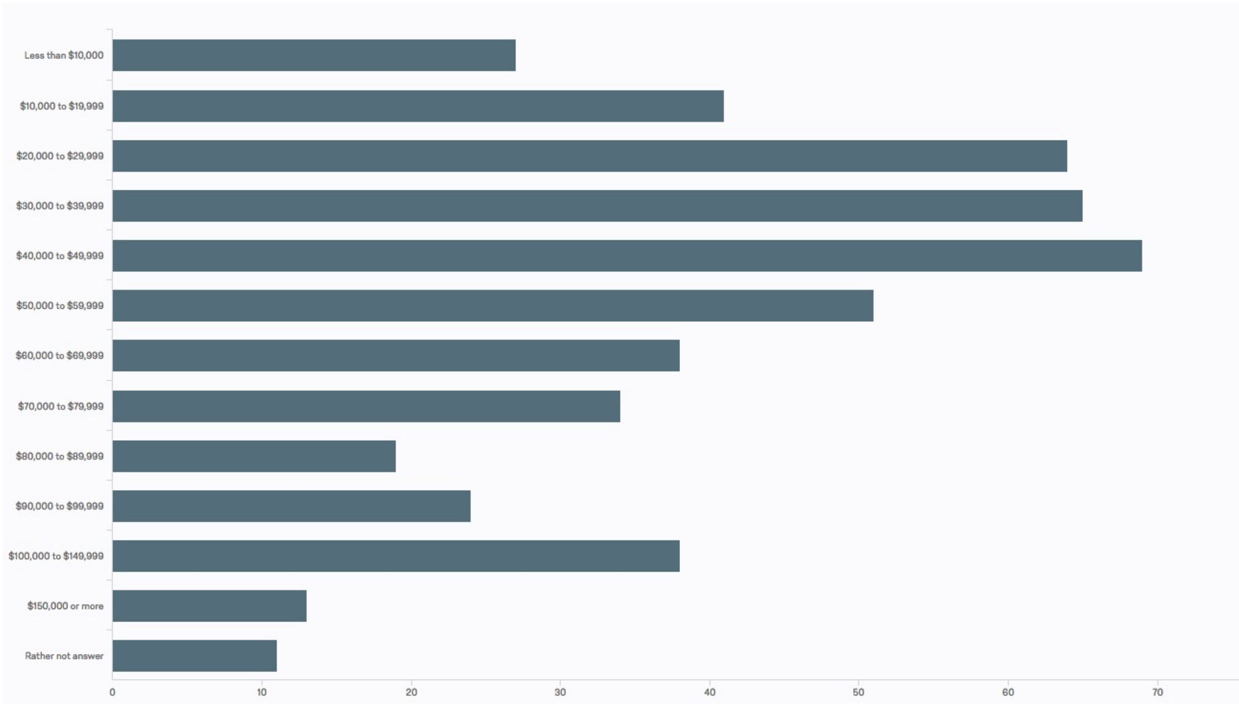


Figure C.6: Distribution of participants household annual income in USD

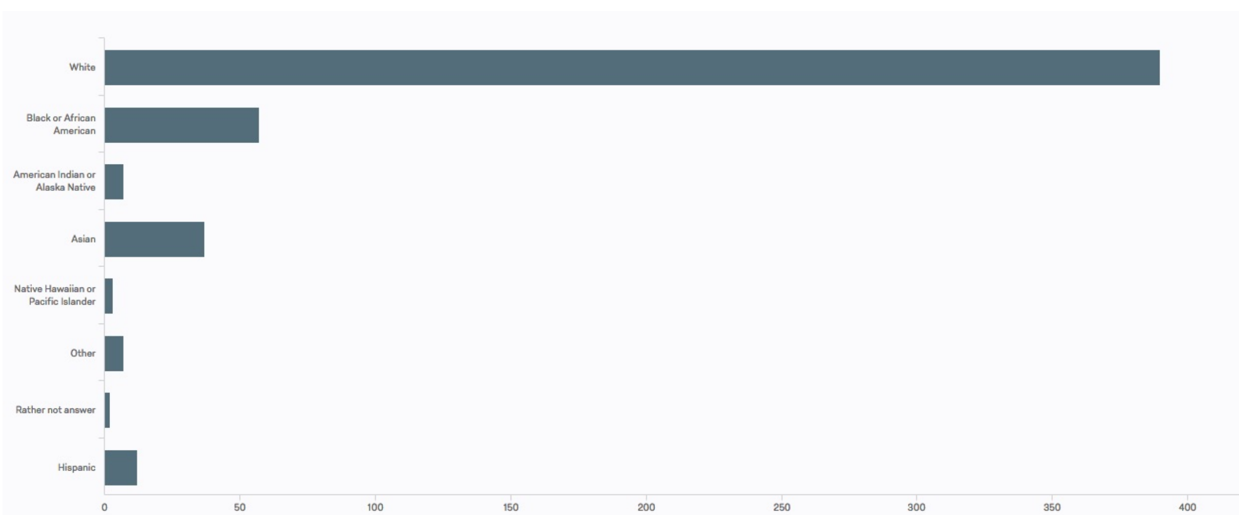


Figure C.7: Distribution of participants races