7-27-2016

Essays on the Interactions of Consumer Behavior and Firm Strategy in Multi-Channel Environments

Bin Li
bin.li2@business.uconn.edu

Essays on the Interactions of Consumer Behavior and Firm Strategy in Multi-Channel Environments

Bin Li, PhD

University of Connecticut, 2016

Abstract

With the increasing popularity of the online channel, both consumers and firms are engaging in more and more multi-channel activities. On the one hand, consumers can integrate information from searches on both online and offline channels, and then decide on the best channel to buy from. On the other hand, firms need to consider consumer behavior in different channels in their strategy design. As a result, cross-channel interactions between consumer behavior and firm strategy can be within the same channel or across different channels. While the within-channel interaction has been studied extensively in the previous literature, there is much less research on the cross-channel interaction. In my dissertation, I add to the understanding of consumer behavior and firm strategy in the multi-channel environment by empirically analyzing their cross-channel interactions.

This dissertation consists of three separate but related essays. The first answers the question: How does consumer behavior affect optimal product portfolio strategies in online versus offline channels? I develop an empirical model to simultaneously identify the cannibalization effect (within a brand) and the competition effect (between different brands) in different retail channels. I further examine how these effects are affected by consumer preferences. The second essay answers the question: How does a

firm's offline strategy affect consumer online behavior? I use a natural experiment to examine how the awareness and convenience effects from opening new retail stores affect the online search. The final essay answers the question: How does online banking affect entry/exit of offline bank branches? I develop and estimate a dynamic entry/exit model examining the relationship between technological advances and market structure evolution. My counterfactual analysis shows that the asymmetric reduction in operating costs is the most significant factor driving recent changes in the U.S. banking industry, followed by increased entry costs and increased deposits for large banks due to greater online presence. My findings provide important implications for firms engaging in multi-channel activities.

Essays on the Interactions of Consumer Behavior and Firm Strategy in Multi-Channel Environments

Bin Li

B.S., Tsinghua University, 1991

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

[2016]

APPROVAL PAGE

Doctor of Philosophy Dissertation

Essays on the Interactions of Consumer Behavior and Firm Strategy in Multi-Channel Environments

Presented by

Bin Li, B.S.

Major Advisor ____

Honju Liu

Associate Advisor ____

Joseph Pancras

Associate Advisor ____

Ting Zhu

Associate Advisor ____

Nicholas Lurie

Associate Advisor ____

William T. Ross, Jr.

University of Connecticut

2016

Acknowledgments

This dissertation could not have been completed without all kinds of advice, guidance, and support from so many people over the past years. I wish to offer my heartfelt thanks to them.

First, I will be forever grateful to my five co-advisors: Dr. Hongju Liu, Dr. Joseph Pancras, Dr. Ting Zhu, Dr. Nicholas Lurie, and Dr. Bill Ross. They have guided me through the doctoral program using their experience in academia, professional knowledge in various areas, and endless encouragement. As the chairman of my doctoral committee, Dr. Liu means a lot to me: a lifetime friend, a gentle guiding hand, a watchful observer, and a silent but forceful motivator. The door to Dr. Pancras is always open to me, and he is patient to discuss and chat on anything I am interested in. The support and encouragement provided by Dr. Zhu cannot be overstated. Her unhesitant willingness to advise a doctoral outside of her institution merely for the sake of scholarship is an example worthy of long admiration. Dr. Ross inspired my research with his wonderful sense of humor and his eternal optimism. He is on top of my mind whenever I have difficulties in the PhD program. I would offer special thanks to Dr. Lurie. My journey in Marketing would not have started without you! Thank you for your continuous encouragement and suggestions in so many years.

I offer my sincere appreciation to Dr. Robin Coulter, Dr. Girish Punj, Dr. Susan Spiggle, Dr. Narasimhan Srinivasan, Dr. Jane Gu, Dr. Kunter Gunasti, Dr. David Norton, and Dr. Hee Mok Park for providing support and valuable suggestions to my research. In UConn Marketing, I have lots of good memories that I will remember forever.

I would like to take this opportunity to express my very deep gratitude to Dr. Xinxin Li from OPIM Department for her inspirations and support to make this dissertation possible. It is always my pleasure to work with her.

I would especially like to thank my family. To my mom, you understanding and endless love made me who I am today. My wife, Junjuan, has been extremely supportive of me throughout this entire process. Thank you for your countless sacrifices to our family and help me get to this point. To my children, Julia and Kelly, you are the pride of my life, and the ultimate motivation to me in finishing my degree.

This dissertation is dedicated to the memory of my dad. I miss him every day!

Table of Contents

Chapter 1 Introduction **1**

Chapter 2 Consumer Preferences, Cannibalization and Competition: Evidence from the Personal Computer Industry **4**

2.1 Introduction 5

2.2 Literature Review 9

2.3 Data Description 11

2.4 Empirical Model 12

2.4.1 Cannibalization and Competition Effects 12

2.4.2 Consumer Preferences 16

2.5 Estimation 20

2.6 Results 22

2.6.1 Cannibalization and Competition Effects 22

2.6.2 Consumer Preferences 25

2.6.3 Discussion 28

2.7 Robustness Checks 32

2.8 Conclusions 33

Chapter 3 Retail Store Entry and Online Consumer Search: The Role of Awareness versus Convenience Effect **36**

3.1 Introduction 37

3.2 Theoretical Framework 39

3.2.1 The Convenience Effect 40

3.2.2 The Awareness Effect 41

3.2.3 The Moderating Role of Consumer-Specific Characteristics 42

3.3 Data 44

3.3.1 Search Behavior 44

3.3.2 Store Entry and Location 45

3.3.3 Online Channel Efficiency 46

3.4 Research Methodology 47

3.4.1 Natural Experiment Design 47

3.4.2 Marketing and Shopping Regions 48

3.4.3 Model Specification 50

3.4.4 Matching 50

3.5 Empirical Results 53

3.5.1 Analysis on the Within Session Behaviors ... *59*
3.6 Conclusion and Implications ... 60
3.6.1 Managerial Implications ... *61*
3.6.2 Limitations and Future Research ... *62*
Chapter 4 Technology and Market Structure: An Empirical Analysis of Entry and Exit in the Banking Industry ... 64
4.1 Introduction ... 64
4.2 Literature Review ... 68
4.3 The U.S. Banking Industry ... 70
4.4 Data ... 72
4.4.1 Sample Markets ... *72*
4.4.2 Banks and Branches ... *74*
4.4.3 Broadband Penetration ... *76*
4.4.4 Descriptive Analysis ... *77*
4.5 Model ... 78
4.5.1 Deposit Demand ... *79*
4.5.2 Cost Structure ... *81*
4.6 Empirical Strategy ... 86
4.6.1 Policy Functions ... *86*
4.6.2 Estimator ... *87*
4.6.3 Computational Details ... *89*
4.7 Results ... 91
4.7.1 Deposit Demand Parameters ... *91*
4.7.2 Structural Parameters for Entry and Exit Decisions ... *94*
4.8 Counterfactual Simulations ... 96
4.9 Conclusion ... 100
Chapter 5 Conclusions and Future Research ... 102
Bibliography ... 103
Appendix 109
Appendix A. Selection of Nest Structure (Chapter Three) ... 109
Appendix B. GMM Estimator (Chapter Three) ... 110

List of Figures

Figure 1. Interactions between consumers and firms in the multi-channel environment 1

Figure 2：Nest Structure of Our Two-level Nested Logit Model 13

Figure 3: Model Free Evidence of Higher Cannibalization Online 29

Figure 4：Comparison between Online and Offline Channels: Percentage Increase in Sales as the Number of Products Increases 31

Figure 5. Spatial distribution of the sample markets 73

Figure 6. Changes in the number of branches.. 75

Figure 7. Comparing the distributions of deposit and log(deposit) 80

List of Tables

Table 1: Descriptive Statistics 20

Table 2: Correlation Matrix 21

Table 3: Estimation Results 23

Table 4: Cannibalization Parameter for Each Sub-Nest from Model M1 24

Table 5: Competition Parameter for Each Nest from Model M1 25

Table 6: Explaining Power of Loyalty and Search Variables... 27

Table 7: Robustness Checks 32

Table 8: Data Sources 44

Table 9: Control and Treatment Groups 48

Table 10: Comparison of Geo-Information Before and After Full Matching 53

Table 11: Descriptive Statistics 53

Table 12: Correlation Matrix 55

Table 13: Number of Visits the Online Store 55

Table 14: Number of Pages per Visit.. 59

Table 15. Summary statistics of demographic variables 73

Table 16. Changes in the number of branches for all the markets... 74

Table 17. Changes in the number of branches for the sample markets 75

Table 18. Broad band penetration.. 76

Table 19. HHI of bank level deposits 77

Table 20. Regression analyses on the number of branches and broadband penetration............... 77

Table 21. Estimates for deposit demand.. 92

Table 22. Estimates for structural parameters... 94

Table 23. Counterfactual simulations for entry/exit probabilities 97

Table 24. Counterfactual simulations for the number of branches 98

Chapter 1 Introduction

The emergence and popularity of the online channel change the way consumers search for information and make decisions in online and offline environments. This has important implications for firms' multi-channel strategies and well as the evolution of markets. On the one hand, consumers can integrate information from searches on both online and offline channels, and then decide on the best channel to buy from. On the other hand, firms need to consider consumer behavior in different channels in their strategy design. As a result, the cross-channel interactions between consumer behavior and firm strategy can be within the same channel or across different channels as shown in Figure 1.

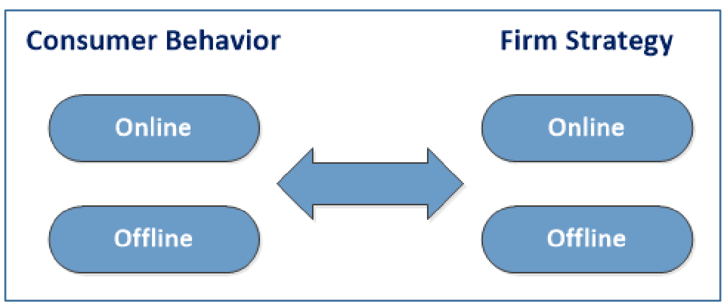


Figure 1. Interactions between consumers and firms in the multi-channel environment

While the within-channel interaction has been studied extensively in the previous literature, there is much less research on the cross-channel interaction. In my dissertation, I add to the understanding of consumer behavior and firm strategy in the multi-channel environment by empirically analyzing their cross-channel interactions from three perspectives.

In the first essay, I develop an empirical model to simultaneously identify the cannibalization effect (within a brand) and the competition effect (between different brands) in different retail channels. I further examine how these effects are affected by consumer preferences. Our results show that the online market exhibits stronger cannibalization and

competition than the offline channel. Moreover, heterogeneity in consumer search behavior and brand loyalty explain a significant fraction of the variation in both cannibalization and competition between the two channels. My findings suggest that firms should offer fewer models in the online channel and when consumers are highly loyal.

In the second essay, I examine how opening a new retail store affects the cognitive costs of online search and the physical cost of offline search. I find that, for consumers with prior experience on the retailer's website, opening a new store leads to a 68% increase in the number of visits to the retailer's website by consumers who live in the broader marketing area of the new store, while it leads to a 49% decrease in the number of visits to the retailer's website by consumers who live in the nearby shopping region. More interestingly, the effect of store entry in the shopping region on decreasing website search is weaker when consumers are more efficient in using the online channel. My results are robust to corrections for endogeneity in the choice of new store locations using different matching methods.

In the third essay, I develop and estimate a dynamic entry/exit model examining the relationship between technological advances and market structure evolution to understand the relationships between these intriguing phenomena. Despite the rise of consumer online banking, there has been little reduction in the number of brick and mortar bank branches in the U.S. At the same time, large national banks have expanded their branch networks at the cost of small local banks. My findings suggest that the advent of online banking provides significant competitive advantages to large national banks over small local banks that lower large banks' offline operating costs and increase consumer deposits; yet increase entry costs. A counterfactual analysis shows that the asymmetric reduction in operating costs is the most significant factor

driving recent changes in the U.S. banking industry, followed by increased entry costs and increased deposits for large banks due to greater online presence.

In sum, the results presented in this dissertation provide insight into the multi-channel business models. These studies also contribute to the literature on product portfolio management and local competition.

Chapter 2 Consumer Preferences, Cannibalization and Competition: Evidence from the Personal Computer Industry

Understanding the degree of cannibalization and competition in online and offline markets is important to firms' product line designs. However, few empirical studies have measured both effects simultaneously or have examined the factors that determine the extent of cannibalization and competition. In this study, we develop an empirical model to identify cannibalization and competition effects simultaneously in different markets, and further examine the impacts of consumer preferences on these two effects in a single integrated framework. Using data from the U.S. personal computer (PC) industry, we find that the online market exhibits stronger cannibalization and competition than the offline market. Both effects are significantly influenced by consumers' search behavior and brand loyalty. Specifically, more active consumer search not only intensifies inter-brand competition but also amplifies intra-brand cannibalization. In addition, search has a higher impact on cannibalization than competition. Stronger consumer brand loyalty mitigates inter-brand competition, but its effect on intra-brand cannibalization varies for different consumer segments. In markets consisting of more high-end consumers, the intra-brand cannibalization increases with consumer brand loyalty, while, in contrast, in markets consisting of more low-end consumers, the intra-brand cannibalization decreases with consumer brand loyalty. The differences in consumer search and brand loyalty explain a significant fraction of the variations in both cannibalization and competition between different PC markets.

2.1 Introduction

Product proliferation is an important component of a firm's competitive strategy and is commonly observed in practice. For example, in the personal computer (PC) industry, we often observe multiple models of desktops or laptops offered by the same firm or brand. In 2001, HP offered 19 desktop models with varying configurations, and this number increased to 35 in 2008. Similar phenomena were observed for other PC manufacturers as well. Product proliferation is not specific to the PC industry. According to Bernard et al. (2006), multiproduct firms account for more than 90% of the output in the U.S. manufacturing sectors.

Product proliferation has at least two competing effects on firm profitability.[1] First, it has an *inter-brand competition effect*: when introducing a greater variety of product offerings, a firm can attract new consumers with heterogeneous tastes and induce consumers to switch from competitors (Bayus and Putsis 1999). Second, it has an *intra-brand cannibalization effect*: products offered by the same firm are often considered by consumers as close substitutes so that "one product's customers are at the expense of other products offered by the same firm" (Mason and Milne 1994, p.163).[2]

Firms need to consider both intra-brand cannibalization and inter-brand competition effects when designing their product lines (Kadiyali et al. 1998; Ruebeck 2005; Wilson 2011). A stronger cannibalization effect gives incentives for firms to shorten their product lines, whereas a fiercer competition between firms may induce them to expand their product lines. Due to the rapid growth of the online market, the product line design in online and offline markets becomes

---

[1] Other effects such as market expansion are not the focus of this study, but will be controlled in our analysis.

[2] In this study, we assume each firm has a single brand, so we use brand and firm interchangeably. Within each brand (e.g. Dell) there can be many products (e.g. Latitude E6400, XPS 8500, etc.).

an increasingly more important issue for multiproduct firms to consider. Earlier work in information systems suggests some important differences between the online and offline markets. For example, the online channel has lower search cost (Bakos 1997; Clemons et al. 2002; Ghose and Yao 2011) and menu cost (Bailey 1998; Brynjolfsson and Smith 2000) compared to brick-and-mortar stores. These differences have important implications to firms' online pricing (Clemons et al. 2002; Ghose and Yao 2011). Recent studies have also started to examine the implications of this new channel to firms' product strategy. For example, Brynjolffson et al. (2011) suggest that the online market creates a long tail in the distribution of sales by substantially increasing the market shares of niche products. A natural question to ask following these findings is: how should firms optimize their product offerings online? Should firms offer more products online than offline given its ability to attract consumers into buying more niche products? Given that the online and offline markets differ in many key characteristics that can potentially affect the degree of intra-brand cannibalization and inter-brand competition, it is important for firms to gauge both cannibalization and competition effects in online and offline markets and to understand their influencing factors to make optimal product line decisions.

Our first objective in this study is to empirically measure intra-brand cannibalization and inter-brand competition and examine whether they vary across online and offline markets. We develop a unified framework to jointly estimate both cannibalization and competition in different markets in the U.S. PC industry. We find significant differences in both effects between online and offline markets. Specifically, the online market exhibits both stronger intra-brand cannibalization and stronger inter-brand competition than the offline market for the PC industry.

This leads to our second research question: *what factors drive intra-brand cannibalization and inter-brand competition in different markets?* In particular, we focus on how consumer brand loyalty and search behavior influence cannibalization and competition, respectively. To answer this question, we model both intra-brand cannibalization and inter-brand competition as functions of consumer preferences. We find that, consistent with intuition, stronger consumer brand loyalty reduces inter-brand competition. The impact of brand loyalty on intra-brand cannibalization, however, varies for different segments of consumers. Specifically, stronger brand loyalty increases intra-brand cannibalization in markets consisting of more high-end consumers (high income, low price sensitivity), but decreases intra-brand cannibalization in markets consisting of more low-end consumers (low income, high price sensitivity). In addition, we find that more active consumer search not only intensifies inter-brand competition but also amplifies intra-brand cannibalization to a larger extent, suggesting that more consumer search before purchase encourages more comparisons within the same brand than between different brands. The differences in consumer search and brand loyalty explain a significant fraction of the variations in both cannibalization and competition between different PC markets.

This study makes the following important contributions. First, our study offers a framework to jointly identify intra-brand cannibalization and inter-brand competition in different markets as well as the impacts of demand side factors on competition and cannibalization in a single integrated model, while at the same time controlling for potential endogeneity issue. Prior studies (e.g., Brynjolfsson et al. 2009; Hui 2004; Watson 2009) have focused on either cannibalization or competition, but have not measured them simultaneously for different markets. Our results suggest that both cannibalization and competition vary across online and offline markets and need to be considered jointly for optimal product line design. Specifically,

while the higher online competition between firms has been recognized in academic research and practice, the higher online cannibalization between products within a firm has been largely overlooked. This could lead to sub-optimal product line design in the online market.

Second, our study highlights important demand side factors (i.e., consumer preferences) that influence both cannibalization and competition, whereas prior studies (e.g., Thomadsen 2007) have mainly focused on the supply side factors (e.g., product characteristics and pricing). Our results suggest that consumer brand loyalty and search behavior have significant impacts on cannibalization and competition, and in addition, that the differences in these consumer preferences play an important role in explaining the variations in cannibalization and competition between different markets.

In particular, when drawing implications about the impact of brand loyalty on product cannibalization, it is important for firms to consider consumers' brand loyalties in different segments because their impacts are different. For example, if a market is mainly composed of loyal consumers from the high-end segment, intra-brand cannibalization can be high and it may be optimal for manufacturers to offer fewer products to the market. In fact, in our data, we observe a higher interaction between loyalty and income in the online channel. Because brand loyalty of high end consumers has a positive impact on cannibalization, the characteristics of the online population in the PC market can make it more likely for us to observe a higher cannibalization online.

While prior IS studies have primarily focused on the impact of lowered search cost on inter-firm competition, we find that the lowered search cost not only affects competition, but also affects cannibalization. More importantly, our result suggests that search in fact can have a higher impact on cannibalization than competition. This result has important implications: while

the increased competition due to lower search cost online may encourage firms to offer more products to attract consumers, it can increase intra-brand cannibalization to a larger extent, limiting the optimal number of products to offer. Therefore, the lowered search cost online does not necessarily favor a longer product line. This can be counter-intuitive since online stores are often not limited by shelf space or capacity constraints, and it is generally cheaper to host a greater product variety online. However, a shorter product line online may be optimal if the higher intra-brand cannibalization online is a major concern.

2.2 Literature Review

Our study is closely related to the literature on product proliferation. Most studies in this research stream develop theoretical models to analyze firms' decisions on product quality (Desai 2001; Katz 1984; Moorthy 1984) or on the length of product line (Bayus and Putsis 1999; Bordley 2003; Liu and Cui 2010; Shugan 1989). When consumers have heterogeneous preferences over quality, firms have incentives to vertically differentiate their products to meet the needs of different consumers. Desai (2001) develops a model for duopolistic competition where consumers have heterogeneous preferences over brand in addition to quality. His results suggest that the brand preferences of consumers with different quality sensitivities affect intra-brand cannibalization differently. Studies on the length of product line suggest that a longer product line can lead to more severe product cannibalization and diminishing marginal contribution to firm performance (Draganska and Jain 2005) or even negative impact on firm profit (Bayus and Putsis 1999). The effectiveness of product proliferation as a marketing strategy thus is clearly limited by the extent of product cannibalization. However, there have been few attempts to study cannibalization, competition and their influencing factors simultaneously, which is a gap in the literature that this study aims to fill.

Prior studies have, however, studied each of the two effects (cannibalization and competition) separately. The research stream on product cannibalization has used different methods to quantify the extent of intra-brand cannibalization. Most papers infer about cannibalization indirectly from cross-product price elasticity (Berry et al. 1995), or from the change of firm performance in response to the change in the length of product line (Bayus and Putsis 1999; Draganska and Jain 2005). Notably Hui (2004) examines a direct measure of intra-brand cannibalization and then links cannibalization to firm brand value. Compared with the cross elasticity between individual products, this measure indicates the degree of cannibalization at an aggregate level. Such a single measure is especially suitable to study our research question, which examines the link between cannibalization and consumer characteristics at the market level. We thus use similar aggregate-level measures in this study. Our study is different from Hui (2004) in two aspects: first, we measure not only intra-brand cannibalization but also inter-brand competition, and second, more importantly, we further investigate the impact of consumer preferences on both cannibalization and competition in a single integrated model.

The recent research stream on inter-brand competition effect focuses on comparing this effect between online and offline markets. Most research in this area focuses on the consequences of lower search costs online and often uses price dispersion as an index of competition (e.g., Baye et al. 2006; Brynjolfsson et al. 2010; Clemons et al. 2002). Some researchers suggest that firms should optimize their product line design by offering more differentiated products in the online market to mitigate online competition (Bar-Isaac et al. 2009; Kuksov 2004). We capture the effect of consumer search on product competition in our model, and different from previous studies, we further compare this effect with the impact of consumer

search on cannibalization and include consumer preference variables to help explain the differences in both cannibalization and competition between different markets.

2.3 Data Description

Our empirical analyses focus on the U.S. PC industry. We choose the PC industry as the empirical context for the following three reasons. First, it is a highly competitive market with a number of major PC makers, each offering a variety of computer models. This combination allows us to study both inter-brand competition and intra-brand cannibalization. Second, products in the PC industry are vertically differentiated with similar functions, so PC vendors compete fiercely on product quality, such as CPU speed. This enables us to apply Lancaster's characteristics theory of value (Lancaster 1966) in our empirical model. Third, this is a dynamic industry characterized by frequent new product introductions and quality improvements, which provide us with more variations in the data that are helpful for parameter identification.

Our data come from two datasets. The first dataset is PC sales data between 2004 and 2008 collected by the International Data Corporation (IDC). For each PC model, IDC records its manufacturer, form factor (desktop or laptop), CPU, average retail price, and unit sales in each quarter. The sales data are available for six distribution channels: Internet, retail, direct inbound, direct outbound, dealer/value-added reseller/systems integrator, and others. In this study, we use data for the Internet and retail channels only,[3] because of two reasons: first, our focus is on comparing the two channels, and second, these two channels mostly reflect consumer purchases and therefore can be matched with the information from our second dataset—the Consumer Technographics Benchmark survey data for the same period from Forrester Research. The

[3] Consumer purchases in other channels will fall under an outside option in our demand model presented later.

benchmark survey is an annual survey of over 50,000 households in North America, from which the households from Canada are excluded from our analysis. The survey includes questions about consumer preferences, general shopping behaviors and purchase decisions for computers, among other products. Detailed information about consumers' PC purchases in the survey is used to match their shopping behaviors to IDC sales data. Specifically, our matching is based on three factors: year of purchase, form factor (desktop or laptop) and channel of purchase (online or offline), given that our consumer preference variables will be measured at the market level. While the IDC sales data allow us to estimate a demand model that quantifies cannibalization and competition in the PC market, the Forrester survey data help explain how consumer preferences affect the extent of cannibalization and competition in the same market.

In our study, we restrict our analyses to firms identified in both the IDC sales data and the Forrester surveys. Our final dataset includes the eight largest PC firms between 2004 and 2008: Acer, Apple, Dell, Gateway, HP, Lenovo, Sony, and Toshiba. These firms accounted for 77% of total sales in the PC industry during the period. In addition, we aggregate the quarterly IDC sales data to a yearly level to match the annual Forrester surveys.

2.4 Empirical Model

2.4.1 Cannibalization and Competition Effects

To address our first research objective that aims to identify both intra-brand cannibalization and inter-brand competition and compare them across online and offline markets, we develop an empirical model under the generalized extreme-value (GEV) discrete choice framework (McFadden 1978). In particular, we consider consumers' PC purchases for different form factors (desktop or laptop) and in different channels (online or offline). We thus set up a

two-level nested logit model as illustrated in Figure 2[4]. In the first level, the four nests correspond to the four markets identified by form factor and channel combinations: laptop-online, desktop-online, laptop-offline, and desktop-offline. The second level sub-nests include all the firms competing in the same market, and each sub-nest has all the computer models from the same firm. With eight firms in our analysis, in total we have 30 sub-nests in the second level because Toshiba desktops are excluded due to very limited sales.

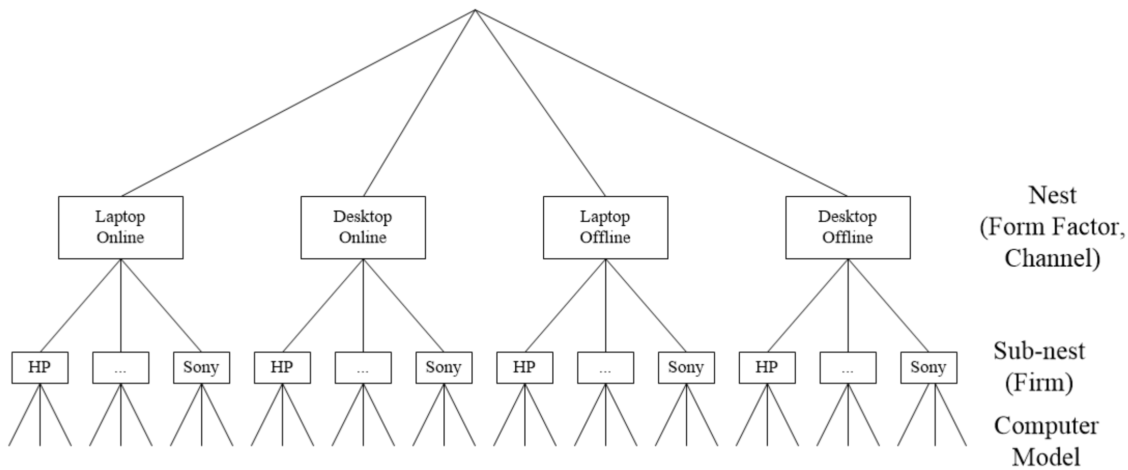


Figure 2: Nest Structure of Our Two-level Nested Logit Model

Let *l* index form factor and *m* index channel. The utility to consumer *h* from purchasing PC model *i* from firm *j* in nest *lm* is specified as:[5]

$$U_{hijlm} = \alpha_p p_{ijlm} + \varphi_1 a_{ijl} + \varphi_2 a_{ijl}^2 + b_j + c_i + d_{lm} + \xi_{ijlm} + \varepsilon_{hijlm} \qquad (1)$$

Here, p_{ijlm} is the price of the PC model. a_{ijl} represents the age of the PC model, measured as the time elapsed since it was first available in either channel. We include the quadratic term of a_{ijl} to capture any nonlinearity in consumer preferences for product vintage. b_j is the brand value of firm *j*. c_i captures consumer preferences for the CPU used in model *i*, which

[4] The rationale behind the choice of our nest structure is explained in detail in Appendix A.
[5] For readability purpose, we suppress all time subscripts.

is a key indicator for PC quality because it is the most important component in determining the performance of a PC (Bresnahan et al. 1997; Hui 2004; Venkatesh and Mahajan 1997).[6] d_{lm} stands for the interactions of the form factor dummies with the channel dummies to capture form factor and channel fixed effects. ξ_{ijlm} captures the unobserved product attributes, while ε_{hijlm} captures the idiosyncratic taste of consumer *h*.

The unconditional market share of PC model *i* from firm *j* in nest *lm* is the product of three terms:

$$s_{ijlm} = s_{i|j|lm} \cdot s_{j|lm} \cdot s_{lm}. \tag{2}$$

Here s_{lm} is the (unconditional) market share of nest *lm*, $s_{j|lm}$ is the market share of firm *j* in nest *lm*, and $s_{i|j|lm}$ is the market share of model *i* within firm *j* in nest *lm*. Following the generalized extreme-value framework, we have:

$$s_{i|j|lm} = \frac{e^{\delta_{ijlm}/1-\sigma_{jlm}}}{G_{jlm}}; \tag{3}$$

$$s_{j|lm} = \frac{{G_{jlm}}^{1-\sigma_{jlm}/1-\rho_{lm}}}{F_{lm}}; \tag{4}$$

$$s_{lm} = \frac{{F_{lm}}^{1-\rho_{lm}}}{\sum_{lm} {F_{lm}}^{1-\rho_{lm}}}; \tag{5}$$

where

$$\delta_{ijm} = \alpha_p p_{ijlm} + \varphi_1 a_{ijl} + \varphi_2 a_{ijl}^2 + b_j + c_i + d_{lm} + \xi_{ijlm}; \tag{6}$$

$$G_{jlm} = \sum_{i \in jlm} e^{\delta_{ijlm}/1-\sigma_{jlm}}; \tag{7}$$

$$F_{lm} = \sum_{j \in lm} {G_{jlm}}^{1-\sigma_{jlm}/1-\rho_{lm}}. \tag{8}$$

As a result, the unconditional market share of model *i* from firm *j* in nest *lm* is given by:

---

[6] Because there are over 100 different CPUs in our dataset, it is empirically challenging to estimate a separate dummy variable for each CPU. Following Gordon (2009) and Sriram et al. (2010), we classify all CPUs into five categories based on their benchmark scores. Our results are robust to the number of CPU categories.

$$s_{ijlm} = \frac{e^{\delta_{ijlm}/1-\sigma_{jlm}}}{G_{jlm}{}^{1-(1-\sigma_{jlm})/(1-\rho_{lm})} \cdot F_{lm}{}^{\rho_{lm}} \cdot \sum_{lm} F_{lm}{}^{1-\rho_{lm}}}. \quad (9)$$

Note that G_{jlm} and F_{lm} are the inclusive values that consumers expect to receive when choosing the best option in sub-nest *jlm* and in nest *lm*, respectively. They indicate the overall attractiveness of the corresponding sub-nest or nest.

Following Berry (1994), we can transform Equation (9) into its equivalent linear form:

$$\log(y_{ijlm}) - \log(y_0) = \alpha_p p_{ijlm} + \varphi_1 a_{ijl} + \varphi_2 a_{ijl}^2 + b_j + c_i + d_{lm} +$$
$$\rho_{lm} \log(S_{j|lm}) + \sigma_{jlm} \log(S_{\mathrm{i}|j|lm}) + \xi_{ijlm}. \quad (10)$$

Here, y_{ijlm} is the unit sales of model *i*, and y_0 is the quantity corresponding to the outside option. To allow for the possibility that the market size may be changing over time, we use a set of yearly dummies to control for the size of outside option in our estimation process.[7]

The parameter σ_{jlm} measures the similarity between the alternatives within sub-nest *jlm* perceived by consumers, and thus captures the intra-brand cannibalization within firm *j* in market *lm*. The rationale is that when consumers choose products to maximize their utility, the extent of competition between products depends on how closely these products are perceived by consumers. The larger the similarity parameter σ_{jlm}, the closer the products are perceived, and the stronger the cannibalization between firm *j*'s own PC models. This can be seen from Equation (10) above. Note that $\log(S_{\mathrm{i}|j|lm})$ is always negative in the $\sigma_{jlm} \log(S_{\mathrm{i}|j|lm})$ term. When σ_{jlm} becomes larger, the sales of each individual product (y_{ijlm}) will become smaller given the same market shares within sub-nest *jlm*, indicating stronger intra-brand

[7] Using our regression results, we also examine the evolution of the outside option in our model. We compare the estimated yearly dummies with the total number of PC models over the years. We find that as the number of models increased, the size of outside option decreased over the years. This pattern may provide a crude indication for the market expansion effect. However, given the focus of this study, we do not pursue to investigate this effect in detail.

cannibalization. Following the same logic, we consider the parameter ρ_{lm} to capture the inter-brand competition between firms in market *lm* since it measures the perceived similarity between the alternatives within nest *lm*. The larger the parameter ρ_{lm}, the more intense the competition between different firms within the market.[8]

To compare cannibalization and competition across online and offline markets, we model both σ_{jlm} and ρ_{lm} as functions of the interacted formfactor and channel dummies. We also include firm fixed effects in the function for σ_{jlm}. To ensure that σ_{jlm} and ρ_{lm} are between 0 and 1, we apply a logistic transformation:

$$\sigma_{jlm} = \frac{e^{\theta_j + \theta_{lm}}}{1 + e^{\theta_j + \theta_{lm}}}; \tag{11}$$

$$\rho_{lm} = \frac{e^{\mu_{lm}}}{1 + e^{\mu_{lm}}}. \tag{12}$$

After estimating σ_{jlm} and ρ_{lm}, we can then compare these two effects between online and offline markets to examine whether the cannibalization and competition effects vary in different markets.

2.4.2 Consumer Preferences

To answer our second research question that aims to examine the impacts of consumer preferences on cannibalization and competition, we further link σ_{jlm} and ρ_{lm} to variables indicating consumer preferences. Literature suggests that there are at least two aspects of

---

[8] To see the connection between our cannibalization parameter and change in sales, we run a simple simulation according to Equation (10) to see how product sales respond to changes in the number of products at different values of σ_{jlm}. By fixing other parameters except the number of products offered by a firm, it shows that, increasing the number of products, e.g., from 5 to 10, doubles the total sales for the firm when $\sigma_{jlm} = 0$, but leads to no gain in total sales for the firm when $\sigma_{jlm} = 1$. Similar insight can be drawn for competition parameter ρ_{lm}. The detail of the simulation is available upon request.

consumer preferences related to intra-brand cannibalization and inter-brand competition: brand loyalty and search behavior.

Multiproduct firms usually differentiate products on quality to serve consumer segments varying in preferences for quality (Moorthy 1984; Moorthy 1988; Mussa and Rosen 2012). Firms are concerned, however, about their low-quality product cannibalizing the demand meant for their high-quality product. The magnitude of this cannibalization is determined by the substitutability between high-quality and low-quality products. This substitutability is contingent on the price and quality of each product, while firms' price and quality decisions are affected by consumers' preferences for different brands. The brand loyalties of price-sensitive (low-end) and quality-conscious (high-end) consumers can have different effects on firms' price and quality decisions, which, as a result, affect intra-brand cannibalization differently (Desai 2001). Based on the results from Desai's analytical model, if the brand preference of price-sensitive consumers is higher, the lack of competition between low-quality products from different firms will likely make them less attractive for the high-end consumers (as a result of firms optimizing both quality and price), thus mitigating intra-brand cannibalization. Conversely, if the brand preference of quality-conscious consumers is higher, the lack of competition between high-quality products will likely make them less attractive and thus increase the high-end consumers' incentives to buy the low-quality product, intensifying the intra-brand cannibalization. Correspondingly, we expect the impact of brand loyalty on intra-brand cannibalization to be positive in markets that consist of more quality-conscious consumers and negative in markets that consist of more price-sensitive consumers.

In addition to affecting intra-brand cannibalization, consumer brand loyalty also affects the competition between brands. Consumers loyal to a brand typically conduct fewer searches for

alternatives, effectively reducing their consideration set and leading to less brand switching (Sambandam and Lord 1995). Moreover, loyal customers become less price sensitive (Krishnamurth and Raj 1991) and less responsive to promotions (Empen et al. 2011; Raju et al. 1990). As a result, stronger brand loyalty can reduce the intensity of inter-brand competition, which applies to both high-end and low-end consumer segments. Due to a lack of theoretical support and empirical evidence, we do not consider the difference between high-end and low-end consumers when modeling the effect of brand loyalty on inter-brand competition.

Consumer search is another factor that may affect both cannibalization and competition. Consumers usually engage in active prior-purchase search when they are uncertain about the attributes or prices of alternative products and must gather further information in order to reach a utility-maximizing choice (Feinberg and Huber 1996; Ratchford 1982). The prior-purchase search is more important for complex decisions, such as purchases of computers (Bettman 1979). After filtering the available alternatives, consumers form their consideration sets that include products that they will consider (Wright and Barbour 1977). Because prior-purchase search may involve comparing products from different firms as well as comparing products from the same firm, it has the potential to affect inter-brand competition as well as intra-brand cannibalization.

To capture the impacts of brand loyalty and search behavior on cannibalization and competition, we further incorporate these consumer preference variables into the functions for σ_{jlm} and ρ_{lm}:

$$\sigma_{jlm} = \frac{e^{\beta_1 L_{lm} + \beta_2 I_{lm} + \beta_3 L_{lm} * I_{lm} + \beta_4 L_{lm} * I_{lm}^2 + \beta_5 S_{lm} + \beta_6 M_{jlm} + \theta_j + \theta_{lm}}}{1 + e^{\beta_1 L_{lm} + \beta_2 I_{lm} + \beta_3 L_{lm} * I_{lm} + \beta_4 L_{lm} * I_{lm}^2 + \beta_5 S_{lm} + \beta_6 M_{jlm} + \theta_j + \theta_{lm}}}; \quad (13)$$

$$\rho_{lm} = \frac{e^{\gamma_1 L_{lm} + \gamma_2 S_{lm} + \mu_{lm}}}{1 + e^{\gamma_1 L_{lm} + \gamma_2 S_{lm} + \mu_{lm}}}. \quad (14)$$

Here, L_{lm} is the average brand loyalty of consumers who shop for form factor *l* in channel *m*. We use the item "*When I find a brand I like, I stick to it*" in the Forrester survey to measure brand loyalty.[9] In the literature, brand loyalty has been operationalized from either behavioral or attitudinal perspectives. Specifically, behavioral loyalty is determined by the observed outcome, such as repeated purchase (Guadagni and Little 1983; Kahn et al. 1986), while attitudinal loyalty focuses on consumers' stated preferences and purchase intentions (Chaudhuri and Holbrook 2001; Jacoby and Kyner 1973). Our measurement falls into the latter category.

I_{lm} represents the average income of consumers who shop for form factor *l* in channel *m*. Consumers' income levels reflect their price sensitivities. High-income (low-income) consumers are expected to be low (high) in price sensitivities. Thus, if a market has a high (low) average income, we expect it to be mainly influenced by high-end (low-end) consumers. Given our earlier discussion that the impact of brand loyalty on intra-brand cannibalization differs across markets consisting of different consumers, we interact the brand loyalty L_{lm} with the average income I_{lm} in both linear and quadratic terms, so that the impact of brand loyalty on cannibalization may vary at different income levels.

S_{lm} captures the average consumer preference for prior-purchase search for the consumers who shop for form factor *l* in channel *m*. A consumer's preference for search is measured by the survey item about consumer tendency to "*research products for purchase*" in the Forrester survey. To control for the impact that the number of models has on product

[9] We also tried using a different survey item "I would pay more for products consistent with an image I like" to measure brand loyalty and obtained similar results.

cannibalization, we add M_{jlm} into our model, which is the ratio of the number of models offered by firm *j* to the total number of models in nest *lm*.

2.5 Estimation

We first replace σ_{jlm} and ρ_{lm} in Equation (10) with Equations (11) and (12) to answer our first research question of whether cannibalization and competition vary across markets, and then replace σ_{jlm} and ρ_{lm} in Equation (10) with Equations (13) and (14) to answer our second research question of how consumer preferences affect cannibalization and competition. This setup jointly estimates cannibalization and competition as well as the impacts of two demand side factors (search and loyalty) on both cannibalization and competition in a single integrated framework. This approach is more efficient than a two-step procedure like the one used by Hui (2004). This setup also allows us to observe the extent of variations in cannibalization and competition that can be explained by the consumer preference variables included in this study. Table 1 presents the summary statistics of the key variables and Table 2 presents their correlations.

Table 1: Descriptive Statistics

	Description	Obs.	Mean	St. Dev.	Min	Max
y_{ijlm}	Unit sales	2,851	65,691	154,819	2	2,160,372
$s_{j\|lm}$	Market share of firm *j* in nest lm	2,851	0.20	0.24	0.00*	0.89
$s_{i\|j\|lm}$	Market share of model *i* in sub-nest jlm	2,851	0.05	0.08	0.00*	0.98
p_{ijlm}	Price (in $1,000)	2,851	0.92	0.48	0.21	5.00
a_{ijl}	Product age (in years)	2,851	0.60	1.01	0.00	11.23
L_{lm}	Consumer loyalty	2,851	3.52	0.06	3.36	3.62
S_{lm}	Consumer search before purchase	2,851	0.16	0.04	0.12	0.29
M_{jlm}	Percent of models offered by firm *j*	2,851	0.21	0.10	0.02	0.56
I_{lm}	Consumer income (in $10,000)	2,851	5.84	1.04	4.48	8.19

(* They are shown to be zero because of truncation.)

Table 2: Correlation Matrix

	y_{ijmt}	$S_{i\|lm}$	$S_{i\|j\|lm}$	p_{ijlm}	a_{ijl}	L_{lm}	S_{lm}	M_{jlm}	I_{lm}
y_{ijlm}	1.0000								
$S_{j\|lm}$	0.3885	1.0000							
$S_{i\|j\|lm}$	0.3426	-0.1226	1.0000						
p_{ijlm}	-0.0591	-0.1155	0.1068	1.0000					
a_{ijl}	0.0725	0.0404	0.1763	-0.0559	1.0000				
L_{lm}	0.0628	0.0911	-0.0473	-0.0699	-0.0402	1.0000			
S_{lm}	-0.0887	0.0108	0.0976	0.0911	0.0745	-0.2169	1.0000		
M_{jlm}	-0.0026	0.3239	-0.3376	-0.2836	-0.0023	0.1711	0.0131	1.0000	
I_{lm}	0.0514	0.0740	0.2056	0.1517	0.1371	-0.2114	0.6399	0.1763	1.000

Note that in Equation (10), the price p_{ijlm} may correlate with the error ξ_{ijlm}, because firms' pricing decisions can be based on some information in ξ_{ijlm} that is unobserved to the econometrician. Moreover, market shares may also be affected by the unobserved product attributes in ξ_{ijlm}. As a result, both price and market shares can be endogenous in our model. We use two sets of instruments to ensure consistent estimates of model parameters. Our first set of instruments uses supply-side cost shifters. Standard differentiated-product models predict that price is a function of marginal cost, so input prices are often used as instrumental variables for prices of end products by prior studies (e.g., Draganska and Jain 2005; Chu et al. 2007). Specifically, we use the Producer Price Index (PPI) for CPUs as an instrument. Given that CPU is a key input to PC production, the PPI of CPU should be highly correlated with PC prices. On the other hand, consumer demand for PCs should not be directly affected by CPU prices after accounting for PC prices. As a result, the PPI of CPU should not be correlated with the error term in the demand system for PC, making it a valid instrument. We download the PPI data for the product category of "Microprocessors" from BLS's website. Our second set of instruments is derived from the observed attributes of related PC models. Berry (1994) shows that the observed

attributes of related products are valid instruments for price and within-nest market share in discrete choice demand models for differentiated products because these attributes are predetermined, or at least determined before consumers' evaluations of the unobserved product attributes are revealed. This type of instruments has been widely used by prior studies (e.g., Berry 1994; Berry et al. 1995; Hui 2004; Zhu and Zhang 2010). Specifically, we use: (1) the sum of product age for the other PC models from the same firm with the same form factor in the same channel; (2) the sum of product age for the PC models from all the other firms with the same form factor in the same channel. To ensure that we have a sufficient number of instruments, we interact these two sets of instruments with the combination of firm, form factor, and channel dummies during the estimation process (Chu et al. 2007; Hui 2004). We have multiple endogenous variables and therefore multiple first-stage regressions. The average first-stage R^2 is 0.6942 and adjusted-R^2 is 0.6813.

A linear model with instrumental variables is typically estimated using a standard two-stage least squares (2SLS) procedure. However, given the logistics transformations applied on σ_{jlm} and ρ_{lm}, the 2SLS estimation procedure is not applicable in our nonlinear setup. Therefore, we use a generalized method of moments (GMM) estimator. The detail of the GMM estimator is in Appendix B.

2.6 Results

2.6.1 Cannibalization and Competition Effects

Table 3 shows our estimation results. Model M1 corresponds to the model specified in Equation (10) with σ_{jlm} and ρ_{lm} replaced with Equations (11) and (12) respectively. The

parameter estimates allow us to calculate the average intra-brand cannibalization and the average inter-brand competition for both online and offline markets.

Table 3: Estimation Results

		M1	M2	M3
	Description	No Consumer Preference Variables (GMM)	Main Model (GMM)	Linear Link Function (2SLS)
Utility function				
$\alpha_p(p_{ijlm})$	Price	-0.8291 ***	-0.3460 ***	-0.3222 ***
$\varphi_1(a_{ijl})$	Product age	0.0122 *	0.0134 ***	0.0106 ***
$\varphi_2(a_{ijl}^2)$	Squared product age	-0.0002 *	-0.0003 ***	-0.0003 **
Competition				
$\gamma_1(L_{lm})$	Consumer loyalty		-6.7689 ***	-0.3172 ***
$\gamma_2(S_{lm})$	Consumer search		6.4832 **	0.3068 **
Cannibalization				
$\beta_1(L_{lm})$	Consumer loyalty		-13.5006 ***	-1.4939 ***
$\beta_2(I_{lm})$	Consumer income		-9.8615 ***	-1.0730 ***
$\beta_3(L_{lm} * I_{lm})$	Consumer loyalty * income		2.2664 ***	0.2559 ***
$\beta_4(L_{lm} * I_{lm}^2)$	Consumer loyalty * squared income		0.0545 ***	0.0048 ***
$\beta_5(S_{lm})$	Consumer search		14.3093 ***	1.5700 ***
$\beta_6(M_{jlm})$	Percent of models offered by firm *j*		1.3033 ***	0.1384 ***
Observations		2,851	2,851	2,851

Note: * p-value<.05; ** p-value<.01; *** p-value<.001. Standard errors are in parentheses. Due to space limitation, we have omitted all the coefficients for dummy variables included in Equations (10) - (14), including firm dummies, interactions of form factor and channel dummies, CPU category dummies, and yearly dummies.

To estimate the difference in cannibalization effect, we first substitute the estimated form factor, channel, and brand dummies into Equation (11) and calculate the cannibalization effects σ_{jlm} for all the form factor-channel-vendor combinations (listed in Table 4). Then we report the difference in average σ_{jlm} between online and offline channels. Given that we use nonlinear functions of the estimators to calculate the difference, the p-value is computed based on a Wald test statistic. To estimate the difference in competition effect, we first substitute the estimated

form factor and channel dummies into Equation (12) and calculate the four ρ_{lm}, one for each form factor and channel combination (listed in Table 5). We then report the difference in average ρ_{lm} between online and offline channels. Again this difference involves nonlinear functions of the estimators. We compute the p-value based on a Wald test statistic. We find that the online market has both higher cannibalization (difference = 0.0722, p-value < 0.001) and higher competition (difference = 0.0460, p-value < 0.001) than the offline market. This answers our first research question by showing that both cannibalization and competition differ across online and offline markets.

Table 4: Cannibalization Parameter for Each Sub-Nest from Model M1

Vendor	Form factor	Online	Offline	Online - Offline
Acer	Desktop	0.9249 (0.0165)	0.7583 (0.0252)	***
Acer	Notebook	0.7574 (0.0047)	0.6292 (0.0337)	*
Apple	Desktop	0.9223 (0.0047)	0.7214 (0.0222)	
Apple	Notebook	0.7459 (0.0594)	0.6629 (0.0718)	
Dell	Desktop	0.9740 (0.0025)	0.8659 (0.0065)	***
Dell	Notebook	0.8946 (0.0200)	0.6924 (0.0259)	
Gateway	Desktop	0.9653 (0.0140)	0.8621 (0.0200)	***
Gateway	Notebook	0.8750 (0.0337)	0.8016 (0.0355)	*
HP	Desktop	0.9496 (0.0086)	0.8030 (0.0137)	***
HP	Notebook	0.8225 (0.0216)	0.7537 (0.0294)	*
Lenovo	Desktop	0.9767 (0.0044)	0.9034 (0.0077)	****
Lenovo	Notebook	0.8997 (0.0200)	0.8523 (0.0205)	*
Sony	Desktop	0.5055 (0.0107)	0.1878 (0.0111)	***
Sony	Notebook	0.3038 (0.0807)	0.2367 (0.0069)	
Toshiba	Notebook	0.9175 (0.0395)	0.8726 (0.0427)	*

Note: * p-value<.05; ** p-value<.01; *** p-value<.001.

Table 5: Competition Parameter for Each Nest from Model M1

	Online	Offline	Online - Offline
Desktop	0.9765 (0.0223)	0.9444 (0.0167)	*
Notebook	0.9587 (0.0131)	0.8882 (0.0222)	***

Note: * p-value<.05; ** p-value<.01; *** p-value<.001.

Model M2 corresponds to the model specified in Equation (10) with σ_{jlm} and ρ_{lm} replaced with Equations (13) and (14) respectively. The "Competition" section lists the variables included in Equation (14) that affect inter-brand competition, while the "Cannibalization" section lists the variables included in Equation (13) that affect intra-brand cannibalization. Using this model, we find a similar result to that of Model M1 in that the online market has both higher cannibalization (difference = 0.1428, p-value < 0.001) and higher competition (difference = 0.0557, p-value <0.001) than the offline market.

2.6.2 Consumer Preferences

To answer our second research question, our result of Model M2 shows that both brand loyalty and consumer search play important roles in influencing cannibalization and competition. In terms of cannibalization, the coefficient of brand loyalty is negative and significant (β_1 = -13.5006, p-value <0.01), whereas the coefficient for the interaction of brand loyalty and income is positive and significant in both linear (β_3= 2.2664, p-value< 0.01) and quadratic terms (β_4 = 0.0545, p-value< 0.01). These estimates suggest that brand loyalty has a negative impact on cannibalization if the average income in the market is low, and the impact is positive if the average income is high. This is consistent with our earlier argument that brand loyalties of high-end and low-end consumers can have opposite effects on intra-brand cannibalization. It is

therefore important to consider consumer composition before drawing conclusions on the effect of brand loyalty on intra-brand cannibalization.

The coefficient for brand loyalty on inter-brand competition, regardless of income, is negative and significant (γ_1 = -6.7689, p-value<0.01), suggesting that stronger consumer loyalty can mitigate the inter-brand competition. It is consistent with the finding in prior literature that when consumers are more loyal, they become less price-sensitive and less likely to switch between brands (Empen et al. 2011; Krishnamurth and Raj 1991; Raju et al. 1990; Sambandam and Lord 1995).

The coefficient for consumer search is positive and significant for both competition (γ_2 = 6.4832, p-value<0.05) and cannibalization (β_5 = 14.3093, p-value < 0.01). We compare the marginal effects of search on competition and cannibalization and further find that search consistently has a higher impact on cannibalization than on competition in both online and offline markets (p-value<0.01). Accordingly, consumer preference for more prior-purchase search intensifies both inter-brand competition and intra-brand cannibalization. Moreover, more active search by consumers before purchase not only facilitates the comparison of products across firms but also amplifies the substitutability of products within a firm to a larger extent. In addition to the consumer preference variables, the coefficient for M_{jlm} is positive and significant (β_6 = 1.3033, p-value<0.01), confirming that a relatively long product line tends to induce higher product cannibalization.

As a robustness test, Model M3 shows the results from a 2SLS estimation with linear link functions instead of the logistic transformations in Equations (13) and (14). Our main results hold qualitatively.

Now that we have shown that consumer preference variables have significant impacts on both cannibalization and competition, an additional question is: how much variation in cannibalization and competition between different markets can be explained by the differences in consumer preference variables? Using estimated parameters from Model M2, we calculate the differences in competition and cannibalization effects between online and offline channels as shown in the first row of Table 6. To see the impact of consumer loyalty and search variables, in the second row of Table 6, we assume that there was no difference in loyalty and search variables between online and offline channels (by replacing their values with the overall average in the market), in which case the differences in competition and cannibalization effects between the two channels come from the channel and form factor dummies only. We can see that our loyalty and search variables explain 53.32% of the difference in competition effect and 48.25% of the difference in cannibalization effect. This finding indicates that our consumer preference variables play a critical role in explaining the cross-market differences in competition and cannibalization.

Table 6: Explaining Power of Loyalty and Search Variables

	Online - Offline			
	Competition		Cannibalization	
Loyalty and search variables set at observed values	0.0557	***	0.1428	***
Loyalty and search variables set at overall average values	0.0260	**	0.0739	***
Percentage explained by variations in loyalty and search variables	53.32%		48.25%	

Note: ** p-value<.01; *** p-value<.001.

We also compare the consumer preference variables between online and offline channels to draw additional implications on their impacts on the differences in cannibalization and competition between the two channels. We find that consumers who buy online tend to search significantly more than consumers who buy offline. We do not find a significant difference in consumer brand loyalty across channels, but the interaction between loyalty and income is higher

in the online channel due to higher income. This suggests that while the overall higher competition in the online channel is primarily driven by the difference in consumer search, the higher cannibalization in the online channel is driven both by the difference in consumer search and by the differential effect of brand loyalty of consumers with different preferences for quality. Even though brand loyalty is not statistically different across channels, the higher average income of the consumers purchasing PCs online suggests a higher consumer preference for quality, ultimately leading to a higher cannibalization in the online channel.

2.6.3 Discussion

Our results indicate that online markets exhibit stronger competition and cannibalization than offline markets. While prior work in information systems has primarily focused on how inter-firm competition differs between online and offline markets, little attention has been paid to the difference in intra-brand cannibalization between online and offline markets. However, both cannibalization and competition are important to firms' online strategies. Our main result that cannibalization is higher online than offline is consistent with the observation from our data that as a firm increases its number of model offerings, its total sales increase at a lower rate in the online market than in the offline market, as shown in Figure 3.

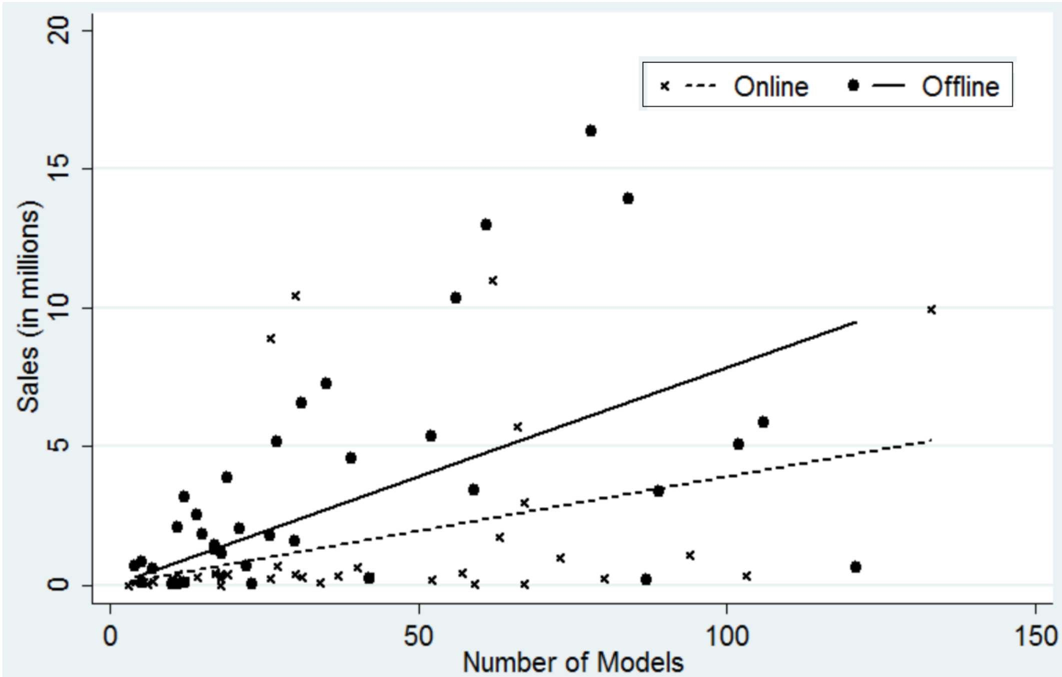


Figure 3: Model Free Evidence of Higher Cannibalization Online

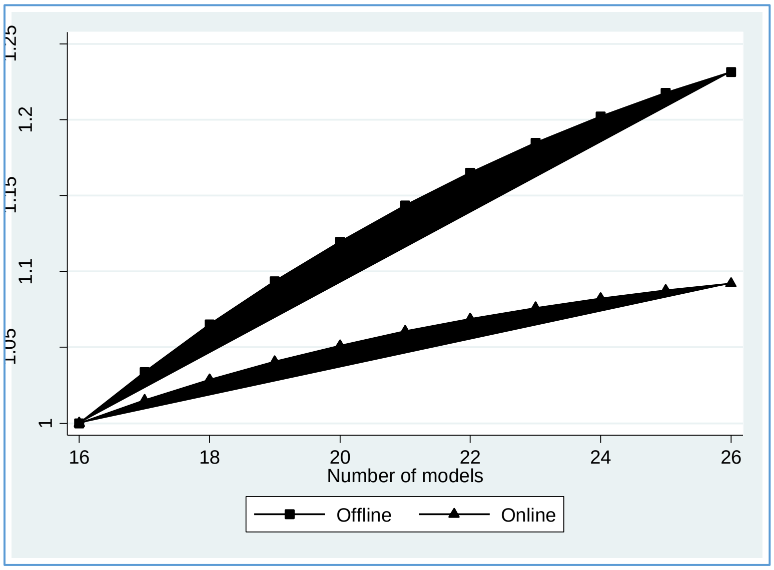
There are at least two reasons why cannibalization can be higher online than offline. First, the online channel has made it a lot easier to search for information. While previous studies have examined the impact of lowered search cost on inter-firm competition, we argue that the lowered search cost, not only affects competition, but also affects cannibalization. More importantly, our result suggests that search in fact can have a higher impact on cannibalization than on competition. The Internet has provided an unprecedented scale of information covering almost every aspect of the product and has also made it very easily accessible. This without doubt intensifies competition between products in the online market, not only across brands but also within the same brand. Because of the amount of details available for the products, it also possibly makes it easier for consumers to distinguish products from different firms and recognize the commonalities between products from the same firm. As a result, while both intra-brand

cannibalization and inter-brand competition are higher online, cannibalization could be affected more than competition.

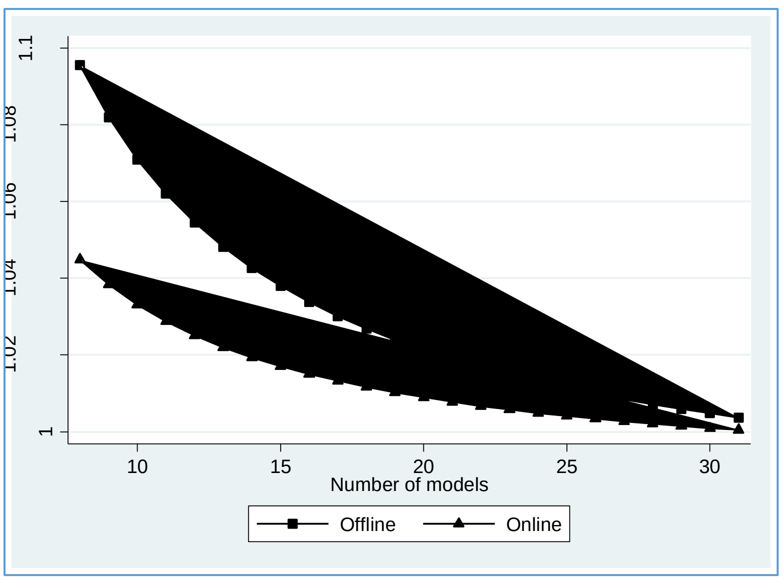
Second, the impact of consumer loyalty on cannibalization can also be different online. Consumer loyalty is not necessarily lower online. The online channel offers features like personalized recommendations and one-click purchasing etc. that can increase loyalty for time conscious consumers. In fact, in our data, we do not find a significant difference in consumer brand loyalty across channels, but the interaction between loyalty and income is higher in the online channel due to the higher income. The people who are more prone to online purchase in the PC market are likely the ones who are more familiar with technology. These people, compared to the general population, can be relatively more knowledgeable and quality conscious. Because brand loyalty of high end consumers has a positive impact on cannibalization, the characteristics of the online population in the PC market make it more likely for us to observe a higher cannibalization online.

Recognizing the importance of both cannibalization and competition for firm's online product strategies, our results also show significant economic impact. For example, we can compare the percentage increase in a firm's total sales after introduction of new products between the online and offline channels through simulations. When we run the simulations, we assume that all the firms are symmetric and all the products are symmetric. We use the mean value for each of our predictor variables to simulate what would happen if one firm increased its number of products. Figure 4a shows the percentage increase in sales as a firm increases its number of products from 16 (which is the average number of computer models per brand in our data) to the number indicated on the X axis. Figure 4b shows the incremental percentage increase in sales if a firm increases its number of products by one assuming its original number of

computer models is the number indicated on the X axis. Both plots show a significant higher increase in sales in the offline channel than in the online channel.



(a) What if the focal firm increases the number of products from 16 to a specific number?



(b) What if the focal firm increases the number of products by 1 from a specific number?

Figure 4：Comparison between Online and Offline Channels: Percentage Increase in Sales as the Number of Products Increases

2.7 Robustness Checks

In this section we explore alternative specifications to check the robustness of our findings. First, intra-brand cannibalization can be affected by the number of computer models. Therefore, instead of using the ratio of the number of models offered by firm *j* to the total number of models in nest lm, in column M4 of Table 7, we directly use the number of models offered by firm *j* in nest lm (N_{jlm}) to control for the impact of product variety on cannibalization and use the total number of models in nest lm (N_{lm}) to control for the impact of product variety on competition. Second, as an additional approach to address the concern that search can be endogenous since consumers are likely to search more when there are more products, in columns M5 and M6 of Table 7, we instrument the search variable in our main model using the total number of models in market lm (N_{lm}) and meanwhile we control for the impact of product variety in Equation (13) using either the ratio variable M_{jlm} or the number of models N_{jlm}. Third, to explicitly control for cross-channel competition within the same form factor, in column M7 of Table 7, we modify our two-level nest structure into a three-level nest structure: Level 1 (Form Factor), Level 2 (Channel), Level 3 (Vendor) and computer models at the bottom. As shown in Table 7, our main results hold qualitatively in all the robustness checks.

Table 7: Robustness Checks

		M4	M5	M6	M7
	Description	Number of Models in Cannibalization and Competition	IV for Search + Ratio of Models in Cannibalization	IV for Search + Number of Models in Cannibalization	3-Layer Model
Utility function					
$\alpha_p(p_{ijlm})$	Price	-0.5282 ***	-0.4268 ***	-0.5386 ***	-0.2687 ***
$\varphi_1(a_{ijl})$	Product age	0.0119 ***	0.0133 ***	0.0085 **	0.0043 *
$\varphi_2(a_{ijl}^2)$	Squared product age	-0.0002	-0.0001	-0.0001	-0.0001
Competition					
$\gamma_1(L_{lm})$	Consumer loyalty	-7.5026 **	-6.8631 ***	-5.2698 *	-6.4624 **

$\gamma_2(S_{lm})$	Consumer search	10.5942 ***	6.0018 **	7.6573 ***	6.4238 *
$\gamma_3(N_{lm})$	Number of models	0.0091 ***			
Cannibalization					
$\beta_1(L_{lm})$	Consumer loyalty	-13.5496 ***	-13.3902 ***	-14.1748 ***	-13.0430 ***
$\beta_2(I_{lm})$	Consumer income	-10.2586 ***	-9.9123 ***	-10.5683 ***	-10.7237 ***
$\beta_3(L_{lm} * I_{lm})$	Consumer loyalty * income	2.6050 ***	2.3405 ***	2.5216 ***	2.4542 ***
$\beta_4(L_{lm} * I_{lm}^2)$	Consumer loyalty * squared income	-0.0373 ***	0.0483 ***	-0.0535 ***	0.0582 ***
$\beta_5(S_{lm})$	Consumer search	13.4020 ***	15.2805 ***	15.1942 ***	14.4817 ***
$\beta_6(M_{jlm})$	Percent of models offered by firm *j*		1.1467 ***		1.0943 ***
$\beta_7(N_{jlm})$	Number of models offered by firm *j*	0.0079 ***		0.0102 ***	
Observations			2,851	2,851	2,851

Note: * p-value<.05; ** p-value<.01; *** p-value<.001.

2.8 Conclusions

In this study, we develop a unified GEV framework to simultaneously measure both intra-brand cannibalization and inter-brand competition in online and offline markets as well as the impacts of search and consumer loyalty on both cannibalization and competition. We find that both intra-brand cannibalization and inter-brand competition are higher in the online market than in the offline market, suggesting stronger competition between products both within a brand and across brands in the online market. While the higher online competition between brands has been recognized in academic research and practice, the higher cannibalization between products of the same brand in the online market has been largely overlooked. Given the importance of both effects in designing optimal product lines, our results highlight the importance for firms to consider both effects when practicing the product proliferation strategy online. If intra-brand cannibalization is more pronounced, it may be optimal for firms to offer fewer products online. This might be the reason why the number of models from HP is consistently smaller in the online

market than in the offline market. This can be counter-intuitive since online stores are often not limited by shelf space or capacity constraint, and it is generally cheaper to host a greater product variety online. However, a shorter product line online may be optimal if the higher intra-brand cannibalization online is a major concern.

Our results also identify demand-side factors that drive cannibalization and competition in different markets. Specifically, we find asymmetric effects of brand loyalty on cannibalization from consumers with different price sensitivities. While higher brand loyalty of high-end consumers can intensify product cannibalization, higher brand loyalty of low-end consumers can have the opposite effect. While higher brand loyalty in general mitigates inter-brand competition, more active consumer search not only intensifies inter-brand competition but also enhances intra-brand cannibalization to a larger extent. We also find that the differences in these consumer preferences can explain a significant fraction of the variations in both cannibalization and competition between different markets. These results highlight the need for firms to consider the impact of consumer loyalty and search behavior on both product cannibalization and competition when designing their product lines. For example, if the market is mainly composed of loyal consumers from the high-end segment, it may be optimal for manufacturers to offer fewer products to the market. Perhaps that is why Apple only offers a few iPhone models, while its major competitor, Samsung, offers consumers far more options. Since consumer search adversely affects both inter-brand competition and intra-brand cannibalization, a lower search cost does not necessarily promote longer product lines, which are often considered effective in reaching and attracting consumers. The long tail phenomenon may not work well for markets mainly consisting of loyal, price insensitive consumers.

This study has a number of limitations that provide opportunities for future research. First, we focus on the PC industry which is a high-technology durable product category. Future research can apply the same framework to examine cannibalization and competition in other industries, such as consumer package goods. While the nest structure and the degree of prior-purchase search or brand loyalty may vary in different industries, the mechanisms by which search and loyalty affects cannibalization and competition, and in turn the direction of our results on how search and loyalty affect cannibalization and competition are expected to be the same. Second, consumer preference is a broad concept and encompasses various dimensions. In this study, we look at only brand loyalty and consumer search, and the survey items we use to construct our variables also may not fully capture these consumer preferences. It will be interesting for future research to examine alternative ways to measure brand loyalty and consumer search and to investigate other dimensions of consumer preference and their effects on market competition and product cannibalization. Lastly, it would also be interesting for future research to incorporate unobserved consumer heterogeneity into the model when examining the impacts of consumer preferences on product cannibalization and firm competition.

Chapter 3 Retail Store Entry and Online Consumer Search: The Role of Awareness versus Convenience Effects

Relatively little research explicitly considers how a firm's channel decisions affect consumers' tradeoffs between the cognitive costs of online search and the physical costs of visiting a retail store. We argue that opening a new retail store affects both cognitive and physical search costs, but that the magnitude of these effects depends on the location of the consumer relative to the new store, and consumers' characteristics. As a result, retail store entry may shift consumers' online behaviors differently. We test these ideas by examining how the opening of multiple retail stores on the same day by a major retailer affects consumer search behavior on the retailer's website. For consumers with prior experience on the retailer's website, we find that opening a new store leads to a 68% increase in the number of visits to the retailer's website by consumers who live in the broader marketing area of the new store, while it leads to a 49% decrease in the number of visits to the retailer's website by consumers who live in the nearby shopping region. More interestingly, the effect of store entry in the shopping region on decreasing website search is weaker when consumers are more efficient in using online channel. In addition to the total number of online sessions, consumers change their within session behaviors after the offline store entry.

3.1 Introduction

Most research on the effect of firm behavior on consumer information search has focused on the physical search costs of gathering information (Balasubramanian, 1998; Forman, Ghose, & Goldfarb, 2009). Relatively little research considers how firm behavior affects consumer tradeoffs between the physical costs of offline search (as determined by the physical effort involved in visiting a retail store; (Balasubramanian, 1998; Moorthy, Ratchford, & Talukdar, 1997) and the costs of online search (as determined by the effort to acquire and process online information; (Hoque & Lohse, 1999; Ratchford, Lee, & Talukdar, 2003). We argue that opening an offline store can increase or decrease the attractiveness of the retailer's online store depending on where the consumer lives. In other words, we propose that offline stores can have complementary as well as substitution effects on online search behavior. For consumers who live in the nearby shopping region, store entry can decrease online search by making offline search more convenient (a substitution effect). However, for consumers store entry can increase online search by creating greater awareness of the retailer's brand (a complementary effect). Further, we propose that the impact of offline store entry depends on consumer-specific characteristics such as their efficiency of using the online channel.

We test these ideas using a unique dataset that combines the online search behavior of a representative sample of US consumers, and census data, with information on the opening of 64 new Kohl's retail stores in October 2006. We compare changes in the online search behavior of consumers living in areas that experience store entry, with that of consumers living in areas with no new store openings. In other words, we treat store openings as a natural experiment to examine how offline store entry affects online search behavior and how this effect is influenced by consumer-specific characteristics.

We find that for consumers with prior experience on the retailer's website, opening a new store leads to a 68% increase in the number of visits to the retailer's website when a new store enters the broader marketing area, while it leads to a 49% decrease in the number of visits to the retailer's website when the new store enters the nearby shopping region. More interestingly, the effect of store entry in the shopping region on decreasing website search is weaker when consumers are more efficient in using the online channel. In addition to the total number of online sessions, consumers change their within session behaviors after the offline store entry.

Although theoretical analyses suggest complementary as well as well as substitution effects of adding new channels (e.g., Lal & Sarvary, 1999), empirical research has generally found that the offline channel competes with the firm's online channel (Anderson, Fong, Simester, & Tucker, 2010; Goldfarb & Tucker, 2011). In addition, prior research has focused on purchase rather than search behavior (Anderson, Fong, Simester, & Tucker, 2010; Avery, Steenburgh, Deighton, & Caravella, 2012; Forman, Ghose, & Goldfarb, 2009; Venkatesan, Kumar, & Ravishanker, 2007; Zettelmeyer, 2000). Understanding the impact of offline store entry on online search is important because many consumers use the Internet to search for product information prior to offline purchase (Verhoef, Neslin, & Vroomen, 2007). Extant research has also tended to focus on differences in the physical costs of search, as determined by the consumer's location relative to the retailer, rather than accounting for differences in the consumer's prior experience with a retailer and their efficiency in using a particular channel. By accounting for factors that affect the costs of online and offline search, we add to prior research by identifying when, and for which consumers, opening a physical store has complimentary versus substitutive effects on online search behavior.

From a managerial standpoint, our results suggest that fears about cannibalization of the online market by offline markets may be overstated. Depending on where consumers live, as well as their efficiency in using the online channel, store entry may increase online search for some consumers while decreasing online search for others. Our results also point to the need to consider the online channel when making decisions about the physical channel. Understanding how specific types of consumers respond to new store openings has implications for store location choices and direct advertising decisions. For example, in addition to considering consumer demographics, shopping, and travel behavior when choosing store locations, a retailer should account for the potential benefits of attracting consumers who live in the broader marketing area to the retailer's website. Similarly, cross-channel effects of entry decisions should be taken into account when assessing the role of each channel in driving sales.

3.2 Theoretical Framework

Despite the dramatic growth of the Internet, consumers continue to rely on traditional retail stores for the majority of their shopping (US Census2010). Prior research based on interviews with consumers suggests that the offline channel is generally preferred to the online channel for search as well as purchase (Frambach, Roest, & Krishnan, 2007). Although the Internet offers a number of advantages for consumer search, including more detailed product information, greater product variety, ease of price comparisons, and the ability to search for products from home (Alba et al., 1997; Anderson, Fong, Simester, & Tucker, 2010; Zettelmeyer, 2000), it has several disadvantages. For example, evaluating products may be difficult online (Kambil & Van Heck, 1998; Overby & Jap, 2009). Further, consumers may vary in their ability to effectively utilize the online channel (Bapna, Goes, Gupta, & Jin, 2004). Traditional retail stores allow consumers to touch and feel products and obtain face-to-face help from a salesperson (Brynjolfsson, Hu, &

Rahman, 2009). In addition, through their physical location and investments in real property, retail stores enhance awareness and trust that the store will be around in the future to service consumer needs (Kirmani & Rao, 2000).

Consequently, we argue that the effect of store entry on online search is likely to depend on how store entry affects consumer awareness of the retailer and the relative convenience of using each channel. More specifically, we distinguish between the effects of new store entry on consumers who live in the store's marketing region and effects on consumers who live in the store's narrower shopping area. For simplicity we refer to these as *awareness* and *convenience* effects, respectively. Marketing regions are larger and defined by high degrees of social and economic integration, while shopping areas are smaller and are defined by the distance the consumer typically travels to visit an offline store.

3.2.1 The Convenience Effect

Prior research suggests that, in the absence of search and travel costs, consumers often prefer to visit physical stores (Frambach, Roest, & Krishnan, 2007). In particular, the online channel is seen as an inferior channel for product information because of the inability to touch and feel products and obtain face-to-face help from a salesperson (Alba et al., 1997; Verhoef, Neslin, & Vroomen, 2007). Further, a physical store engenders trust in the retailer by serving as an observable signal of quality (Kirmani & Rao, 2000). Although consumers may be willing to spend more time traveling for hedonic than utilitarian goods (Okada, 2005), or for greater proportional differences in prices (Kahneman & Tversky, 1984), there is a point at which the costs of traveling to the physical store exceed the perceived costs of obtaining information through the online channel. It is well known that consumers primarily visit stores that are located in the vicinity of their residence (Bell, Ho, & Tang, 1998; Brynjolfsson & Smith, 2000; Huff,

1964). Thus, consumers with a new offline store entry within their narrower shopping area are more likely to shift their search for product information from the online to the retail store, and therefore less likely to visit the retailer's online store.

Hypothesis H1: *For households located within the shopping region of a retailer's new offline store, store entry will reduce search at the retailer's website.*

3.2.2 The Awareness Effect

Store entry into the marketing region where a consumer lives increases consumer awareness of the retailer as the consumer is more likely to see the physical store, be exposed to word-of-mouth about the retailer from other consumers, encounter references to the physical store in Internet and directory searches, and be exposed to advertising for the offline store. For example, retailers often run television advertising and produce newspaper circulars in areas where offline stores are located (Kohl's Corporation, 2010). In other words, the presence of an offline store provides the consumer with greater exposure to vivid information that is more accessible in memory and enhances consumer confidence in the retailer (Berger & Mitchell, 1989; Herr, Kardes, & Kim, 1991). This should increase the likelihood that the consumer considers the retailer when searching for product information online.

New offline store entry in the consumer's marketing region should also reduce information gathering and evaluation costs associated with consideration. In particular, because exposure to the store and its products may be incidental rather than effortful, the presence of a physical store should lower the costs and increase the likelihood of gathering information (Shapiro, Macinnis, & Heckler, 1997). Incidental exposure should also lower information search costs by allowing consumers to use memory-based consideration processes (Nedungadi, 1990); this should also enhance the likelihood of consideration (Kardes, Kalyanaram, Chandrashekaran, & Dornoff,

1993). Store entry should also reduce the cost of evaluating information (Shugan, 1980) by enhancing the consumer's ability to assess the type of products the retailer offers; this should make it more likely for consumers to consider the retailer when searching for product information online (Hauser & Wernerfelt, 1990).

Finally, investing in a new offline store should enhance the consumer's trust in the retailer by serving as an observable signal of quality (Kirmani & Rao, 2000). In addition, since consumers' online behavior may be influenced by other consumers who live in the same neighborhood or zip code (Bell & Song, 2007; Forman, Ghose, & Wiesenfeld, 2008; Jank & Kannan, 2005), new store entry should allow consumers to learn from the experience of others. The trust engendered through physical presence, the ability to talk to "a real person," and the ability to learn from other consumers should make consumers more likely to consider the retailer when shopping for products. For all these reasons—greater awareness, greater information accessibility, lower evaluation costs, and increased trust—store entry in the marketing region where the consumer lives should make consumers more likely to consider the retailer when shopping for products. This greater likelihood of being considered should increase consumer search on the retailer's website.

Hypothesis H2*: For households located within the marketing region of a retailer's new offline store, store entry will increase search on the retailer's website.*

3.2.3 The Moderating Role of Consumer-Specific Characteristics

In addition to being affected by the consumer's location relative to the new store, the effect of store entry on search behavior should also depend on consumer-specific characteristics: the consumer's efficiency in using the online channel and the consumer's experience to the

retailer. The first should affect the relative reduction in perceived offline versus online search costs. The second should affect the awareness effect in considering the retailer.

We have argued that consumers make tradeoffs between the costs associated with searching the online channel and the physical search costs associated with traveling to the physical store. We have also argued that, for consumers living in the narrow shopping region of the new store, store entry reduces the perceived physical costs of search thereby lowering the appeal of visiting the retailer's website. This effect should be greater for those who face higher search costs (i.e., are less efficient) when using the online channel. For such consumers, the reduced physical costs of visiting the new retail store are likely to be lower than the costs of online search, and they are more likely to substitute online search with visiting the retail store. In contrast, consumers who are more efficient at using the online channel are likely to search (and shop) online simply out of habit, obtain hedonic benefits from online search, and browse multiple retailers online when searching for product information. Such consumers are less likely to be affected when a retail store opens in their shopping region. That is, the effect of store entry in reducing online search by those living in the shopping region of the store should be weaker for consumers who are more efficient in using the online channel.

Hypothesis H3: *The effect of new store entry in the narrow shopping region, on decreasing the likelihood of visit to the retailer's website, will be lower for consumers who are more efficient in using the online channel.*

We have also argued that store entry will enhance the awareness of the retailer. That is, incidental exposure to the retailer through the offline store increases the likelihood of inclusion in the consideration set of the consumer. This effect should be stronger for consumers who have not visited the retailer's website before and were less aware prior to store entry. In contrast,

experienced consumers are more likely to already have the retailer in their consideration set, and further incidental exposure to the retailer should have less effect on increasing awareness for these consumers.

Hypothesis H4: *The effect of new store entry in the broader marketing area, on increasing the likelihood of visit to the retailer's website, will be lower for consumers who have used the retailer's online channel.*

3.3 Data

3.3.1 Search Behavior

Table 8 summarizes the data sources used in this research. Our primary data source for consumer browsing is the comScore Media Metrix 2006 dataset of website visitations. ComScore recruits a large random sample of Internet users (88,814 households in the 2006 dataset) and installs a program on each user's computer that tracks their Internet usage over the entire year. The data set contains the name of every domain (website) visited, a time stamp for the visit, the number of pages visited within the domain, and the total time spent on the domain during that visit. Page-level data is not disclosed to protect consumer privacy. The data set also contains demographic information for each household in the panel including zip code, income, education, number of children, and Internet connection speed.

Table 8: Data Sources

Data	Data Source	Details
Consumer Browsing	comScore Media Metrix 2006	Website visitations of 87,773 households over a span of one year (2006). Details available at www.comscore.com
Kohl's store locations and store openings	Manually collected from Kohl's website and news sources	The location of 817 Kohl's stores in 45 states in 2006 including location of 64 stores that opened on October 5 and 3 stores that opened on November 15, 2006
Geo-demographic Information	The U.S. Census Bureau (http://www.census.gov/)	Zip code level information, including, including percentage of high school graduates, percentage of

	bachelor degrees, median income, population, number of establishments, percentage of female, and population density

ComScore data have been used in prior research (Huang, Lurie, & Mitra, 2009; Moe & Fader, 2004; Montgomery, Li, Srinivasan, & Liechty, 2004). Unlike server-side logs (Bucklin & Sismeiro, 2003; Nicholas, Huntington, Jamali, & Dobrowolski, 2007), or aggregate-level data (Forman, Ghose, & Goldfarb, 2009), comScore provides information at the household level over time, as well as household demographics, and is not limited to a single domain or website. Since the data allow us to observe actual online browsing, we avoid the recall problems associated with assessing product search through consumer surveys.

3.3.2 Store Entry and Location

To evaluate the impact of offline store entry on online browsing behavior, we focus on the simultaneous opening of multiple stores by the well-known retailer Kohl's. At the end of 2006, there were 817 Kohl's stores located in 45 states. Sixty-four of these stores opened on the same day, October 5, 2006. Three additional stores opened in mid-November of 2006 and we dropped all households in the marketing and shopping regions of these stores from the data set because we did not have a sufficient period of post store-entry observations for these households. We calculated the distance between the households in the comScore data set and the nearest Kohl's store based on their zip codes.

We focus on the retailer Kohl's for two reasons. First, Kohl's opened a large number of new stores on a single day in 2006. This natural experiment allows us to evaluate the impact of offline store entry on online browsing behavior while controlling for consumer heterogeneity and time-varying effects such as seasonality. Second, most of the products sold by Kohl's (such as shoes, apparel, home furniture and home accessories) can be classified as search goods (Nelson,

1970, 1974). This is important since previous research has found important differences in consumer online browsing behavior between experience and search goods (Huang, Lurie, & Mitra, 2009). In addition, because the Kohl's website and stores carry almost identical products (Kohl's 2010), the effects we observe are likely due to differences between channels of distribution rather than product lines. Further, based on the comScore transaction data, over 90% of the online orders at the Kohl's website in 2006 were for apparel, shoes and similar products. The focus on a single retailer and a homogeneous set of products reduces concerns about retailer and product-level heterogeneity.

3.3.3 Online Channel Efficiency

We capture consumers' efficiency in using the online channel following Johnson et al. (2003), who show that improvements with practice at performing online shopping tasks are linear in the log-log space. More specifically, they empirically demonstrate that log(T) = log(B) - α Log(N), where T is the time required to complete a task, N is the number of trials, and B is the baseline intercept term (Johnson, Bellman, & Lohse, 2003, pp. 63). To capture consumers' efficiency in using the online channel, we define $Page\_time_i$ as the time required by consumer i to cognitively process a page of online information at the beginning of our study period (July 2006), and we calculate it as follows. We first estimate the α_i and β_i coefficients in the equation: $\log(T_{iv}) = \beta_i - \alpha_i * \log(v)$ through random coefficient models, where T_{iv} is the time spent per page by consumer i during the v^{th} visit to any retailer website in the same genre as Kohl's during the first six months of 2006. We then use the estimated α_i and β_i coefficients to calculate the predicted time spent per page of online information by consumer i at the start of our study period. That is, $Page\_time_i = \exp(\beta_i + \alpha_i * \log(N_i))$ where N_i is the cumulative number of prior visits to retailer websites at the beginning of our study period (July 2006) by consumer i. Thus,

$Page\_time_i$ predicts the speed at which consumer i can process a page of online information at the beginning of our study period. In our following empirical analysis, we use -1 * log($Page\_time_i$) as the proxy for consumer efficiency in using the online channel (Johnson, Bellman, & Lohse, 2003), which means the less time used to process the information in one web page, the higher efficiency for a consumer.

3.4 Research Methodology

3.4.1 Natural Experiment Design

We take the offline store entry of Kohl's as a natural experiment to draw inference on how consumers change their online behaviors. One of the advantages with natural experiment design is that we can control for unobserved events that coincide with the opening of new offline stores, but may affect consumer online behaviors. For example, if there were any nationwide economic fluctuations, they may intensify or attenuate the effect of new store entry. These uncontrolled events do not pose any problems to the identification of treatment effects in our natural experiment design because both treatment and control groups will suffer from the same effect. In addition, the use of natural experiment allows us to control the seasonal fluctuations of consumer search behaviors. It is critical for our study because our data conver the shopping season of Thanksgiving and Christmas holidays.

Empirically, we select and categorize households into two treatment groups and one control group based on the entry (on October 5, 2006) of new Kohl's stores within the shopping and marketing regions defined later (details are shown in Table 9). To make the groups as comparable as possible, and cleanly assess the impact of store entry, we limit our analysis to areas where there were no nearby Kohl's stores within the marketing region prior to Fall 2006. The control group consists of households that did not have any Kohl's stores within their

marketing and shopping regions for the entire year. The first treatment group (marketing entry) had no Kohl's stores within the marketing or shopping regions prior to October 5th 2006, and had one or more Kohl's stores located in their marketing region (but outside their shopping region) after October 5th 2006. Likewise, the second treatment group (shopping entry) had no Kohl's stores within their marketing or shopping regions prior to October 5th 2006, but had one or more Kohl's stores located in their shopping (and consequently marketing) region after October 5th 2006.

Table 9: Control and Treatment Groups

Groups	Number of Stores before Oct. 5, 2006		Number of Stores after Oct. 5, 2006		Description
	Marketing Region	Shopping Region	Marketing Region	Shopping Region	
Control	0	0	0	0	No offline stores in the marketing or shopping area for the whole year
Marketing Entry	0	0	≥ 1	0	Offline store entry in the marketing (but not shopping area) in October 2006
Shopping Entry	0	0	≥ 1	≥ 1	Offline store entry in shopping area (and hence also in the marketing) in October 2006

In addition, we separate consumers in our data set into two groups according to whether they visited the Kohl's retail website during 2006. Our study period consists of the last six months (July–December) of 2006 and we use the first six months (January–June) of the year to calculate variables related to the consumer's prior experience at Kohl's and efficiency in online information search.

3.4.2 Marketing and Shopping Regions

We define the *marketing region* of an offline store, which captures the awareness effect of offline store presence, according to its distance to consumers. According to the findings by

(Zentner, Smith, & Kaya, 2013), the impact of offline store on consumers' behaviors on the retailer's website diminishes with the distance of the offline store. As a baseline analysis, we use 50 miles in the definition of marketing region for an offline store, but we have tested our results to be robust to alternative distances. This is also the strategy used by (Gallino & Moreno, 2014), which assumes the marketing region to be a 50 miles circle centered at the offline store.

The *shopping region* of an offline store, which captures the substitution effect of offline store presence on online browsing behavior, is determined based on the average distance that a consumer is likely to travel to visit an offline store. While distance is widely acknowledged to be an important predictor of offline store visits (Bell, Ho, & Tang, 1998; Brynjolfsson & Smith, 2000), there are few estimates of the average distance that a consumer is willing to travel to shop at a store, and such estimates vary by product type and consumer beliefs about travel time (Kang, Herr, & Page, 2003; Murdie, 1965). For example, Murdie (1965) estimates that most customers purchase shoes (a product category that is similar to the products purchased from Kohl's website) from stores that are located within 10 miles of their residence. The U.S. Department of Transportation provides a similar estimate of 7 miles for the average distance a consumer travels to shop at a retail store (US Department of Transportation2008). Brynjolfsson and Smith (2000) estimate that the average distance that a consumer travels to purchase books and CDs is 5.4 miles, also consistent with a maximum distance of 10 miles to purchase at an offline store. Accordingly, we use a distance of 10 miles from the offline store location to define the shopping region. Since Kohl's stores are located in large metropolitan areas, the shopping region is contained within the marketing region for almost all households in the sample.

3.4.3 Model Specification

Our primary interest is to understand how an entry of offline store affects consumer search in the retailer's website, and how consumers' characteristics moderate such a shift in their online behaviors. Hence, we include both the indicators for store entry in marketing and shopping regions, and their interactions with consumer prior visits and online efficiency in the model. Finally we specify the model as:

$$\begin{aligned} Y_{it} = \beta_i + \beta_1 * SHP_{it} + \beta_2 * MKT_{it} + \beta_3 * PriorVisit_i * SHP_{it} + \beta_4 * PriorVisit_i * MKT_{it} \\ + \beta_5 * SHP_{it} * Efficiency_i + \beta_6 * Month_t + \beta_7 * PriorVisit_i * Month_t + \varepsilon_{it} \end{aligned} \quad (1)$$

In Equation (1), Y_{it} represents the number of visits for consumer i in week t on Kohls.com. SHP_{it} is an indicator for the store entry within the shopping region of consumer i, which is set to be zero before October 6th. Similarly, MKT_{it} is an indicator for store entry in the marketing region. $PriorVisit$ is another indicator for consumers who have visited Kohls.com between January and June in 2006. As a result, β_1 and β_2 represent the effects of store entry on consumers without any experiences on Kohls's website, while β_3 and β_4 capture the incremental effects for experienced consumers. β_5 tests how consumers' online efficiency moderates the effect of store entry within the shopping regions. β_6 and β_7 capture the seasonality effects. In addition, we add fixed effects β_i to control for consumer heterogeneity from unobserved characteristics, such as income, education, etc.

3.4.4 Matching

A possible concern with the nature experiment design is that Kohl's location choices for new offline stores are not random. Although we have partially controlled for individual level heterogeneity in Equation (1) through the consumer level constant β_i since we are using panel

data methods. However, there may be systematic differences between consumers in locations where Kohl's opens new stores (the marketing entry and shopping entry groups in Table 2) versus consumers in other locations (the control group in Table 2). For example, retailers are more likely to open new stores in regions with more favorable geo-demographic conditions, such as more consumers with higher purchase power. Unfortunately, our natural experiment cannot rule out the effect of such systematic differences on consumers' online search behaviors.

Thus, following the recent research in marketing(Wangenheim & Bayón, 2007), we apply matching methods in our empirically study. The objective of matching method is to formulate a control group that is as similar as possible to the treatment groups based on covariates that may be correlated with Kohls's store entry decisions and consumer search behaviors. Following Avery et al. (2012), we match the control and treatment groups at the zip code level. Our matching process involves the following two steps.

We first construct a large pool of zip codes as the candidates for the control group. This pool constitutes of all the areas in America (excluding Alaska and Hawaii) without a Kohl's store throughout the year of 2006. Then we collect the geo-demographic information for all the zip codes in the pool, including percentage of high school graduates, percentage of bachelor degrees, median income, population, number of establishments, percentage of female, and population density. We also include the number of clothing stores and department stores as the proxy for competition against Kohl's in these zip codes.

Then we quantitatively match each zip code in the treatment group with zip codes in the pool created in the first step using all the eight variables as geo-demographic information. Because there are various algorithms in applying matching method, and researchers in marketing have just applied the matching method in marketing (Wangenheim & Bayón, 2007), there is no

conclusive answer to which algorithms is the best. Thus we test four widely used algorithms and report the results in the next section.

The *nearest neighbor* matching algorithm selects the best zip code in the control group pool for each individual zip code in the treatment based on a pre-defined distance measure between the two zip codes. Here, the distance measure is based on the probability regression of store entry on the eight selected covariates.

The *optimal* matching algorithm(Hansen, 2004) is an extension to nearest neighbor procedure. While nearest neighbor algorithm tries to find the closest control match for each treated unit at each step, it does not try to minimize a global distance measure. On the contrary, the optimal algorithm finds the matched control group with the minimum average distance across all the matched pairs.

The *full* matching algorithm(Hansen, 2004; Rosenbaum, 2002) is a kind of subclassification algorithm, which first form subclasses from the candidate pool such that in each of them the distribution of (rather than the exact values) of covariates for the treated and control groups are as similar as possible. Different from other matching algorithms, a fully matched sample is composed of matched sets each of which contains one treated zip code unit and one or more control zip codes. As a result, the full matching procedure tries to use all the candidates except those who are outside the range of common support. In addition, full matching is optimal which minimizes the estimated distance measure between each treated subject and each control subject within each subclass.

The *genetic* matching algorithm(Abadie & Imbens, 2006; Diamond & Sekhon, 2013) uses the genetic searching technique developed by Abadie and Imbens (2006). This algorithm tries to find a set of weights for each covariate such that optimal balance is achieved after

matching, where the balance is determined by two different tests: paired t-tests for binary variables and a non-parametric Kolmogorov-Smirnov test for category and continuous variables.

After matching, we include all the households in the matched zip codes into the control groups. With matching method, we can isolate the effect of store entry by ruling out possible alternative explanations for the change of consumer search behaviors because of the self-location of Kohl's in their store locations. Table 10 shows the comparison of geo-demographic variables before and after the matching. We can see the treatment and control groups are much similar in terms of the used covariates after matching.

Table 10: Comparison of Geo-Information Before and After Full Matching

	Mean (Treatments)	Before Matching		After Matching	
		Mean (Control)	Difference	Mean (Control)	Difference
%High School	82.8045	82.2748	0.5297	82.7625	0.0419
%Bachelor	18.7841	17.8685	0.9156	18.9450	-0.1608
Median Income	46411	42184	4227	46525	-114
Population	12135	9454	2681	12092	43
# of Establishments	51.5793	44.3650	7.2142	51.1167	0.4625
# of clothing stores	6.5504	5.6074	0.9431	6.4481	0.1023
# of Department Stores	0.4107	0.3451	0.0655	0.4078	0.0029
Population density	661.4314	457.9833	203.4481	727.0021	-65.5707
%Female	0.5089	0.5049	0.0041	0.5086	0.0003

3.5 Empirical Results

Table 11 and Table 12 show descriptive statistics for variables in the empirical models and their corresponding correlations. Since the data set will be different for different matching methods, we list here the summary statistics from the results using full matching algorithm. Our empirical analysis is based on the last six months of the year (July–December 2006) with weekly time buckets, while browsing data from the first six months (January–June 2006) are incorporated in the calculation of consumers' prior experience on Kohl's website.

Table 11: Descriptive Statistics

		Obs.	Mean	Standard Deviation	Min	Max
Y_{it}	Number of visits to Kohl's website in week i	164,700	0.0068	0.1070	0	[illegible]
$Pages_{it}$	Average pages / visit	164,700	0.0865	2.2408	0	[illegible]
$Duration_{it}$	Average duration / visit (Minutes)	164,700	0.0395	1.056	0	[illegible]
SHP_{it}	Shopping entry	164,700	0.0778	0.2680	0	
MKT_{it}	Marketing entry	164,700	0.2503	0.4332	0	
$PriorVisit_i$	Prior visits to Kohl's website	164,700	0.0421	0.2009	0	
$Efficiency_i$	Efficiency in using online channel	164,700	0.3031	0.3386	[illegible]	

Table 12: Correlation Matrix

		Y_{it}	$Pages_{it}$	$Duration_{it}$	SHP_{it}	MKT_{it}	$PriorVisit_i$	$Efficiency_i$
Y_{it}	Number of visits to Kohl's website	1.00						
$Pages_{it}$	Average pages / visit	0.41 (0.00)	1.00					
$Duration_{it}$	Average duration / visit (Minutes)	0.40 (0.00)	0.83 (0.00)	1.00				
SHP_{it}	Shopping entry	0.02 (0.00)	0.01 (0.00)	0.01 (0.00)	1.00			
MKT_{it}	Marketing entry	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)	0.50 (0.00)	1.00		
$PriorVisit_i$	Prior Visits to Kohl's website	0.10 (0.00)	0.06 (0.00)	0.07 (0.00)	0.00 (0.37)	-0.01 (0.01)	1.00	
$Efficiency_i$	Efficiency in using online channel	0.03 (0.00)	0.02 (0.00)	0.01 (0.00)	0.06 (0.00)	0.01 (0.00)	0.14 (0.00)	1.00

Note: Significance levels in parentheses.

Table 13: Number of Visits the Online Store

	Variable	No Matching	Matching			
			Full	Genetic	Nearest	Optimal
β_1	SHP_{it}	-0.0020 (0.0048)	-0.0019 (0.0045)	-0.0019 (0.0045)	-0.0019 (0.0047)	-0.0019 (0.0053)
β_2	MKT_{it}	0.0024** (0.0010)	0.0021* (0.0011)	0.0039*** (0.0012)	0.0020* (0.0011)	0.0013 (0.0014)
β_3	$PriorVisit_i * SHP$	-0.0312*** (0.0082)	-0.0321*** (0.0078)	-0.0321*** (0.0078)	-0.0321*** (0.0081)	-0.0321*** (0.0091)
β_4	$PriorVisit_i * MKT_{it}$	0.0465*** (0.0050)	0.0606*** (0.0056)	0.0408*** (0.0060)	0.0533*** (0.0053)	0.0397*** (0.0066)
β_5	$SHP_{it} * Efficiency_i$	0.0116 (0.0071)	0.0116* (0.0067)	0.0116* (0.0068)	0.0116* (0.0069)	0.0116* (0.0069)
β_6	$Month_t$	Included				
β_7	$PriorVisit_i * Month_t$	Included				
Number of obs.		344,979	164,700	138,375	244,242	151,416
Number of households		12,777	6,100	5,125	9,046	5,608
# of Zip codes	Shopping Entry	130	128			
	Marketing Entry	571	566			
	Control	2,787	694	517	1,388	694

	Entry_10	918	916	916	916	916
# of Households	Entry_50	2,037	2,029	2,029	2,029	2,029
	Control	9,822	3,155	2,180	6,101	2,663

Table 13 shows the effects of new offline store entry in the marketing and shopping areas on the number of visits to the Kohl's website. We report the results from four matching methods as well as the one using the original data which include all the zip codes without matching. The bottom two sections report the number of zip codes and the number of households for the five data sets in the estimation. Please notice that some of the zip codes in the treatment groups (with shopping entry or marketing entry) are dropped by the matching methods because there are no comparable households in the control group. Since all the results are qualitatively similar, the following discussion is based on results after applying the full matching method.

Shopping region entry. We focus on the effect of shopping entry on consumers with prior experience on Kohl's website, because the number of visits cannot decrease from zero for consumers who have never search online before. Thus we conduct post estimation of SHP_{it} + $PriorVisit_i * SHP$, which is negative and significant using the full matching method ($\beta_1 + \beta_2 = -0.0340, p < .01$), indicating that opening a new offline store in the shopping area decreases the likelihood of visit to the online store for consumers with prior visits during the first half year of 2006. Thus, we find support for Hypothesis H1. Compared to the mean value of number of visits per week to Kohl's website for these customers before the store entry which is 0.0688, the effect of new offline store entry in the shopping region means a 49% decrease in the consumers' online search on Kohls' website.

Marketing region entry. The coefficient for MKT_{it} ($\beta_3 = .0021, p < .10$) is positive and significant using the full matching method, indicating that opening a new offline store in the marketing area increases the likelihood of visit to the online store for both groups of consumers with or without prior visits during the first half year of 2006. Thus, we find support for Hypothesis H2. Compared to the mean value of number of visits per week to Kohl's website for

all the customers before the store entry which is 0.0046, the effect of new offline store entry in the marketing region means a 46% increase in the consumers' online search on Kohls' website.

Consumer efficiency in using online channel. The coefficient for the interaction term $PriorVisit_i * MKT_{it}$ is significant and positive ($\beta_5 = .0016$, $p < .10$) indicating that the decrease in the likelihood of visit is weaker for consumers who are more efficient in using the online channel. Thus, we find support for Hypothesis H3. As shown in Table (4), the standard deviation of efficiency is 0.3386, which means that the increase of one standard deviation in efficiency reduces the effect of shopping entry from -0.0040 to -0.0035 (= -0.0040 + 0.3386*0.0016), a change of 14%.

Consumer prior experience on Kohl's website. The coefficient for the interaction term $PriorVisit_i * MKT_{it}$ is significant and positive ($\beta_4 = .0606$, $p < .01$) indicating that the increase in the likelihood of visit is stronger for consumers who have visited Kohl's website during the first half year of 2006. It is the direct opposite of our prediction in Hypothesis H4. One possible reason is that these consumers have accumulated enough knowledge on how to use Kohl's website to search product information and make purchases online, thus they would use Kohl's website even more after the store entry in the marketing region. Post estimation of MKT_{it} + $PriorVisit_i * MKT_{it}$ ($\beta_1 + \beta_2 = .0627, p < .01$), which means that the effect of new offline store entry in the marketing region leads to 68% increase in the search on Kohls' website for consumers with prior online experience compared to the mean value of number of visits per week before the store entry for these consumers (0.0917).

3.5.1 Analysis on the Within Session Behaviors

As indicated by Bucklin & Sismeiro (2003), the number of pages within a session is another important variable characterizing consumer online search besides the number of visits, which represents the amount of information gained from one online visit. Thus, we conduct another analyses to show a more complete picture of how the offline store entry affects average number of pages per visit in the online channel. We replace the dependent variable in Equation (1) with the average number of pages per visit, then run the similar regressions for the new variables. Table 14 shows the similar results as the number of visits. Specifically, consumers with prior experience on Kohl's website browse less pages per visit after a store entry within the shopping region, but search more pages per visit after a store entry within the marketing region. For consumers without prior experience on Kohl's website, a store entry leads them to visit more pages. However, we do not find the moderate effect of online efficiency on the effect of marketing entry.

Table 14: Number of Pages per Visit

	Variable	No Matching	Matching			
			Full	Genetic	Nearest	Optimal
β_1	SHP_{it}	0.1724* (0.0966)	0.1751* (0.0981)	0.1751* (0.0987)	0.1751* (0.0962)	0.1751* (0.1011)
β_2	MKT_{it}	0.0398* (0.0207)	0.0312 (0.0245)	0.0332 (0.0267)	0.0403* (0.0216)	0.0356 (0.0262)
β_3	$PriorVisit_i * SHP$	-0.2789* (0.1656)	-0.2903* (0.1684)	-0.2903* (0.1695)	-0.2903* (0.1653)	-0.2903* (0.1737)
β_4	$PriorVisit_i * MKT_{it}$	0.6059*** (0.1017)	0.9459*** (0.1211)	0.5276*** (0.1301)	0.7739*** (0.1071)	0.1928*** (0.1251)
β_5	$SHP_{it} * Efficiency_i$	-0.1517 (0.1442)	-0.1519 (0.1464)	-0.1519 (0.1474)	-0.1519 (0.1437)	-0.1519 (0.1510)
β_6	$Month_t$			Included		
β_7	$PriorVisit_i * Month_t$			Included		
Number of obs.		344,979	164,700	138,375	244,242	151,416

Number of households		12,777	6,100	5,125	9,046	5,608
# of Zip codes	Shopping Entry	130	128			
	Marketing Entry	571	566			
	Control	2,787	694	517	1,388	694
# of Households	Entry_10	918	916	916	916	916
	Entry_50	2,037	2,029	2,029	2,029	2,029
	Control	9,822	3,155	2,180	6,101	2,663

3.6 Conclusion and Implications

Despite the growth of the Internet, traditional stores still account for the vast majority of retail sales and retailers continue to open new stores to gain access to new markets. Our results show that opening a retail store affects consumer search on the retailer's website. In contrast to prior research demonstrating that offline store entry reduces online purchases (Forman, Ghose, & Goldfarb, 2009; Liu, Gupta, & Zhang, 2006), our results show that store entry can both increase and decrease online search depending on consumer-specific characteristics that affect the cognitive costs of online search. For consumers in the marketing area of a new store, entry increases the likelihood of visiting the retailer's website (a complementary effect); for consumers in the shopping region of a new store, entry reduces the likelihood of website visits (a substitution effect). Further, we find that the substitution effect of new store entry on online search is weaker for consumers who are more efficient in using the online channel. To correct for the endogenous selection of store location, we apply newly developed matching methods to select comparable control groups in our natural experiment design.

These results are important because they 1) provide empirical support for the idea that the online and offline channels can be complements and not just substitutes (as demonstrated by prior research on offline store entry) and 2) highlight the impact of consumer characteristics on

the tradeoff between the costs of online and offline search. In summary consumers change their online search behavior not only based on offline store entry but also based on the physical locations of the entering retailer and their efficiency in using the online channel.

3.6.1 Managerial Implications

From a managerial standpoint, these results highlight the need to consider how changes in one distribution channel affect consumer behavior in other channels and how this impact depends on consumer characteristics. In other words, although it seems fairly straightforward that opening a new retail store will draw nearby consumers to that store, it is not clear a-priori that this will increase use of the website for some consumers (a complementary effect) and reduce it for others (a substitution effect). Knowing that store entry will likely reduce use of the online channel by consumers in the shopping region while increasing online search by those in the marketing region, and that these effects will depend on consumer-specific characteristics that affect the search costs, can help managers better understand and allocate online and offline resources in a symbiotic and proactive fashion. For example, if many of a firm's online customers live in the shopping region of a new store, a drop in website visits by these customers may not be a sign of trouble. Similarly, if many of a firm's online customers live outside the shopping region, but within the marketing region, managers should plan for increased visitations by those customers and think of ways to capitalize on this new interest.

Our results indicate that fears about cannibalization of the online market by offline markets may be overstated. Further, our results suggest that, in evaluating returns on retail and online infrastructure investments, firms should account for cross channel effects. In particular, the offline channel should be given credit for driving online search and the online channel should be given credit for facilitating pre-purchase search that likely leads to purchase at the retail store.

Our results also have implications for the location of new retail stores. A priori, it is not clear whether retailers should open new stores in areas with more or fewer consumers that visit their online store. Since the substitution effect of store entry is greater for consumers that are less efficient in using the online channel, retailers can choose locations that have a greater concentration of such consumers. Clickstream data can facilitate their identification.

After the opening of a new retail store, managers can better target consumers by taking into account the consumer's location relative to the new store, their loyalty to the retailer, and their efficiency in using the online channel. For example, consumers within the shopping region of the new store who are less efficient in using the online channel are better targets for the new retail store, while those who live in the marketing region (but outside the shopping region) and are infrequent visitors to the online store, are better targets for the online store.

3.6.2 Limitations and Future Research

Our approach has a number of strengths and limitations. Our quasi-experimental approach, in which we look at how multiple store openings on a single day in Fall 2006 by a major retailer affects online search at the retailer's website and compare these changes to those for consumers in markets where there was no store entry, allows us to control for seasonality and other time-based effects. Our use of fixed-effect panel data models allows us to control for consumer heterogeneity. Our use of behavioral data over a six-month period overcomes the drawbacks of using self-reported behavior. At the same time, our focus on a single retailer means the effects we observe may not be generalizable to other retailers. In particular, the effects we observe may be stronger for retailers that integrate their web and retail store strategies to focus on a single target consumer and weaker for retailers that focus on distinct consumer segments in these different channels. Another limitation is that we only observe online behavior and do not

have data on consumer offline behavior. Hopefully future datasets will allow researchers to track consumer behavior over time across multiple channels.

Future research could build on our findings in a number of ways. For example, in addition to consumer efficiency in using the online channel, other individual-level factors such as overall Internet buying experience, income, and the need the need to “touch” certain products prior to purchase (Peck & Childers, 2003) may be important moderators of the effects of store entry on online consumer behavior. Although we focus on store entry, other research could examine the effects of other interventions such as store closings, cross-channel advertising, and new product introductions in the retail store on online search behavior.

Chapter 4 Technology and Market Structure: An Empirical Analysis of Entry and Exit in the Banking Industry

Most U.S. banks started to offer Internet banking in late 1990s. Since then the retail banking industry expected to substitute the costly branch network with the far more cost-efficient Internet channel. However, we find that the expansion of online banking did not reduce the total number of brick-and-mortar branches, and furthermore, large, national banks expanded their branch network at the cost of small, local banks. Using detailed data on branch location and performance, we estimate a dynamic entry/exit model to investigate the relationship between the technology advancement and the market structure evolution. Our findings suggest that the advent of online banking has provided significant competitive advantages to large banks over small banks. Specifically, large banks are in a better position to take advantage of the increasing residential broadband penetration rate by investing more in online banking services, and hence improve efficiency and reduce the costs in operating offline branches. Our model can disentangle how different factors contribute to the market structure evolution. Through counterfactual simulations, we show that the reduction in operating costs for large banks is the most significant factor driving the recent change in the U.S. banking industry, followed by increased entry costs and increased deposits due to greater online presence.

4.1 Introduction

Technological improvement is one of the fundamental determinants of market structure evolution (Schumpeter, 1942a). Besides improving productivity and lowering operating costs (Casolaro & Gobbi, 2007; Thatcher & Oliver, 2001), a new technology may also help firms increase demand by identifying consumers unsatisfied with the current technology (Schmookler,

1966), entering new potential markets (Vernon, 1966), or gaining first-mover advantages over competitors (Powell & Dent-Micallef, 1997). Following Schumpeter's idea that appropriability from technology increases more than proportionally with firm size, economists generally believe that technology improvements are scale-increasing and leads to an increase in market concentration. However, some researchers argue that smaller firms may benefit more from the changing technology because they are more adaptive to turbulent environments, and moreover, a new technology may facilitate industrial specialization, which can also be more beneficial to smaller firms (Baumol, Blinder, & Wolff, 2003).

The dissensus in the theoretical literature has inspired a large amount of empirical research on the relationship between technology progress and market structure (Brynjolfsson, Malone, Gurbaxani, & Kambil, 1994; Feng & Serletis, 2010; Ferrier & Lovell, 1990; Green, Harris, & Mayes, 1991; Maloney, 2001; Mansfield, 1983). Interestingly, existing empirical evidence has been mixed. According to Mansfield (1983), technology changes may lead to a decrease in concentration in some major industries, despite that economists tend to view it as a concentration-increasing force. Additional empirical evidence is needed to examine how technological changes in a particular industry impact large versus small firms, and what factors drive this impact.

In this research, we develop an estimable econometric model of dynamic discrete game in the spirit of Ericson and Pakes (1995b). Using data on exogenous technology changes, our model of dynamic entry/exit can help us address some common issues in the previous literature. First, most empirical studies estimate economies of scale from a new technology by using a time-trend index on cost (Daly & Rao, 1985; Green, Harris, & Mayes, 1991; Hunter & Timme, 1991; Maloney, 2001; Murray & White, 1983) or on production (Feng & Serletis, 2010; Ferrier &

Lovell, 1990; Kumbhakar, 1987) for active firms in the market. They generally focus on the effect of technology-induced efficiencies on the evolution of firm sizes, but in large part pay little attention to the effect of technology changes on firm entry or exit, which motivates us to develop a dynamic entry/exit model in this research. Second, the prior literature often focuses on internal technological innovations and firms' incentives to invest, in which case both technology and market structure affect each other and hence evolve endogenously. Our empirical strategy circumvents this simultaneity issue by using an exogenous source of technology change to identify the effect of technology on market structure.

We apply the model to study the U.S. banking industry and evaluate how its market structure is affected by the increasing penetration of residential broadband Internet connections, an exogenous technology change driven by industries other than banking. Despite increasing broadband penetration and adoption of online banking, the total number of brick-and-mortar branches in the U.S. has remained stable in recent years. This phenomenon is intriguing because it contradicts the conventional wisdom that the Internet channel cannibalizes the offline channel. In particular, as consumers switch their banking activities from offline to online channels, the number of branches is expected to decrease, as has happened in the book retailing industry (Goldmanis, Hortacsu, Syverson, & Emre, 2010) and the newspaper industry (Deleersnyder, Geyskens, Gielens, & Dekimpe, 2002). Meanwhile, large banks are continuously expanding their branch networks while small bank networks are shrinking, leading to a higher concentration in the overall banking industry.

Drawing on the previous literature, we separate the effect of technology changes on market structure evolution into three factors: demand, operational efficiency, and entry cost. First, on the demand side, online banking offers customers an alternative channel with increased

convenience, higher accessibility and reduced waiting time (Bitner, Brown, & Meuter, 2000; Curran, Meuter, & Surprenant, 2003), effectively lowering transaction costs from customers' perspectives (Campbell & Frei, 2010). Also, online banking can better satisfy customer needs by allowing customers to have more control over the service delivery (Dabholkar, 1991). As a result, online banking customers have been found to be more loyal with higher retention rates (Campbell & Frei, 2010) and tend to maintain higher asset and liability balances (Hitt & Frei, 2002), which lead to an increase in market share for banks with better online services (Campbell & Frei, 2010). Second, online banking is widely believed to have the potential to increase banks' operational efficiency when customers switch transactions from the offline to online channel. For example, online banking allows customers to serve themselves without face-to-face interactions with bank employees in the branches. It is estimated that the cost to process an account transfer can be as low as $0.01 with online banking compared to $1.07 in a branch (Hitt, Xue, & Chen, 2007). Finally, offering online banking services may require a larger amount of investment to open a branch (Advisors, 2001), e.g., on better IT infrastructure and additional training to employees.

Our econometric model enables us to identify how the broadband penetration affects demand and cost factors separately. We find that these effects are asymmetric for firms with different sizes. Although large banks pay higher entry costs than smaller banks, they benefit from online banking in the long run through significant reduction in their operating costs and increase in consumer deposit demand. Our counterfactual experiments show that the effect of broadband penetration on operating costs of large banks is the most important factor driving the recent market structure change in the U.S. banking industry, followed by its effects on entry costs and deposit demand.

Our research makes important contributions to the existing literature. First, we add to the literature on technology and market structure by offering a framework that disentangles how technology changes affect different factors that drive market structure evolution. Such a framework is general enough to be applied to various industries, so that researchers can use it to examine why technology is scale-increasing in some industries but scale-decreasing in others. Second, we empirically explain how the market structure of the U.S. banking industry evolved in recent years in response to the diffusion of broadband residential Internet connections. Our results highlight the tradeoff between immediate investment on higher entry costs and long-term benefit in lower operating costs. A pure myopic static model or a model focusing on production efficiency only cannot fully characterize the dynamic tradeoff in this industry.

4.2 Literature Review

Our research is built on several streams of existing literature. First, this study is closely related to the research on technology progress and its impact on the efficiency improvements in financial services, which has a long history in economics and finance literature. This stream of literature has been dominated by the estimation of cost (Altunbas, Goddard, & Molyneux, 1999; Hunter & Timme, 1991; Lang & Welzel, 1996; Mckillop, Glass, & Morikawa, 1996) or production functions (Feng & Serletis, 2010; Ferrier & Lovell, 1990). Theoretically, these two approaches are equivalent under certain regularity conditions, as shown by Shephard's (1953) famous duality theorem. Most studies in this research stream have documented scale-increasing effects of technology in the banking industry of various countries. For example, Hunter and Timme (1991) find that technology progress leads to a 1% annual decrease in the production cost of U.S. banks from 1980 to 1986, and larger banks benefit more from technological changes than smaller banks. Other researchers have found similar scale-increasing effects of technology for

the banking industry in Japan (Mckillop, Glass, & Morikawa, 1996) and European countries (Altunbas, Goddard, & Molyneux, 1999; Maudos, Pastor, & Quesada, 1996). In contrast, Lang and Welzel (1996) find a larger reduction in production cost for smaller banks with data from German cooperative banking industry. Our paper extends this literature by offering a framework that decomposes the scale-increasing effects of online banking into demand and cost factors that drive the market structure evolution. Also, we focus on a specific technology progress, whereas previous studies typically estimate the effect of technology as the derivative of the cost or production function with respect to a time trend variable, which may reflect the progress in information and telecommunication technologies, financial innovations, demographic changes in local markets, etc.

The prior literature generally limits its focus to changes in firm sizes but, in large part, pays little attention to changes in the number of firms in the market. It is not enough to focus only on the distribution of firm sizes only, because firms' entry/exit decisions are at least equally important when studying how the market structure evolves over time. To address this issue, researchers have developed entry/exit models to explicitly take into account the change in the number of banks and branches. For example, Nam and Elliger (2008) estimate a nested logit model using the U.S. banking data between 2003 and 2006. They find that bank size, structure, and market characteristics are important factors related to bank's branch expansion. De Elejalde (2009) estimates a dynamic entry model to compare the operating profit and sunk cost of entry between large multi-market banks and single-market banks. His results show that single-market banks pay a higher entry cost, but enjoy a higher operating profit than multi-market banks. Our research uses a similar methodology, but focuses on the role of technology advancement in affecting different factors related to branch entry/exit, such as operating and entry costs.

Another stream of related literature studies the relationship between IT investment and productivity (Beccalli, 2007; Brynjolfsson, 1993; Brynjolfsson & Hitt, 2000; Brynjolfsson, Malone, Gurbaxani, & Kambil, 1994; Hernando & Nieto, 2007; Hitt & Brynjolfsson, 1996), but the results are far from conclusive. Hitt and Brynjolfsson (1996) find that IT investment has a positive effect on the productivity for 370 large firms. Others find IT investment has little relationship with the improvement of bank performance (Beccalli, 2007; Hernando & Nieto, 2007). But none of them studies the impact of IT investment on market structure with one exception from Brynjolfsson et al. (1994), which finds that IT investment is associated with a decrease in firm size in America since 1975. One difficulty in these empirical studies of IT productivity is that IT investment may be endogenous – more profitable firms may have more resources and incentives to invest on IT infrastructure. To circumvent such issues, our measurement of technology advancement is exogenous to the focal banking industry.

4.3 The U.S. Banking Industry

A bank is a financial intermediary that accepts deposits and makes loans. In the U.S., the banking industry is one of the largest sectors in the economy, which accounted for 7.2% of the country's Gross Domestic Product (GDP) when combined with insurance firms in 2013 (Bea, 2013) . With over 7,000 banks and 66,000 branches in 2013, the U.S. banking industry is said to be the largest one in the world in terms of the number of banks. Further analysis shows that the banking industry is very fragmented with the co-existence of a large number of small to medium size institutions and a few very large banks.

The U.S. banking industry is also the most heavily regulated in the world (Besley & Brigham, 2011). The key regulators include the Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC), the Federal Reserve systems (FRS), and

state bank regulators (Saunders, 2000). Historically, the state regulators played a critical role in interstate branching restrictions which prohibited a bank to operate across states, because a state received no charter fees from banks incorporated in other states. In 1994, the congress passed the Riegle-Neal Interstate Banking and Branching Efficiency Act (RN Act), removing barriers to opening bank branches across state lines.

Technology has long been regarded as one of the most important drivers to the market structure change in the banking industry (Ecb, 1999). In terms of information technology, the banking industry represents the highest proportion of investment at the industrial level in this country (Advisors, 2001). Traditionally, banks offer their services to customers through brick-and-mortar branches. More recently, with the rapid diffusion of broadband Internet connections, banks offer online banking as an alternative channel to branches for most of their services, such as funds transfer, bill payment and account management. Since its first inception in New York (Cronin, 1997), online banking has grown into a widely used tool with about 47% of American adults paying bills online in 2009 (Whitney, 2009).

4.4 Data

We construct a yearly panel data set for the period of 2008 to 2013[10], including information on the market structure change, bank branch performance, and broadband penetration rate in each market.

4.4.1 Sample Markets

Following Seim (2006), we define banking markets based on the concept of place from the Census Bureau, which is basically "a concentration of population"[11]. In addition, a market should satisfy the following two criteria: (1) It should be large enough and isolated from neighboring markets, so that consumers generally do not use financial services from banks outside the local market; (2) It should be small enough so that there are no submarkets within the local market. To satisfy the first criterion, we first merge places with common boundaries using Topologically Integrated Geographic Encoding and Referencing (TIGER) geodatabases for all places in America[12], because not all Census places are isolated from each other. Then, we choose candidate markets from places after boundary merging whose: (a) largest neighboring place within 10 miles has a population in 2010 no more than 15,000; (b) largest neighboring place within 20 miles has a population in 2010 no more than 30,000; (c) population in 2010 is greater than 1,400. The distances between two places are calculated using their population weighted centroids. To satisfy the second criterion, we select places with a population no more than 50,000 in 2010.

---

[10] We focus on the period of 2008 to 2013 because FCC began to publish tract level residential broadband penetration rate data from 2008.
[11] https://www.census.gov/geo/reference/gtc/gtc_place.html
[12] http://www.census.gov/geo/maps-data/data/tiger-geodatabases.html

In addition, we drop markets with multiple branches from the same bank. The two isolated states, Alaska and Hawaii, are dropped from our sample. Our final sample consists of 1,104 local markets. **Error! Reference source not found.** Figure 5 is a map of the U.S. with identified local markets in red. The map shows that our sample covers almost all the states in the country, with more representation of the South and Midwest regions.

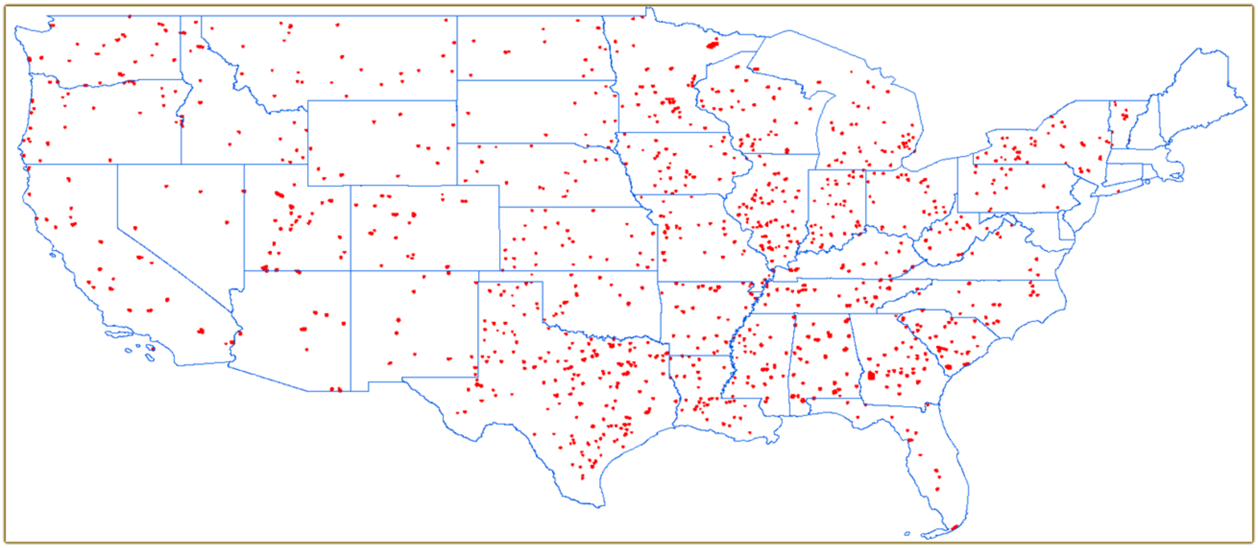


Figure 5. Spatial distribution of the sample markets

We collect market-level demographic data from the U.S. Census Bureau. The income information is drawn from American Community Survey (ACS)[13], while the population information is obtained from Census City and Town Intercensal Estimates[14]. Table 15 shows the descriptive statistics of population and income in 2010. The table shows that the local markets in our sample are generally small, but with substantial variation in income and population.

Table 15. Summary statistics of demographic variables

Variable	Obs.	Mean	Std. Dev.	Min	Max	Median
Income	1,104	18,990	5,527	3,367	101,244	18,586
Population	1,104	3,723	2,993	1,448	35,180	2,730

[13] http://www.census.gov/acs/www/
[14] http://www.census.gov/popest/

4.4.2 *Banks and Branches*

Our bank and branch level data contain information on all the commercial banks and thrifts in the U.S. during the period of 2008 - 2013. We obtain the data set from the Summary of Deposits (SOD) from the Federal Deposit Insurance Corporation (FDIC), which is a U.S. government corporation responsible for insuring deposits, examining and supervising financial institutions, and managing receiverships. SOD is an annual survey of branch office deposits for all institutions insured by FDIC. It contains detailed branch level information as of June 30th in the reported year, including ownership, deposits, and location. We use this data set to construct branch entry/exit decisions in local markets and the performance of active branches. To map the addresses of branches to our identified local markets, we use Google Geocoding API[15] and Yahoo BOSS PlaceFinder API[16] to translate the reported text-based addresses to their corresponding latitudes and longitudes, which can be processed directly by ArcGIS Geographic Information System.

Table 16 and Figure 6 show that the total number of branches is relatively stable from 2008 to 2013. However, the pattern is asymmetric for large and small banks. The total number of branches for large banks, which we define as those with total deposits more than 1 billion U.S. dollars, increased by 4.35% from 2008 to 2013, but meanwhile this number decreased by 15.62% for small banks. We see a similar pattern in our sample markets as indicated by Table 17.

Table 16. Changes in the number of branches for all the markets

Year	2008	2009	2010	2011	2012	2013
All banks						

[15] https://developers.google.com/maps/documentation/geocoding/
[16] https://developer.yahoo.com/boss/placefinder/

Number of branches	98,239	98,653	97,671	97,397	96,546	95,563
Change from last year		414	-982	-274	-851	-983
Percentage change from last year		0.42%	-1.00%	-0.28%	-0.87%	-1.02%
Large banks						
Number of branches	63,429	64,789	64,956	65,964	66,142	66,191
Change from last year		1,360	167	1008	178	49
Percentage change from last year		2.14%	0.26%	1.55%	0.27%	0.07%
Small banks						
Number of branches	34,810	33,864	32,715	31,433	30,404	29,372
Change from last year		-946	-1149	-1282	-1029	-1032
Percentage change from last year		-2.72%	-3.39%	-3.92%	-3.27%	-3.39%

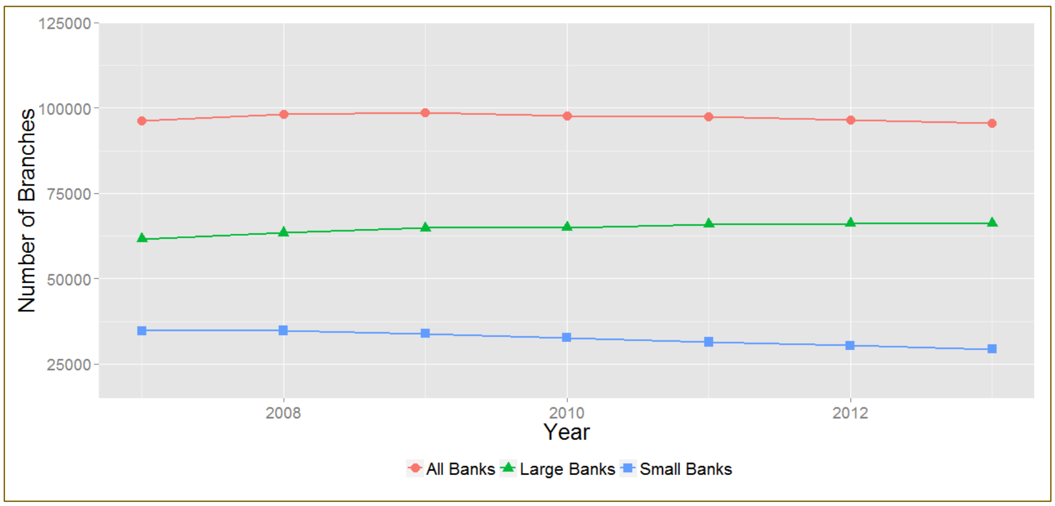


Figure 6. Changes in the number of branches

Table 17. Changes in the number of branches for the sample markets

Year	2008	2009	2010	2011	2012	2013
All banks						
Number of branches	2,784	2,787	2,795	2,802	2,794	2,784
Number of Entrants	62	81	55	50	54	
Number of Exits	69	73	48	58	64	
Net Percentage change from last year		0.11%	0.29%	0.25%	-0.29%	-0.36%
Large banks						
Number of branches	1,159	1,192	1,206	1,227	1,241	1,253
Number of Entrants	41	41	30	32	36	
Number of Exits	8	27	9	18	24	
Net Percentage change from last year		2.85%	1.17%	1.74%	1.14%	0.97%
Small banks						

Number of branches	1,625	1,595	1,589	1,575	1,553	1,531
Number of Entrants	21	40	25	18	18	
Number of Exits	51	46	39	40	40	
Net Percentage change from last year		-1.85%	-0.38%	-0.88%	-1.40%	-1.42%

4.4.3 Broadband Penetration

Our data on broad band penetration is drawn from the Federal Communications Commission (FCC) Form 477, "Local Telephone Competition and Broadband Reporting"[17]. All providers of broadband connections to end user locations are required to file the FCC Form 477 semi-annually. This form contains information on the number of broadband connections per 1,000 households. According to the FCC, an Internet access service is defined as broadband if it is faster than 200 kbps in at least one direction. We map the Census tract level penetration data to Census place level using weighting matrix from Missouri Census Data Center[18].

Table 18 presents the average broadband penetration rate of our sample markets. To summarize the findings in **Error! Reference source not found.** and **Error! Reference source not found.**, we observe the following patterns from 2008 to 2013: (1) Broadband penetration increased by 51.65%; (2) The number of branches increased by 6.53% for large banks but decreased by 5.99% for small banks.

Table 18. Broad band penetration

Year	Obs.	Mean	Std. Dev.	Min	Max
2008	1104	0.3419	0.1800	0.00	0.90
2009	1104	0.3927	0.1898	0.00	0.90
2010	1104	0.4294	0.1963	0.10	0.90
2011	1104	0.4800	0.1684	0.00	0.90
2012	1104	0.5160	0.1647	0.10	0.90

[17] http://transition.fcc.gov/wcb/iatd/comp.html
[18] http://mcdc2.missouri.edu/websas/geocorr12.html

4.4.4 *Descriptive Analysis*

Table 19 presents the Herfindahl-Hirschman index (HHI) for bank level deposits. We can see that HHI was very stable at the Census place level but increased by 75% at the national level from 2008 to 2013. This pattern implies that large banks were entering more local markets during the period.

Table 19. HHI of bank level deposits

Year	National	All Census places	Sample Markets
2008	0.0230	0.6750	0.6235
2009	0.0286	0.6733	0.6208
2010	0.0335	0.6746	0.6193
2011	0.0367	0.6737	0.6182
2012	0.0373	0.6734	0.6207
2013	0.0387	0.6743	0.6255

The primary objective of this study is to investigate how the exogenous change in broadband penetration rate affects the market structure in the U.S. banking industry. As a first step, we conduct reduced form analyses to find out how the increase in broadband penetration rate is correlated with the changes in branch numbers. Specifically we run regression analyses using three different dependent variables: the percentage of branches belonging to large banks; the number of large bank branches; and the number of small bank branches. We include market fixed effects to control for unobserved characteristics in local markets. Table 20 shows the results.

Table 20. Regression analyses on the number of branches and broadband penetration

	% of Large Branches	# of Large Branches	#of Small Branches

Penetration	0.0496*** (0.0122)	0.1264*** (0.0260)	-0.1152*** (0.0307)
Population (10,000)	0.1036 (0.1324)	0.4081 (0.2814)	-0.9840*** (0.3326)
Income ($10,000)	0.0402*** (0.0120)	0.0669*** (0.0254)	-0.0068 (0.0301)
Market fixed effects	included	included	included
Obs.	5,482	5,520	5,520
R-squared	0.9452	0.9731	0.9526

* p<.10; ** p<.05; *** p<.01.

The broadband penetration rate is positively correlated with the percentage of large bank branches (coef. = 0.0496, $p < 0.01$) and the number of large bank branches (coef. = 0.1264, $p < 0.01$), but negatively correlated with the number of small bank branches (coef. = -0.1153, $p < 0.01$). This asymmetry seems to suggest that broadband penetration plays an important role in driving the recent market structure changes in the U.S. banking industry. However, this simple linear analysis reveals no information about how broadband penetration might drive market structure changes. In the next section, we set up a structural model to further our understanding of this issue.

4.5 Model

Our model allows the broadband penetration to affect the market structure evolution through its asymmetric impact on large and small banks in terms of deposit demand, operating costs and entry costs. We first specify how the single-period deposit demand is affected by broadband penetration, and then recover the cost structure through a dynamic entry/exit model.

In principle, we could develop a structural model of deposit demand if we had direct observations on prices (interest rates) and quantities (deposits). However, given that only deposit information is available, we follow the prior literature (Berry, 1992; Singh & Zhu, 2008) to use a

reduced form specification of deposit demand, which controls for competition effects and various market characteristics in a static setup.

Because we do not observe bank branches' operating costs and entry costs, we estimate a dynamic entry/exit model to recover the cost structure of different branches. The identification strategy is to find the set of costs that best rationalizes the branch level entry/exit decisions. Estimating the structural model serves another purpose in allowing for the counterfactual experiments used to investigate how the increasing broadband penetration rate leads to the recent market structure changes.

4.5.1 Deposit Demand

To test if branches of large banks earn a deposit premium from higher broadband penetration rate after controlling for market characteristics including population and income, we regress the deposit amount on these market specific factors as well as the number of competitors from large and small banks. Specifically, the deposit to branch *i* in market *m* in time period *t* is specified as:

$$
\begin{aligned}
Log\_DEP_{imt}^{L} = {} & \alpha_m^{MF} + \alpha_1^L * POP_{mt} + \alpha_2^L * INC_{mt} + \alpha_3^L * PEN_{mt} \\
& + \alpha_4^L * I(n_{mt}^L >= 1) + \alpha_5^L * I(n_{mt}^L >= 2) + \alpha_6^L * Max(0, n_{mt}^L - 2) \\
& + \alpha_7^L * I(n_{mt}^S >= 1) + \alpha_8^L * I(n_{mt}^S >= 2) + \alpha_9^L * Max(0, n_{mt}^S - 2) \\
& + e_{imt}^L \\
Log\_DEP^{S} = {} & \alpha_m^{MF} + \alpha_1^S * POP_{mt} + \alpha_2^S * INC_{mt} + \alpha_3^S * PEN_{mt} \\
& + \alpha_4^S * I(n_{mt}^L >= 1) + \alpha_5^S * I(n_{mt}^L >= 2) + \alpha_6^S * Max(0, n_{mt}^L - 2) \\
& + \alpha_7^S * I(n_{mt}^S >= 1) + \alpha_8^S * I(n_{mt}^S >= 2) + \alpha_9^S * Max(0, n_{mt}^S - 2) \\
& + e_{imt}^S
\end{aligned}
\tag{1}
$$

Here Log_DEP^L and Log_DEP^S are respectively the amount of deposits received by branches of large and small banks after log transformation. We choose a log-linear regression model according to the empirical distribution of branch deposit shown in Figure 7. The left panel

is the distribution of deposit before log transformation, and the right panel is close to a normal distribution after log transformation. The superscripts *L* and *S* indicate whether it is for large or small banks. The same convention will be followed hereafter.

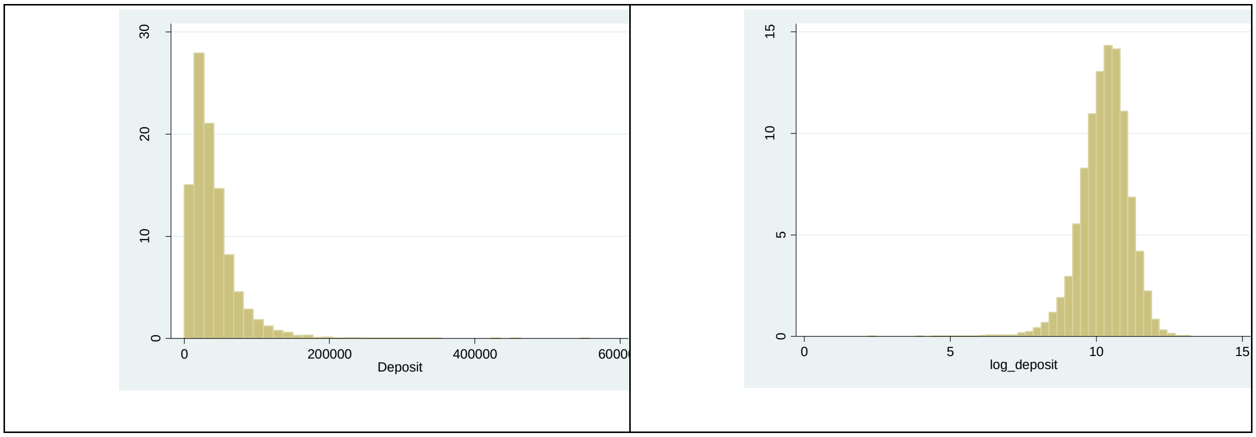


Figure 7. Comparing the distributions of deposit and log(deposit)

We include market fixed effects α_m^{MF} to control for persistent but unobserved market demand conditions. The observed market characteristics include population POP_{mt}, income INC_{mt}, and broadband penetration rate PEN_{mt}. The two key parameters are α_3^L and α_3^S, which capture the effect of broadband penetration rate on the deposit demand. We expect them to be significantly different from each other if there exists asymmetric effect of online banking on the deposit demand for small and large banks.

Competition effects depend on the number of competitors in each type, n_{mt}^S and n_{mt}^L. α_4 captures the competition effect from the first large bank branch. α_5 captures the incremental competition effect from the second large bank branch. When the number of competing large bank branches is greater than two, we assume their competition effects to increase linearly with a coefficient of α_6. Similarly, α_7, α_8 and α_9 are the competition effects from competing branches of small banks. Lastly, e_{imt}^L and e_{imt}^S are i.i.d. normal error terms.

4.5.2 Cost Structure

Recovering the unobserved cost structure requires a structural model. Our empirical framework starts with a dynamic discrete game played by a number of branches in local markets. Incumbent branches decide whether to remain active in the market or exit. Potential entrants decide whether to enter or stay out of the market. The discrete entry/exit decisions are based on the expected discounted flow of payoffs. We learn this expected discounted value and how it varies over different states from the observed decisions of different branch types across different markets. Specifically, operating costs are identified from the difference in entry/exit probabilities for branches under different market conditions, and entry costs are identified from the difference between incumbent branches and potential entrants in their probabilities of choosing to be active.

By examining the variation of operating and entry costs under different broadband penetration rates, we can find out whether large banks are receiving cost advantages from the increasing broadband penetration rate in recent years.

We model bank branches' entry and exit decisions as a dynamic discrete game in local markets, following the spirit of Ericson and Pakes (1995a) and Pakes, Ostrovsky and Berry (2007). Each market has a set of branches, with branch *i* characterized by two indicators: whether it is an incumbent branch (x_{it} = 1) or a potential entrant (x_{it} = 0), and its type (large or small bank). Because we do not observe multiple entrants in our data, we assume there is one potential entrant of each bank type in each market. Each branch chooses to be active or inactive for the next period, and then active branches compete in the local markets.

Each market is described by a vector of state variables, which determines the branches' payoffs. These state variables are common to all branches in the market. We denote this common state vector as s_{mt} . It includes population, income, broadband penetration rate, as well as the endogenous market structure variables, namely the number of branches of large and small banks. In addition, we include the estimated market fixed effects from deposit demand estimation into the state vector because it directly affects the payoffs to active branches.

Besides the market level state variables, branches also observe some private information that affects their profits in the coming period. They consider this private information together with the common state vector in deciding whether to be active or inactive. The private information may include a branch's managerial ability, and/or fluctuations in market conditions. We denote this one-period shock by $\varepsilon_{it}(a_{it})$.

Denote all the payoff related observed states for branch *i* as a vector w_{it} , which contains a 6×1 vector of common market conditions (s_{mt}) and a 2×1 vector of branch status (active or not;

large or small banks). Let $a_{it} \in \{0,1\}$ indicate the branch's decisions on being active or not. The one-period payoff is assumed to be:

$$\pi_{it}(a_{it} \mid w_{it}, \varepsilon_{it}, a_{\text{-}it}) = OR(a_{it} \mid w_{it}, a_{\text{-}it}) - a_{it} * OC(w_{it}) - a_{it} * (1 - x_{it}) * EC(w_{it}) + \varepsilon_{it}(a_{it}) \qquad (2)$$

The private information $\varepsilon(a_{it})$ is assumed to be additively separable. Notably while the operating revenue depends on competitors' decisions, operating and entry costs do not. Because the sell-off value of an existing branch cannot be separately identified from operating and entry costs (Aguirregabiria & Suzuki, 2014), we normalize the sell-off value to be zero following the common practice in literature (Aguirregabiria & Mira, 2007; Collard‐Wexler, 2013; Dunne, Klimek, Roberts, & Xu, 2013; Ellickson, Misra, & Nair, 2012). With such normalization, the meanings of estimated values for operating and entry costs may change slightly. For example, the estimated entry costs are now in fact "ex-ante" sunk entry costs, which equals to the real entry costs minus the sell-off value for branches at the same state (Aguirregabiria & Suzuki, 2014).

We model branches' decisions in a game with discrete time and infinite horizon. The timing of the game is described as following. At the beginning of each period, both incumbent branches and potential entrants observe the market state and draw their private profit shocks. Based on such information, incumbent branches decide whether to remain active in the market or exit and potential entrants decide whether to enter the market or stay out. Once these decisions are made, active branches compete in the local markets, and receive operating revenue OR while incurring operating costs OC. New entrants also pay one-time entry costs EC. At the end of each period, all the state variables evolve.

Operating revenue. We use the estimated deposit demand to derive the operating revenue of an active branch. Specifically, we assume that

$$\begin{aligned} OR^L(a_{it}=1\,|\,w_{it},a_{\text{-}it}) &= \alpha_0^L * \mathrm{E}_{s'_{mt}}\left[DEP(s'_{mt}\,|\,s_{mt},\mathrm{a}_{it}=1)\right] \\ OR^S(a_{it}=1\,|\,w_{it},a_{\text{-}it}) &= \alpha_0^S * \mathrm{E}_{s'_{mt}}\left[DEP(s'_{mt}\,|\,s_{mt},\mathrm{a}_{it}=1)\right] \end{aligned} \qquad (3)$$

We take the expectation of deposits with respect to s'_{mt} because branch *i* does not know how many branches will be operating at the time of decision making. α_0^L and α_0^S measure the profitability of large and small bank branches from a certain amount of deposits. These two coefficients may be different because, e.g., smaller banks tend to focus on retail banking services such as consumer loans and residential mortgages, while larger banks may engage more in wholesale banking services such as industrial and commercial lending (Saunders, 2000).

Operating and entry costs. We specify operating costs *OC* and entry costs *EC* to be

$$\begin{aligned} OC^L(w_{it}) &= \beta_0^{SF} + \beta_1^L * PEN_{mt} \\ OC^S(w_{it}) &= \beta_0^{SF} + \beta_1^S * PEN_{mt} \end{aligned} \qquad (4)$$

$$\begin{aligned} EC^L(w_{it}) &= \gamma_0^{SF} + \gamma_1^L * PEN_{mt} \\ EC^S(w_{it}) &= \gamma_0^{SF} + \gamma_1^S * PEN_{mt} \end{aligned} \qquad (5)$$

β_0^{SF} and γ_0^{SF} are state fixed effects that capture systematic differences in cost related economic factors across states in the U.S., such as regulations for banking industry, tax rates, land price, and labor costs. β_1^L and β_1^S measure the effect of broadband penetration (PEN_{mt}) on a branch's operating cost. β_1^L and β_1^S could potentially be different from each other, implying the effect of broadband penetration on operating costs is asymmetric for large and small banks. Similarly, γ_1^L and γ_1^S capture any possible asymmetric effect of broadband penetration on entry costs.

Equilibrium. We focus on the equilibrium concept of Markov Perfect Equilibrium (MPE). Denote $\sigma = \{\sigma_i(\mathrm{w}_i, \varepsilon_i)\}$ as the set of strategies with each player's strategy as a mapping $\sigma_i : (\mathrm{w}_i, \varepsilon_i) \rightarrow a_i \in \{0,1\}$. Given these strategies, by the principle of optimality, each branch's decision process satisfies:

$$V_i^{\sigma}\left(w_{it}, \varepsilon_{it}\right) = \max_{\alpha_{it} \in \{0,1\}} \left\{ \pi_{it}(a_{it} \mid w_{it}, \varepsilon_{it}, a_{-it}) + \beta \mathrm{E}_{w_{i,t+1}, \varepsilon_{i,t+1}} \left[\mathrm{V}_i^{\sigma}\left(w_{i,t+1}, \varepsilon_{i,t+1} \mid w_{it}, \varepsilon_{it}, a_{it}\right) \right] \right\}$$

(6)

$\beta \in (0,1)$ is a discount factor.[19] The value function $\mathrm{V}_i^{\sigma}\left(w_{it}, \varepsilon_{it}\right)$ is indexed by strategy profile σ, which implies that it is the value to branch i when all branches in the same market behave according to σ. The inclusion of private information in the model guarantees the existence of at least one pure strategy MPE, as shown by Doraszelski & Satterthwaite (2010).

[19] In our empirical estimation we fix the discount factor to be 0.95.

4.6 Empirical Strategy

In this section, we discuss our model assumptions and estimation strategy. Until recently, estimating a model of dynamic discrete game has been considered a formidable task. The reason is that solving for the equilibrium of the game is too computationally demanding, and the process has to be repeated many times for different parameter sets. To make it even worse, the equilibrium is not guaranteed to be unique. Due to recent developments by (Aguirregabiria & Mira, 2007) and (Bajari, Benkard, & Levin, 2007), two-stage methods have significantly reduced the computational burden. In these methods, reduced form entry/exit policy functions are first estimated directly from the data. By assuming these policy functions to reflect the equilibrium outcome, two-stage methods avoid solving the game fully for every set of parameters encountered. These methods have been applied in recent empirical works, e.g., on price repositioning in the retail industry (Ellickson, Misra, & Nair, 2012), cost of environmental regulation in the cement industry (Ryan, 2012), and demand fluctuations in the ready-mix concrete industry (Collard‐Wexler, 2013). Our empirical strategy follows this tradition. In the following subsections, we first discuss how we estimate the policy functions, and then how we construct the maximum likelihood estimator based on these policy functions. Finally, we discuss how we tackle various computational issues, including dimensionality reduction and the algorithm for computing the transition matrix.

4.6.1 Policy Functions

The first step is the estimation of policy functions governing the entry/exit decisions. As shown in Hotz and Miller (1993) and Aguirregabiria and Mira (2007), a set of strategies can be expressed as its equivalent policy functions or Conditional Choice Probabilities (CCPs). By

assuming the private information to follow an i.i.d. Type-I Extreme-Value distribution, these CCPs are estimated using a simple Logit model:

$$p(a_{it}=1\,|\,\mathrm{w}_{it})=\frac{\exp(\widehat{w}_{it}\theta)}{1+\exp(\widehat{w}_{it}\theta)} \qquad (7)$$

We choose this parametric specification for policy function estimation instead of a non-parametric frequency estimator because in our data entry/exit decisions are unobserved for many states. A similar specification of CCP has been applied in previous literature (Ryan, 2012). To make the CCP specification more flexible, we include both linear and quadratic forms of state variables into $\widehat{w}_{it}$. Similar to the operating/entry cost functions, $\widehat{w}_{it}$ also includes state fixed effects. In addition, we estimate two sets of parameters, one for each type of banks, to allow for different policies for large and small banks.

4.6.2 Estimator

Next we describe the likelihood function for our model estimation. For each possible choice of a_{it}, denote its corresponding choice specific value function as:

$$v_i^{\sigma}\left(w_{it},\alpha_{it}\right)=\tilde{\pi}_{it}(w_{it},a_{it})+\beta\mathrm{E}_{w_{i,t+1}}\left[\tilde{V}_i^{\sigma}(w_{i,t+1}\,|\,w_{it},a_{it})\right] \qquad (8)$$

Here $\tilde{\pi}_{it}(w_{it},a_{it})$ is the expected one-period payoff after dropping the private shock in Equation (2). $\tilde{\mathrm{V}}_i^{\sigma}(w_{i,t+1}\,|\,w_{it},a_{it})$ is the integrated value function derived by integrating out the private shock from the value function in Equation (6):

$$\tilde{\mathrm{V}}_i^{\sigma}(w_{it})\equiv\int \mathrm{V}_i^{\sigma}\left(w_{it},\varepsilon_{it}\right)g(\varepsilon_{it})d\varepsilon_{it} \qquad (9)$$

The integrated value function is the ex-ante value to the branch in state w_{it} before the private information is revealed. Because the policy functions are the best responses, we can rewrite the integrated value function as:

$$\tilde{V}_i^{\sigma}(w_{it}) = \sum_{a_{it}\in\{0,1\}} \left\{ p(a_{it} \mid w_{it}) * \left[\tilde{\pi}_{it}(w_{it}, a_{it}) + \mathrm{E}\left(\varepsilon_{it}(w_{it}, a_{it})\right)\right]\right\} + \beta \mathrm{E}_{w_{i,t+1}}\left[\tilde{V}_i^{\sigma}(w_{i,t+1} \mid w_{it})\right] \qquad (10)$$

$\mathrm{E}\left(\varepsilon_{it}(w_{it}, a_{it})\right)$ is the expectation of the private information conditional on the state w_{it} and that the alternative a_{it} is optimal for the branch. When the private information is Type-I Extreme-Value distributed, $\mathrm{E}\left(\varepsilon_{it}(w_{it}, a_{it})\right)$ equals to $Euler - \ln\left[\mathrm{p}(a_{it} \mid w_{it})\right]$ where $Euler$ is the Euler's constant.

After discretizing the continuous state variables in w_{it}, the calculation of the integrated value function can be written as a set of linear equations:

$$\tilde{V}_i^{\sigma}(w_{it}) = \sum_{a_{it}\in\{0,1\}} \left\{ p(a_{it} \mid w_{it}) * \left[\tilde{\pi}_{it}(w_{it}, a_{it}) + \mathrm{E}\left(\varepsilon_{it}(w_{it}, a_{it})\right)\right]\right\} + \beta \sum_{w_{i,t+1}} \left\{\tilde{V}_i^{\sigma}(w_{i,t+1}) f(w_{i,t+1} \mid w_{it})\right\}$$

(11)

Here $f(w_{i,t+1} \mid w_{it})$ is the probability density of $w_{i,t+1}$ given w_{it}. The solution to this system of linear equations can be written in vector form as:

$$\tilde{\mathrm{V}}^{\sigma} = (I - \beta F)^{-1} \sum_{a_{it}\in\{0,1\}} \left\{ \mathrm{p}(a_{it}) * \left[\tilde{\pi}_{it}(a_{it}) + \mathrm{E}(\varepsilon_{it}(a_{it}))\right]\right\} \qquad (12)$$

I is an identity matrix. *F* is the transition matrix for all the state variables, which we will discuss in detail later. Given the integrated value function, the choice specific value function in Equation (8) is readily available. Our empirical model implies that the optimal decision is:

$$\alpha_i^*(w_i) = \arg\max_{\alpha_i\in\{0,1\}} \left\{ v_i^{\sigma}\left(w_i, \alpha_i\right) + \varepsilon_i\left(\alpha_i\right)\right\} \qquad (13)$$

The probability for a branch choosing to be active in the next stage can be written as:

$$p^*(a_{it}=1|\mathrm{w}_{it})=\frac{\exp\left(v_i^\sigma(w_{it},\alpha_{it}=1)-v_i^\sigma(w_{it},\alpha_{it}=0)\right)}{1+\exp\left(v_i^\sigma(w_{it},\alpha_{it}=1)-v_i^\sigma(w_{it},\alpha_{it}=0)\right)} \quad (14)$$

Matching this probability to our data, we get the log likelihood for our MLE estimator:

$$ll=\sum_t\sum_i \log\left\{a_{it}*p^*(a_{it}=1|\mathrm{w}_{it})+(1-a_{it})*\left[1-p^*(a_{it}=1|\mathrm{w}_{it})\right]\right\} \quad (15)$$

4.6.3 Computational Details

In our estimation, the calculation of the integrated value function in Equation (12) involves invert the matrix $(I-\beta F)$. Unfortunately, direct inversion of this matrix is computationally infeasible because of its high dimension. Note that our state vector w_{it} is 8×1. After discretization it will result in a huge matrix that needs about 3,564GB of computer memory to store.[20] To address this high dimensionality issue, we compute $(I-\beta F)^{-1}$ for each market and bank type, because the market fixed effects from deposit demand estimation and the bank type are invariant over time. This effectively reduces the dimension of each matrix to 38,880×38,880, which needs about 12GB memory and can be handled by a desktop computer.

Since we need to repeat the inversion for 2,208 times (1,104 markets and two types of banks), we take several steps to improve the computational efficiency. First, we program our estimation in a low level programming language.[21] Second, we use an efficient open source library from Basic Linear Algebra Subprograms (BLAS) to compute the inversion. Third, we incorporate Intel Math Kernel Library (MKL) into our program, which can optimize the execution of our code according to the specific structure of Intel CPUs and allow parallel

[20] We discretize *PEN* into 5 points, whereas *POP*, *INC*, market fixed effects into 9 points each. The maximum number of branches in a market from large banks and small banks is 7 and 5, respectively.

[21] We use C++. Another option for scientific computation is Fortran.

computing as appropriate. With these techniques, we effectively reduce the computational time of $(I-\beta F)^{-1}$ to several minutes, compared to several hours with MATLAB on the same computer.

Computation of the transition matrix is based on the discretized state space. The transition kernel for exogenous state variables are computed directly from data. For population and income, we assume that their logarithm transformed values follow AR(1) processes:

$$\log(POP_{mt}) = \eta_m^{pop} + \rho^{pop}\log(POP_{m,t-1}) + u_{mt}^{pop} \qquad (16)$$

$$\log(INC_{mt}) = \eta_m^{inc} + \rho^{inc}\log(INC_{m,t-1}) + u_{mt}^{inc} \qquad (17)$$

In Equation (16) and (17), we assume the autoregressive parameters ρ^{pop} and ρ^{inc} are homogenous but the mean values η_m^{pop} and η_m^{inc} may vary across markets. Because these AR(1) processes are highly persistent[22], we use the method developed by Rouwenhorst (1995) to obtain the transition kernel of population and income. The transition kernel for penetration is computed directly from the observed data using a non-parametric frequency method.

The transition of market structure variables are computed based on the estimated policy functions. Given that all players are making decisions simultaneously and independently, the expected number of active branches in the next period is distributed binomially according to the CCPs. For example, the probability that $n_{m,t+1}^L$ large bank branches will operate in the next period is given by:

$$f^{\sigma}(n_{m,t+1}^L \mid w_{it}) = \sum_{n=0}^{n_{m,t+1}^L}\left\{B(n; n_{mt}^L, p_1) * B(n_{m,t+1}^L - n; N_{mt}^L - n_{mt}^L, p_2)\right\} \qquad (18)$$

---

[22] ρ^{pop} = 0.9996 and ρ^{inc} = 0.9656 from our estimation.

Here p_1 is the probability for a large incumbent to continue operating in the market, and p_2 is the probability for a large potential entrant to enter the market. Both p_1 and p_2 can be calculated directly from the estimated policy functions. $B(k;n,p)=\binom{n}{k}p^k(1-p)^{n-k}$ is the binomial probability density function and $\binom{n}{k}=\frac{n!}{k!(n-k)!}$. So $B(n;n_{mt}^L,p_1)$ is the probability that n out of n_{mt}^L incumbent large bank branches continue operating in the next period, and $B(n_{m,t+1}^L-n;N_{mt}^L-n_{mt}^L,p_2)$ is the probability that $(n_{m,t+1}^L-n)$ out of $(N_{mt}^L-n_{mt}^L)$ potential entrants enter the market, where N_{mt}^L is the total number of branches from large banks including incumbents and potential entrants.[23] The probabilities for the number of small bank branches can be computed in a similar way.

4.7 Results

4.7.1 Deposit Demand Parameters

Table 21 shows the results from the deposit demand regression. As one might expect, all the competition effects are negative. For example, when a market changes from a monopoly with one large bank branch to a duopoly with two large bank branches, the expected deposits at each branch decrease by (1 - exp(α_4^L)) = 22.10%. Similarly, when a market changes from a monopoly with one large bank branch to a duopoly with one large and one small bank branch, the expected deposit for the large branch decreases by (1 - exp(α_7^L)) = 19.29%. The coefficients of population

[23] As mentioned previously, for empirical estimation we assume one potential entrant of each type.

for large and small banks are positive and comparable in magnitude. The coefficients of income are also positive as expected.

More relevant and interesting results come from the coefficients of broadband penetration, which is positive and significant for large banks (α_3^L = 0.1396, p = 0.04), but insignificant for small banks (α_3^S = -0.0265, p = 0.67). Our result shows that a 10% increase in broadband penetration leads to a 1.41% increase in deposit for large banks, but has little impact on small banks. It may indicate that broadband consumers value the better online banking services from large banks and make more deposits in them. Our results are also consistent with a recent survey by ath Power Consulting, which shows that larger banks are improving their customer service at a faster rate than local community banks over the past few years because of their superior technology offerings.[24]

Table 21. Estimates for deposit demand

	Meaning	Coef.	Std. Err.	t	P>t
Large Bank					
α_1^L	Population (10,000)	1.0183	0.5394	1.89	0.06
α_2^L	Income ($10,000)	0.0284	0.0536	0.53	0.60
α_3^L	Penetration	0.1396	0.0678	2.06	0.04
α_4^L	First large competitor	-0.2498	0.0526	-4.75	0.00
α_5^L	Second large Competitor	-0.3391	0.0617	-5.50	0.00
α_6^L	Number of large competitors - 2	-0.1611	0.0569	-2.83	0.01
α_7^L	First small competitor	-0.2143	0.0530	-4.05	0.00
α_8^L	Second small Competitor	-0.3153	0.0505	-6.24	0.00
α_9^L	Number of small competitors - 2	-0.2803	0.0516	-5.44	0.00
Small Bank					
α_1^S	Population	0.8601	0.5410	1.59	0.11
α_2^S	Income	0.1015	0.0532	1.91	0.06

[24] http://www.americanbanker.com/news/consumer-finance/big-banks-use-technology-to-boost-customer-satisfaction-1073118-1.html

α_3^S	Penetration	-0.0265	0.0620	-0.43	0.67
α_4^S	First large competitor	-0.2030	0.0541	-3.75	0.00
α_5^S	Second large Competitor	-0.3009	0.0536	-5.62	0.00
α_6^S	Number of large competitors - 2	-0.1987	0.0493	-4.03	0.00
α_7^S	First small competitor	-0.2182	0.0512	-4.26	0.00
α_8^S	Second small Competitor	-0.1847	0.0513	-3.60	0.00
α_9^S	Number of small competitors - 2	-0.3763	0.0639	-5.89	0.00

4.7.2 *Structural Parameters for Entry and Exit Decisions*

We have estimated four models with different configurations. In model M1, we omit state fixed effects for operating and entry costs, and set the revenue coefficients of deposits for both types of banks to be the same ($\alpha_L = \alpha_S$). In model M2, we include state fixed effects, but still keep the revenue coefficients to be the same. In model M3, we omit state fixed effects, but estimate different revenue coefficients for large and small banks. In model M4, we include state fixed effects and allow revenue coefficients to be different. For subsequent discussions we focus on the results from M4. Table 22 shows the results.

Table 22. Estimates for structural parameters

		M1	M2	M3	M4
β_1^L	Penetration on large bank operating costs	-0.3326***	-0.3116***	-0.2969***	-0.2810**
		(0.1097)	(0.1192)	(0.1111)	(0.1202)
β_1^S	Penetration on small bank operating costs	-0.1086	-0.0750	-0.1509	-0.1119
		(0.1052)	(0.1149)	(0.1073)	(0.1166)
γ_1^L	Penetration on large bank entry costs	0.5789	0.7285*	0.6231*	0.7806**
		(0.3675)	(0.3859)	(0.3655)	(0.3847)
γ_1^S	Penetration on small bank entry costs	0.4448	0.0125	0.4025	-0.0255
		(0.4920)	(0.5366)	(0.4925)	(0.5369)
α_L	Large bank deposit			0.0029***	0.0026***
		0.0011***	0.0010***	(0.0008)	(0.0008)
		(0.0004)	(0.0004)	0.0007*	0.0007*
α_S	Small bank deposit			(0.0004)	(0.0004)

State fixed effects for operating costs	N/A	Included	N/A	Included
State fixed effect for entry costs	N/A	Included	N/A	Included

* p<.10; ** p<.05; *** p<.01.

Deposit. The coefficients are positive and significant (α_L = 0.0026, p < 0.01; α_S =0.0007, p < 0.01). It is not surprising because one of the major functions for bank branches is to collect deposits from customers. With more deposits, banks earn more profits by making loans to firms or households, or investing in other financial products.

Operating Costs. Our results show that the effects of broadband penetration on banks' operating costs are asymmetric for large and small banks. While it can significantly reduce operating costs and increase profits for large banks, it has little impact on small banks. This is consistent with previous findings in the banking literature (Hunter & Timme, 1991; Maudos, Pastor, & Quesada, 1996; Mckillop, Glass, & Morikawa, 1996), and with Schumpeter's theoretical conjecture that technology is generally a scale-increasing factor to market structure change (Schumpeter, 1942b). To quantify the effect of broadband penetration on operating costs, we divide β_1^L by α_L, which shows that for large banks a 1% increase in broadband penetration leads to a saving in operating costs equivalent to profits from $1 million deposits.

Our results are consistent with the data from Reports of Condition and Income (Call Report), which is filed by all regulated financial institutions to FDIC on a quarterly basis and contains basic financial information such as balance sheet and income statement. The average noninterest expense per branch decreased by 13.55% for large banks, but only 2.69% for small banks from 2008 to 2013. In terms of labor costs, the average number of employees per branch

decreased by 40.04% for large banks, but only 8.01% for small banks in the same period.[25] Clearly the operational efficiency of large banks improved significantly in recent years. Our estimation results suggest that the technology change may be an important reason.

Entry Cost. Our estimated coefficients for entry costs confirm the old adage that there isn't such a thing as a free lunch. While large banks benefit from broadband penetration through the reduction in operating costs, they pay higher entry costs than smaller banks as broadband penetration increases. Our finding is consistent with the observation that IT related investment measured as the percentage of total assets for the 45 largest banks in the U.S. surpassed that of small banks (Cooke, 1997). Comparing the coefficients of penetration on operating costs (β_1^L) and on entry costs (γ_1^L), large banks may recover the higher entry costs from broadband penetration in about 3 years.

4.8 Counterfactual Simulations

One important advantage of estimating a structural model is that it enables researchers to conduct policy experiments with the estimated primitives. Our primary interest in this research is to disentangle the effect of broadband penetration on deposit demand, branch operating costs and entry costs in an effort to understand how the U.S. banking industry evolves with IT technology advancement. To achieve this goal, we analyze how different factors contribute to the change of market structure from 2008 to 2013. In the following analysis, we set all effects of penetration on small banks to be zero ($\alpha_3^S = \beta_1^S = \gamma_1^S = 0$) because their estimated results are not statistically significant. In effect we focus on the effects of broadband penetration on large banks only.

[25] The total noninterest expense and number of employees come from RIAD4093 and RIAD4150 in Call Reports.

Specifically, we conduct five counterfactual experiments. In our first experiment (C1), we implement a counterfactual analysis with zero effect of broadband penetration on deposit demand, operating costs and entry costs ($\alpha_3^L = \beta_1^L = \gamma_1^L = 0$). We will use the results of the first experiment as a benchmark to quantify the relative contribution of demand side and cost side factors. In the second experiment (C2), we use the estimated coefficient of broadband penetration on deposit demand while keeping other coefficients of penetration to zero, which enables us to quantify how the effect of penetration on deposit demand contributes to the entry/exit patterns. In the third experiment (C3), we use the estimated effect of broadband penetration on operating costs to show how the change in penetration affects market structure through changing operating cost. In the fourth experiment (C4), we consider a counterfactual analysis with estimated effect of penetration on entry costs only. Finally in (C5), we use all the estimated coefficients to simulate how the market structure evolves over time. In all simulations, we use the observed information at the beginning of 2008 as a starting point, and simulate the market evolution for 5 years. In each year, we compute the MPE of our model for all the 1,104 local markets.

We report the average results of 500 simulation runs in Table 23 and Table 24. Table 23 shows the entry/exit probabilities, while Table 24 shows the average number of new entrants and exits. Because the patterns are similar in the two tables, we focus on the results in Table 23. The first half of Table 23 shows the average exit probabilities for incumbent branches and the average entry probabilities for potential entrants. The second half of the table shows percentage change of these probabilities relative to the base model (C1) where the effects of broadband penetration are fixed to be zero.

Table 23. Counterfactual simulations for entry/exit probabilities

		C1 (Base)	C2 (Demand)	C 3 (Operation)	C 4 (Entry)	C 5 (Aggregate)
Probability						
	Exit – Large Bank	2.07%	1.97%	0.51%	1.92%	0.39%
	Entry – Large Bank	1.73%	1.83%	6.59%	1.38%	6.20%
	Exit – Small Bank	3.30%	3.31%	3.42%	3.30%	3.41%
	Entry – Small Bank	1.04%	1.04%	1.01%	1.04%	1.01%
Change from C1						
	Exit – Large Bank		-0.10%	-1.56%	-0.15%	-1.68%
	Entry – Large Bank		0.10%	4.87%	-0.34%	4.47%
	Exit – Small Bank		0.01%	0.12%	0.00%	0.11%
	Entry – Small Bank		0.00%	-0.03%	0.00%	-0.03%

Table 24. Counterfactual simulations for the number of branches

		C1 (Base)	C2 (Demand)	C 3 (Operation)	C 4 (Entry)	C 5 (Aggregate)
Probability						
	Exit – Large Bank	23.74	22.64	6.56	21.89	4.96
	Entry – Large Bank	19.09	20.20	72.74	15.38	68.39
	Exit – Small Bank	50.80	51.11	52.55	50.96	52.56
	Entry – Small Bank	11.58	11.42	10.99	11.66	11.14
Change from C1						
	Exit – Large Bank		-1.10	-17.18	-1.85	-18.78
	Entry – Large Bank		1.11	53.65	-3.71	49.30
	Exit – Small Bank		0.31	1.75	0.16	1.76
	Entry – Small Bank		-0.16	-0.59	0.08	-0.44

Deposit demand. Because large banks may receive more consumer deposits with higher broadband penetration, the exit rate of large banks decreases by 0.10% from 2.07% in C1 to 1.97% in C2. Meanwhile, the entry rate of large banks increases by 0.10% from 1.73% in C1 to 1.83% in C2. Given that there are 1,159 large bank branches in 1,104 markets in 2008, this simulation results imply an increase of 23 large bank branches in the 5 years due to the effect of broadband penetration on deposit demand. On the small bank side, entry/exit probabilities change slightly due to competition effects from large banks.

Operating Costs. Based on our simulation results, the effect of broadband penetration on operating costs turns out to be the most important factor driving the market structure change. A comparison of C3 to C1 shows that, with the increase in broadband penetration, the exit rate of large bank branches decreases by 1.56% from 2.07% to 0.51%, while the entry rate increases by 4.87% from 1.73% to 6.59%. The reduction in operating costs from online banking leads to 72 more branches for large banks. At the same time, this hurts small banks and results in a 0.12% higher exit rate and a 0.03% lower entry rate.

Entry Costs. Higher entry costs for large banks have asymmetric effects on incumbent branches and potential entrants. As shown in C4, the exit rate of large incumbent branches decreases by 0.15%, and the entry rate of large potential entrants also decreases by 0.34%. The change in the entry rates can be explained by the increase in entry costs. The change in exit rate illustrates a nice feature of our dynamic model: because the incumbent branches expect less competition from fewer entrants and the increase in re-entry costs in the future once they exit, the incumbent branches will be more likely to stay in the market. Due to the small change in the

number of large bank branches, the impact on the entry/exit rate of small banks is minimal in our simulation.

Aggregate Effect. As shown in C5, the aggregate effect of broadband penetration is consistent with the observed market structure change: an increase in large bank branches and a decrease in small bank branches. In summary, our simulation results show that the reduction in operating costs for large banks is the most important factor driving the recent market structure change in the U.S. banking industry.

4.9 Conclusion

In this research, we propose and estimate a dynamic discrete game framework to understand how technology advancement affects the evolution of market structure. A nice feature of this framework is that it can disentangle the contribution of demand, operating cost, and entry cost factors to the market structure change. When applying this framework to the U.S banking industry, we address the endogeneity of technology investment by collecting data on the exogenous technology penetration.

Our estimation results show that broadband penetration has asymmetric effects on banks of different sizes. Specifically, large banks can take advantage of the increasingly higher residential broadband penetration in the U.S. by investing more on online banking services. As a result, they can improve efficiency and reduce costs in operating offline branches. Moreover, they can attract more customers and receive more deposits by offering better online banking services. Our counterfactual analysis shows that the reduction in operating costs in the most significant factor driving the recent change in the U.S. banking industry, followed by higher entry costs and deposit demand for large banks.

Our research has important implications for bank managers and policy makers. On the one hand, it may be appropriate to revisit the Riegle-Neal Act which requires that no bank holding company can control more than 10 percent of the nation's total deposits. To the extent that the increased concentration is a natural result of technology change, the strict limit on the deposit share of large banks might negatively affect social welfare. On the other hand, small banks may need a better strategy to benefit from the increasing Internet penetration. Given that the long term return from the improved efficiency in large banks can easily justify the increased entry costs, smaller banks can learn from the successful practice of large banks.

Several limitations of this study may warrant future research. First, we use a reduced form approach to model the deposit demand because we do not observe branch-level interest rates and loans. Given adequate data one might develop a more accurate model of local competition on banking services. Second, we abstract from any network effects that may be related to the entry/exit decision of a bank in a specific market (Bresnahan & Reiss, 1991; Mazzeo, 2002). A model with global players competing in local markets is beyond the scope of this research. Third, we do not explicitly model mergers and acquisitions in this framework. Future research may incorporate mergers and acquisitions and see how technology advancement affects merger decisions in the U.S. banking industry.

Chapter 5 Conclusions and Future Research

This dissertation provides a comprehensive investigation on both consumer and firm behaviors in the multi-channel environment. Our results highlight the importance of considering consumers' cross channel behaviors in designing firm strategies.

Although we have rich findings in the three essays, there are some important limitations in this dissertation that deserve future research. For example, my dissertation focuses on online and offline channel only because of data limitations. However, more and more consumers do not discriminate between the tradition online and the new mobile channels in information search and marking purchases. With uninterrupted access to consumers, the mobile channel offers an exciting opportunity for marketers to build valuable relationships with consumers by identifying their personal identities, geographic locations, and social communication patterns. Unfortunately, Future research can extend our ideas to evaluate how consumers and firms interact in the online, offline and mobile channels. We believe our basic frameworks, such as the nested logit structure in the first essay and the entry/exit model, still hold after including the mobile channel. Other interesting future research may include: to investigate how the offline store entry affects consumers' purchase decisions as well as information search; to formally model the merge and acquisition in the banking industry after the introduction of online banking.

In summary, we hope this dissertation will inspire more studies on consumer and firm behaviors in the multi-channel environment, contribute to the Marketing literature and provide managerial implications for firms when introducing more channels.

Bibliography

Abadie, A., & Imbens, G. W. 2006. Large Sample Properties of Matching Estimators for Average Treatment Effects. ***Econometrica***, 74(1): 235-267.

Advisors, C. o. E. 2001. ***The Annual Report of the Council of Economic Advisors***. Washington, DC.: US Government Printing Office.

Aguirregabiria, V., & Mira, P. 2007. Sequential Estimation of Dynamic Discrete Games. ***Econometrica***, 75(1): 1-53.

Aguirregabiria, V., & Suzuki, J. 2014. Identification and Counterfactuals in Dynamic Models of Market Entry and Exit. ***Quantitative Marketing and Economics***, 12(3): 267-304.

Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., & Wood, S. 1997. Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces. ***Journal of Marketing***, 61(3): 38-53.

Altunbas, Y., Goddard, J., & Molyneux, P. 1999. Technical Change in Banking. ***Economics Letters***, 64(2): 215-221.

Anderson, E. T., Fong, N. M., Simester, D. I., & Tucker, C. E. 2010. How Sales Taxes Affect Customer and Firm Behavior: The Role of Search on the Internet. ***Journal of Marketing Research***, 47(2): 229-239.

Avery, J., Steenburgh, T. J., Deighton, J., & Caravella, M. 2012. Adding Bricks to Clicks: Predicting the Patterns of Cross-Channel Elasticities over Time. ***Journal of Marketing***, 76(5): 96-111.

Bajari, P., Benkard, C. L., & Levin, J. 2007. Estimating Dynamic Models of Imperfect Competition. ***Econometrica***, 75(5): 1331-1370.

Balasubramanian, S. 1998. Mail Versus Mall: A Strategic Analysis of Competition between Direct Marketers and Conventional Retailers. ***Marketing Science***, 17(3): 181-195.

Bapna, R., Goes, P., Gupta, A., & Jin, Y. W. 2004. User Heterogeneity and Its Impact on Electronic Auction Market Design: An Empirical Exploration. ***MIS Quarterly***, 28(1): 21-43.

Baumol, W. J., Blinder, A. S., & Wolff, E. N. 2003. ***Downsizing in America: Reality, Causes, and Consequences***: Russell Sage Foundation.

BEA. 2013. Gdp by Industry 1997-2013.

Beccalli, E. 2007. Does It Investment Improve Bank Performance? Evidence from Europe. ***Journal of Banking & Finance***, 31(7): 2205-2230.

Bell, D. R., HO, T. H., & TANG, C. S. 1998. Determining Where to Shop: Fixed and Variable Costs of Shopping. ***Journal of Marketing Research***, 35(3): 352-369.

Bell, D. R., & Song, S. 2007. Neighborhood Effects and Trial on the Internet: Evidence from Online Grocery Retailing. ***Quantitative Marketing and Economics***, 5(4): 361-400.

Berger, I. E., & Mitchell, A. A. 1989. The Effect of Advertising on Attitude Accessibility, Attitude Confidence, and the Attitude-Behavior Relationship. ***Journal of Consumer Research***, 16(3): 269-279.

Berry, S. T. 1992. Estimation of a Model of Entry in the Airline Industry. ***Econometrica: Journal of the Econometric Society***: 889-917.

Besley, S., & Brigham, E. 2011. ***Principles of Finance***: Cengage Learning.

Bitner, M. J., Brown, S. W., & Meuter, M. L. 2000. Technology Infusion in Service Encounters. ***Journal of the Academy of Marketing Science***, 28(1): 138-149.

Bresnahan, T. F., & Reiss, P. C. 1991. Entry and Competition in Concentrated Markets. ***Journal of Political Economy***: 977-1009.

Brynjolfsson, E. 1993. The Productivity Paradox of Information Technology. ***Communications of the ACM***, 36(12): 66-77.

Brynjolfsson, E., & Hitt, L. M. 2000. Beyond Computation: Information Technology, Organizational Transformation and Business Performance. ***The Journal of Economic Perspectives***: 23-48.

Brynjolfsson, E., Hu, Y., & Rahman, M. S. 2009. Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition. ***Management Science***, 55(11): 1755-1765.

Brynjolfsson, E., Malone, T. W., Gurbaxani, V., & Kambil, A. 1994. Does Information Technology Lead to Smaller Firms? ***Management Science***, 40(12): 1628-1644.

Brynjolfsson, E., & Smith, M. D. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. ***Management Science***, 46(4): 563-585.

Bucklin, R. E., & Sismeiro, C. 2003. A Model of Web Site Browsing Behavior Estimated on Clickstream Data. ***Journal of Marketing Research***, 40(3): 249-267.

Campbell, D., & Frei, F. 2010. Cost Structure, Customer Profitability, and Retention Implications of Self-Service Distribution Channels: Evidence from Customer Behavior in an Online Banking Channel. ***Management Science***, 56(1): 4-24.

Casolaro, L., & Gobbi, G. 2007. Information Technology and Productivity Changes in the Banking Industry. ***Economic Notes***, 36(1): 43-76.

Collard-Wexler, A. 2013. Demand Fluctuations in the Ready-Mix Concrete Industry. ***Econometrica***, 81(3): 1003-1037.

Cooke, S. D. 1997. Structural Change in the Us Banking Industry: The Role of Information Technology. ***Washington: US Department of Commerce. Retrieved August***, 29: 2003.

Cronin, M. J. 1997. ***Banking and Finance on the Internet***: John Wiley & Sons.

Curran, J. M., Meuter, M. L., & Surprenant, C. F. 2003. Intentions to Use Self-Service Technologies: A Confluence of Multiple Attitudes. ***Journal of Service Research***, 5(3): 209-224.

Dabholkar, P. A. 1991. Using Technology-Based Self-Service Options to Improve Perceived Service Quality. ***Enhancing knowledge development in marketing***: 534-535.

Daly, M. J., & Rao, P. S. 1985. Productivity, Scale Economies, and Technical Change in Ontario Hydro. ***Southern Economic Journal***: 167-180.

De Elejalde, R. 2009. Local Entry Decisions in the Us Banking Industry: Working Paper, Universidad Carlos III de Madrid.

Deleersnyder, B., Geyskens, I., Gielens, K., & Dekimpe, M. G. 2002. How Cannibalistic Is the Internet Channel? A Study of the Newspaper Industry in the United Kingdom and the Netherlands. ***International Journal of Research in Marketing***, 19(4): 337-348.

Diamond, A., & Sekhon, J. S. 2013. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. ***Review of Economics and Statistics***, 95(3): 932-945.

Doraszelski, U., & Satterthwaite, M. 2010. Computable Markov-Perfect Industry Dynamics. ***The RAND Journal of Economics***, 41(2): 215-243.

Dunne, T., Klimek, S. D., Roberts, M. J., & Xu, D. Y. 2013. Entry, Exit, and the Determinants of Market Structure. ***The RAND Journal of Economics***, 44(3): 462-487.

ECB. 1999. ***The Effects of Technology on the Eu Banking Systems***. Frankfurt: European Central Bank.

Ellickson, P. B., Misra, S., & Nair, H. S. 2012. Repositioning Dynamics and Pricing Strategy. ***Journal of Marketing Research***, 49(6): 750-772.

Ericson, R., & Pakes, A. 1995a. Markov-Perfect Industry Dynamics: A Framework for Empirical Work. ***The Review of Economic Studies***, 62(1): 53-82.

Ericson, R., & Pakes, A. 1995b. Markov-Perfect Industry Dynamics: A Framework for Empirical Work. ***The Review of Economic Studies***, 62(1): 53.

Feng, G., & Serletis, A. 2010. Efficiency, Technical Change, and Returns to Scale in Large Us Banks: Panel Data Evidence from an Output Distance Function Satisfying Theoretical Regularity. ***Journal of Banking & Finance***, 34(1): 127-138.

Ferrier, G. D., & Lovell, C. K. 1990. Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence. ***Journal of econometrics***, 46(1): 229-245.

Forman, C., Ghose, A., & Goldfarb, A. 2009. Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live. ***Management Science***, 55(1): 47-57.

Forman, C., Ghose, A., & Wiesenfeld, B. 2008. Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. ***Information Systems Research***, 19(3): 291-313.

Frambach, R. T., Roest, H. C. A., & Krishnan, T. V. 2007. The Impact of Consumer Internet Experience on Channel Preference and Usage Intentions across the Different Stages of the Buying Process. ***Journal of Interactive Marketing***, 21(2): 26-41.

Gallino, S., & Moreno, A. 2014. Integration of Online and Offline Channels in Retail: The Impact of Sharing Reliable Inventory Availability Information. ***Management Science***, 60(6): 1434-1451.

Goldfarb, A., & Tucker, C. E. 2011. Privacy Regulation and Online Advertising. ***Management Science***, 57(1): 57-71.

Goldmanis, M., Hortacsu, A., Syverson, C., & Emre, Ö. 2010. E-Commerce and the Market Structure of Retail Industries*. ***The Economic Journal***, 120(545): 651-682.

Green, A., Harris, C., & Mayes, D. 1991. Estimation of Technical Inefficiency in Manufacturing Industry. ***Applied Economics***, 23(10): 1637-1647.

Hansen, B. B. 2004. Full Matching in an Observational Study of Coaching for the Sat. ***Journal of the American Statistical Association***, 99(467): 609-618.

Hauser, J. R., & WERNERFELT, B. 1990. An Evaluation Cost Model of Consideration Sets. ***Journal of Consumer Research***, 16(4): 393-408.

Hernando, I., & Nieto, M. J. 2007. Is the Internet Delivery Channel Changing Banks' Performance? The Case of Spanish Banks. ***Journal of Banking & Finance***, 31(4): 1083-1099.

Herr, P. M., KARDES, F. R., & KIM, J. 1991. Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. ***Journal of Consumer Research***, 17(4): 454-462.

Hitt, L., Xue, M., & Chen, P.-Y. 2007. The Determinants and Outcomes of Internet Banking Adoption. ***MSI Reports***: 07-122.

Hitt, L. M., & Brynjolfsson, E. 1996. Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value. ***MIS quarterly***: 121-142.

Hitt, L. M., & Frei, F. X. 2002. Do Better Customers Utilize Electronic Distribution Channels? The Case of Pc Banking. ***Management Science***, 48(6): 732-748.

Hoque, A. Y., & Lohse, G. L. 1999. An Information Search Cost Perspective for Designing Interfaces for Electronic Commerce. ***Journal of Marketing Research (JMR)***, 36(3): 387-394.

Hotz, V. J., & Miller, R. A. 1993. Conditional Choice Probabilities and the Estimation of Dynamic Models. ***The Review of Economic Studies***, 60(3): 497-529.

Huang, P., Lurie, N. H., & Mitra, S. 2009. Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods. ***Journal of Marketing***, 73(2): 55-69.

Huff, D. L. 1964. Defining and Estimating a Trading Area. ***Journal of Marketing***, 28(3): 34-38.

Hunter, W. C., & Timme, S. G. 1991. Technological Change in Large Us Commercial Banks. ***Journal of Business***: 339-362.

Jank, W., & Kannan, P. K. 2005. Understanding Geographical Markets of Online Firms Using Spatial Models of Customer Choice. ***Marketing Science***, 24(4): 623-634.

Johnson, E. J., Bellman, S., & Lohse, G. L. 2003. Cognitive Lock-in and the Power Law of Practice. ***Journal of Marketing***, 67(2): 62-75.

Kahneman, D., & Tversky, A. 1984. Choices, Values, and Frames. ***American Psychologist***, 39(4): 341-350.

Kambil, A., & van Heck, E. 1998. Reengineering the Dutch Flower Auctions: A Framework for Analyzing Exchange Organizations. ***Information Systems Research***, 9(1): 1-19.

Kang, Y. S., HERR, P. M., & PAGE, C. M. 2003. Time and Distance: Asymmetries in Consumer Trip Knowledge and Judgments. ***Journal of Consumer Research***, 30(3): 420-429.

Kardes, F. R., Kalyanaram, G., Chandrashekaran, M., & Dornoff, R. J. 1993. Brand Retrieval, Consideration Set Composition, Consumer Choice, and the Pioneering Advantage. ***Journal of Consumer Research***, 20(1): 62-75.

Kirmani, A., & Rao, A. R. 2000. No Pain, No Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality. ***Journal of Marketing***, 64(2): 66-79.

Kohl's Corporation. 2010. 10-K Annual Report.

Kumbhakar, S. C. 1987. The Specification of Technical and Allocative Inefficiency in Stochastic Production and Profit Frontiers. ***Journal of Econometrics***, 34(3): 335-348.

Lal, R., & Sarvary, M. 1999. When and How Is the Internet Likely to Decrease Price Competition? ***Marketing Science***, 18(April): 485-503.

Lang, G., & Welzel, P. 1996. Efficiency and Technical Progress in Banking Empirical Results for a Panel of German Cooperative Banks. ***Journal of Banking & Finance***, 20(6): 1003-1023.

Liu, Y., Gupta, S., & Zhang, Z. J. 2006. Note on Self-Restraint as an Online Entry-Deterrence Strategy. ***Management Science***, 52(11): 1799-1809.

Maloney, M. T. 2001. Economies and Diseconomies: Estimating Electricity Cost Functions. ***Review of Industrial Organization***, 19(2): 165-180.

Mansfield, E. 1983. Technological Change and Market Structure: An Empirical Study. ***The American Economic Review***, 73(2): 205-209.

Maudos, J., Pastor, J. M., & Quesada, J. 1996. ***Technical Progress in Spanish Banking: 1985-1994***: Institut Valencià d'Investigacions Econòmiques.

Mazzeo, M. J. 2002. Product Choice and Oligopoly Market Structure. ***RAND Journal of Economics***: 221-242.

McKillop, D. G., Glass, J. C., & Morikawa, Y. 1996. The Composite Cost Function and Efficiency in Giant Japanese Banks. ***Journal of Banking & Finance***, 20(10): 1651-1671.

Moe, W. W., & Fader, P. S. 2004. Dynamic Conversion Behavior at E-Commerce Sites. ***Management Science***, 50(March): 326-335.

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. 2004. Modeling Online Browsing and Path Analysis Using Clickstream Data. ***Marketing Science***, 23: 579-595.

Moorthy, S., Ratchford, B. T., & Talukdar, D. 1997. Consumer Information Search Revisited: Theory and Empirical Analysis. ***Journal of Consumer Research***, 23(4): 263-277.

Murdie, R. A. 1965. Cultural Differences in Consumer Travel. ***Economic Geography***, 41(3): 211-233.

Murray, J. D., & White, R. W. 1983. Economies of Scale and Economies of Scope in Multiproduct Financial Institutions: A Study of British Columbia Credit Unions. ***The Journal of Finance***, 38(3): 887-902.

Nam, S., & Ellinger, P. N. 2008. Branch Expansion of Commercial Banks in Rural America. ***Urbana***, 51: 61801.

Nedungadi, P. 1990. Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations. ***Journal of Consumer Research***, 17(3): 263-276.

Nelson, P. 1970. Information and Consumer Behavior. ***Journal of Political Economy***, 78(March-April): 311-329.

Nelson, P. 1974. Advertising as Information. ***Journal of Political Economy***, 82(July-August): 729-754.

Nicholas, D., Huntington, P., Jamali, H. R., & Dobrowolski, T. 2007. Characterising and Evaluating Information Seeking Behaviour in a Digital Environment: Spotlight on the 'Bouncer'. ***Information Processing & Management***, 43(4): 1085-1102.

Okada, E. M. 2005. Justification Effects on Consumer Choice of Hedonic and Utilitarian Goods. ***Journal of Marketing Research***, 42(1): 43-53.

Overby, E., & Jap, S. 2009. Electronic and Physical Market Channels: A Multiyear Investigation in a Market for Products of Uncertain Quality. ***Management Science***, 55(6): 940-957.

Pakes, A., Ostrovsky, M., & Berry, S. 2007. Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples). ***The RAND Journal of Economics***, 38(2): 373-399.

Peck, J., & Childers, T. L. 2003. Individual Differences in Haptic Information Processing: The “Need for Touch” Scale. ***Journal of Consumer Research***, 30(3): 430-442.

Powell, T. C., & Dent-Micallef, A. 1997. Information Technology as Competitive Advantage: The Role of Human, Business, and Technology Resources. ***Strategic management journal***, 18(5): 375-405.

Ratchford, B. T., Lee, M.-S., & Talukdar, D. 2003. The Impact of the Internet on Information Search for Automobiles. ***Journal of Marketing Research***, 40(2): 193-209.

Rosenbaum, P. R. 2002. Observational Studies, ***Observational Studies***: 1-17: Springer.

Rouwenhorst, K. G. 1995. Asset Pricing Implications of Equilibrium Business Cycle Models. In T. F. Cooley (Ed.), ***Frontiers of Business Cycle Research***: Princeton University Press.

Ryan, S. P. 2012. The Costs of Environmental Regulation in a Concentrated Industry. ***Econometrica***, 80(3): 1019-1061.
Saunders, A. 2000. ***Financial Institutions Management: A Modern Perspective***: McGraw-Hill College.
Schmookler, J. 1966. Invention and Economic Growth.
Schumpeter, J. A. 1942a. Capitalism, Socialism and Democracy.
Schumpeter, J. A. 1942b. ***Socialism, Capitalism and Democracy***: Harper and Brothers.
Seim, K. 2006. An Empirical Model of Firm Entry with Endogenous Product-Type Choices. ***The RAND Journal of Economics***, 37(3): 619-640.
Shapiro, S., MacInnis, D. J., & Heckler, S. E. 1997. The Effects of Incidental Ad Exposure on the Formation of Consideration Sets. ***Journal of Consumer Research***, 24(1): 94-104.
Shugan, S. M. 1980. The Cost of Thinking. ***Journal of Consumer Research***, 7(2): 99-111.
Singh, V., & Zhu, T. 2008. Pricing and Market Concentration in Oligopoly Markets. ***Marketing Science***, 27(6): 1020-1035.
Thatcher, M. E., & Oliver, J. R. 2001. The Impact of Technology Investments on a Firm's Production Efficiency, Product Quality, and Productivity. ***Journal of Management Information Systems***, 18(2): 17-45.
US Census. 2010. Quarterly Retail E-Commerce Sales: 4th Quarter 2009. Washington D.C.: US Census.
US Department of Transportation. 2008. Transportation Statistics Annual Report. Washington D.C.: US Department of Transportation.
Venkatesan, R., Kumar, V., & Ravishanker, N. 2007. Multichannel Shopping: Causes and Consequences. ***Journal of Marketing***, 71(2): 114-132.
Verhoef, P. C., Neslin, S. A., & Vroomen, B. 2007. Multichannel Customer Management: Understanding the Research-Shopper Phenomenon. ***International Journal of Research in Marketing***, 24(2): 129-148.
Vernon, R. 1966. International Investment and International Trade in the Product Cycle. ***The quarterly journal of economics***: 190-207.
Wangenheim, F. v., & Bayón, T. 2007. Behavioral Consequences of Overbooking Service Capacity. ***Journal of Marketing***, 71(4): 36-47.
Whitney, L. 2009. Online Banking Is Booming: CNET.
Zentner, A., Smith, M., & Kaya, C. 2013. How Video Rental Patterns Change as Consumers Move Online. ***Management Science***, 59(11): 2622-2634.
Zettelmeyer, F. 2000. Expanding to the Internet: Pricing and Communications Strategies When Firms Compete on Multiple Channels. ***Journal of Marketing Research***, 37(3): 292-308.

Appendix

Appendix A. Selection of Nest Structure (Chapter Three)

Our nest structure is chosen by comparing the model fit between alternative nest structures. By putting form factor, channel, and firm onto different levels, there could be six alternative two-level structures and another six alternative three-level structures. We estimate all the twelve specifications and select the one that provides the best fit to our data. It is worth noting that a nested logit model corresponds to certain correlation structure in the error term, rather than imposing a customer choice process. For example, Hensher et al. (2005) point out that "…the entire purpose in creating a nested form is to try and accommodate violation of IID/IIA. It has nothing to do with any behavioral belief in the way that alternatives are assessed in the process of making a choice. This distinction is very important because many analysts think that nested choice models are a way of defining the decision process that links behavioral choices." Researchers have shown how a nested logit model can be developed from correlations in the variance components without any assumptions on the customer decision process (e.g. Berry 1994; Cardell 1997). We therefore follow Sriram et al. (2006) to use the sum of squared errors as the criterion for comparison. As reported in Table A.1, the nest structure depicted in Figure 1 gives the smallest sum of squared errors. We believe this specification captures reasonably well the correlations between sub-sets of alternative PC models, and also facilitates our study of intra-brand cannibalization within a firm and inter-brand competition between firms within a market.

Table A.1: Model Fit Comparison

First Level	Second Level	Third Level	Sum of Squared Errors

Form Factor + Channel	Vendor		369.6405
Form Factor	Channel + Vendor		380.9223
Form Factor	Vendor	Channel	401.6364
Form Factor	Channel	Vendor	424.0485
Channel	Form Factor + vendor		641.0817
Channel	Form Factor	Vendor	784.9967
Vendor	Form Factor	Channel	2702.4180
Vendor	Form Factor + Channel		2744.0263
Form Factor + Vendor	Channel		3011.6502
Channel	Vendor	Form Factor	3934.1357
Channel + Vendor	Form Factor		6133.2251
Vendor	Channel	Form Factor	15673.0967

Appendix B. GMM Estimator (Chapter Three)

GMM was developed by Hansen (1982), who proved the consistency and asymptotic normality of a GMM estimator under general conditions. Unlike a maximum likelihood estimator which requires complete knowledge of the data distribution, GMM only relies on a set of moment conditions E[g(x,θ)]=0 where x represents the data and θ represents the model parameters. In order to estimate θ, one can replace E[g(x,θ)] by its empirical analog

$$\hat{g}(x,\theta) = \frac{1}{n}\sum_{i=1}^{n} g(x_i,\theta)$$

and then minimize the objective function:

$$\hat{\theta}_{\mathrm{GMM}} = \underset{\theta}{argmin}\left\{\hat{g}(x,\theta)'\widehat{W}\hat{g}(x,\theta)\right\}.$$

Here, W can be any positive-definite weighting matrix. With right choice of the weighting matrix a GMM estimator is asymptotically efficient.

For this paper, what is relevant to us is that GMM provides a useful tool to address the endogeneity in non-linear models using instruments. It becomes two-stage least squares (2SLS) when applied to linear models. Note that the identifying assumption of 2SLS is the orthogonality between the error term (ξ) and instruments (z). The same orthogonality condition becomes the moment condition in GMM for non-linear models, i.e., $g(x,\theta)=\xi(\theta)\cdot z$. By minimizing the GMM objective function, we can obtain the estimates for all model parameters, including ρ_{lm} and σ_{jlm}. Specifically, in our model, let $\xi(\theta)$ be the vector of all error terms (ξ_{ijlm}) as a function of model parameters θ. Let Z be a matrix that contains all observations of the instrumental and exogenous variables. Our GMM estimator can be written as:

$$\hat{\theta}_{\mathrm{GMM}} = \underset{\theta}{argmin}\{\xi'(\theta)Z\widehat{W}Z'\xi(\theta)\}.$$

We obtain an optimal weighting matrix by following the two-step procedure proposed by Hayashi (2000).