EXAMINING THE INFLUENCE OF ONLINE USER GENERATED CONTENT ON CONSUMER CHOICE: VARIETY SEEKING, LEARNING, AND SHARING

WANG QINGLIANG

(B. Eng, Renmin University of China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF INFORMATION SYSTEMS

SCHOOL OF COMPUTING

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

王清亮

Wang Qingliang

14 November 2013

ACKNOWLEDGEMENTS

I would like to express my heartfelt appreciation to many people for their guidance, support, and encouragement throughout the dissertation process.

First and foremost, I am particularly grateful to my dissertation advisor, Professor Goh Khim Yong, who patiently provided the vision, encouragement and advice necessary for me to proceed through the doctoral program and complete my dissertation. He has been a strong and supportive adviser to me throughout my graduate school career. Without his guidance and persistent help this dissertation would not have been possible.

Professor Chu Junhong and Huang Ke-Wei have served as my thesis committee advisors. They have devoted much time and effort to helping me to improve the quality of this research. I am grateful to them for their insightful comments and constructive suggestions. I also wish to express my gratitude to Professor Teo Hock Hai, Professor Chen Yuanyuan, and Professor Phan Tuan Quang, Professor KIM Seung Hyun, and all the rest of the faculty members in the Department of Information Systems for providing me good advice and warm encouragement to upgrade the quality of my research.

Special thanks to all professors and Ph.D. students in the Economics track in IS Department. I have greatly benefited from the discussion and communication in the brown-bag seminar.

I am indebted to all my friends for their enormous support and help. My heartfelt appreciation goes to Cai Shaojiang, Feng Yuanyue, Cheng Yihong, Luo

Cheng, Chen Qing, Cui Tingru, Liu Na, Tong Yu, Koh Ruilin Elizabeth, Zhang Xing, and Jiang Zhiying.

Last, but not least, I thank my parents for supporting my academic aspiration and their unconditional love. Their love gives me inspiration and is the driving force in this dissertation process. I owe them everything and wish I could show them just how much I love and appreciate them. Without them, I would not be where I am now. Any success that I achieve will be theirs as well.

TABLE OF CONTENTS

DECLARATION i

ACKNOWLEDGEMENTS ii

SUMMARY vi

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER 1. INTRODUCTION 1

1.1. Background and Motivation 1
1.2. UGC Literature 2
1.3. Research Overview and Contributions 5

CHAPTER 2. STUDY ONE 12

2.1. Introduction 12
2.2. Literature Review 17
2.2.1. Online UGC, Reviews and Word-of-mouth 17
2.2.2. Variety Seeking Behavior 19
2.3. Research Hypotheses 22
2.3.1. Information Effect: Consumer Awareness 22
2.3.2. Experience Effect: Consumer Prior Experiences 23
2.3.3. Competition Effect 26
2.4. Data Descriptions 28
2.4.1. Research Context and Data 28
2.4.2. Consumer Search and Consideration Set 30
2.4.3. Consumer Trip Level Panel Data 31
2.4.4. Consumer Search and Variety Seeking 32
2.5. Econometric Model Specification 35
2.5.1. First-Stage Decision: Whether to Patronize a New Restaurant? 36
2.5.2. Second-Stage Decision: Which Restaurant to Patronize? 38
2.6. Model Estimation and Findings 41
2.6.1. Identification and Estimation Methods 41
2.6.2. Reduced-Form Model Analysis Results 43
2.6.3. Two-Stage Structural Model Analysis Results 45
2.6.4. Hypotheses Tests Results 49
2.7. Discussion and Conclusion 53

CHAPTER 3. STUDY TWO 62

3.1. Introduction 62
3.2. Literature Review 66
3.2.1. UGC and Firms' Marketplace Performance 66
3.2.2. Quality Learning and Consumer Choice 69
3.3. The Econometric Model 70
3.3.1. Research Scenario and Utility Specification 70
3.3.2. Quality Signals from Consumption Experience 72
3.3.3. Quality Signals from Online UGC 73
3.3.4. Bayesian Learning from Online UGC 75
3.3.5. Bayesian Learning from Consumption Experiences 77
3.3.6. Choice Probability 77

3.4. Data Description **79**
3.5. Estimation **83**
3.6. Results and Managerial Implications **85**
3.6.1. Reduced-Form Analysis 85
3.6.2. Structural Model Results 87
3.6.3. Counterfactual Simulation 91
3.7. Conclusion **94**
CHAPTER 4. STUDY THREE **97**
4.1. Introduction **97**
4.2. Related Literature **101**
4.3. Theoretical Background **104**
4.4. Research Hypotheses **106**
4.4.1. Basic Assumptions 106
4.4.2. Dyadic Level Timing Effect 107
4.4.3. Timing Effect on the Popularity of Online Contents 108
4.5. Research Context **110**
4.6. Econometric Model Specifications **111**
4.7. Data Description **112**
4.8. Estimation Results **116**
4.9. Discussion and Conclusion **120**
CHAPTER 5. CONCLUSION **123**
BIBLIOGRAPHY **128**
APPENDICES **140**

SUMMARY

User-generated content (UGC), such as online product reviews, chat rooms, recommendation sites, and wikis, has grown rapidly on the Internet as a result of the pervasiveness of online Web 2.0 technologies. Consumers share their purchase and consumption experiences of a wide assortment of products through online information channels. This large-scale sharing of consumption experiences can help inexperienced consumers learn about new products and identify the products that best match their idiosyncratic preferences. Online UGC supplements the traditional information channels and has become a pivotal source of product quality information for consumers. The influence of UGC has attracted the attention of both practitioners and researchers alike. Since UGC has the potential to attract consumer visits, increase the time spent on the site, and create a sense of community among frequent shoppers, an increasing number of firms began offering UGC services, such as Amazon, Yelp, Dianping, and Epinions. These firms provide millions of reviews of diverse products and services on their websites and attract large number of visits every day.

This dissertation has two main objectives. First, we examine how online UGC influences individual consumers' purchase decisions. We are particularly interested in the impact of online UGC on individual consumers' new product exploration and product quality learning. Second, we examine the factors that affect the diffusion of UGC on social media platforms (SMP). We specifically investigate the timing effect of UGC diffusion on SMP by adopting a temporal networks modeling approach. Applying a wide variety of theories and techniques

drawing from economics, marketing, information systems, and psychology, we propose and empirically validate the mechanisms through which UGC diffuses on SMP and influences individual consumers' new product exploration and learning behavior.

The dissertation consists of three studies. Study One investigates the underlying process of how individual consumers perceive and use online UGC information to guide their new product exploration and purchase decisions. We propose that online UGC influences an individual consumer's new product exploration and purchase by (1) informing consumers of more choice alternatives in a market (information effect); (2) highlighting new choice alternatives that have a higher expected utility than that of their prior choices (experience effect); and (3) signaling the quality of competing choice alternatives (competition effect). Using a unique data set that consists of online reviews of restaurants on a popular consumer review website, consumers' information search and clickstream records on the same website, and consumers' actual patronage data on restaurant dining transactions, we specify and estimate a structural discrete choice model to empirically evaluate the influence of online UGC on individual consumers' decisions with respect to visiting restaurants. Our model assumes that consumers follow a two-stage decision process. In the first stage, consumers decide whether to explore a new restaurant. In the second stage, consumers decide which specific restaurant to patronize. Our model estimation approach accounts for observed and unobserved consumer heterogeneity, as well as for the potential endogeneity of consumer search. Our results show that consumers are more likely to sample a new restaurant after being exposed to more UGC of previously unvisited new restaurants. Furthermore, they are also more likely to do so when online UGC of restaurants highlights new alternatives with a higher expected utility than that of previously patronized restaurants. Consumers are also more price sensitive and assign more positive weight to UGC volume when they explore new product alternatives.

Study Two examines how consumers' experiential learning moderates the informational role of online UGC on an individual consumer's purchase decision regarding frequently purchased products. We propose a structural model to capture consumer learning from both online UGC and consumption experiences. Adopting the Bayesian updating framework, we demonstrate how individual consumers perceive and interpret the information embedded in online UGC to update their quality perceptions of products. Our model assumes that consumers learn both the average product quality and the precision of UGC signals. We apply our model to the context of consumer dining choice by combining data from online reviews of restaurants and consumers' restaurant dining records. Our analysis leads to two important findings. First, consumers are able to learn about restaurant quality from both online UGC and their own consumption experiences regarding dining choice. There is a significant amount of consumer learning from the consumers' own consumption experiences, indeed, much more than from online UGC. Second, neglecting consumers' experiential learning can result in over-estimation of the impact of online UGC on consumers' restaurant choice. We demonstrate how our model can be used for firms' decisions on word-of-mouth marketing. Our policy simulation results suggest that the impact of online UGC on consumer decisions decreases with the number of consumers' consumption trips. Thus, online UGC promotions may be influential only for new products and it is possible that the impact would be of short duration.

Study Three examines the factors which affect the diffusion of UGC on SMP. Users' attention is generally allocated in a rather unbalanced manner on SMP. An important question for both researchers and practitioners is as follows: how and why does the popular online content become popular? Previous studies have investigated this question from a variety of perspectives. In this study, we propose that the time when the content is generated has a significant impact on its popularity. We investigate this timing effect of information diffusion on SMP by adopting a temporal networks modeling approach. Our research hypotheses focus on examining how users' active time periods may affect the spread of information at the dyadic level and how the temporal order of information diffusion may affect the popularity and velocity of transmission of online content at the global level. Using data from a popular micro-blog website, we find strong evidence that the timing of when a piece of online content is posted has a significant effect on the popularity of the content.

LIST OF TABLES

Table 2-1: Summary Statistics of Restaurant Reviews.. 32
Table 2-2: Summary Statistics of Consumer Search 32
Table 2-3: Consumer Search and Variety Seeking Statistics............................... 35
Table 2-4: Description of Model Variables 37
Table 2-5: Model Estimation Results from Reduced-Form Analysis................... 44
Table 2-6: Model Estimation Results from One-Stage Choice Model................. 47
Table 2-7: Model Estimation Results from Two-Stage Choice Models............... 48
Table 3-1: Consumer Switching Frequency 80
Table 3-2: Consumer Switching Patterns 80
Table 3-3: Summary Statistics for Restaurant Reviews 81
Table 3-4: Summary Statistic for Restaurant Reviews.. 86
Table 3-5: Multinomial Logit Model Results 88
Table 3-6: Estimation Results from Structural Learning Models......................... 89
Table 3-7: Simulation Results.. 94
Table 4-1: User Profile Information 115
Table 4-2: Descriptive Statistics for Model 1 117
Table 4-3: Estimation Results for Model 1.. 118
Table 4-4: Descriptive Statistics for Model 2 119
Table 4-5: OLS Results for Model 2.. 120

LIST OF FIGURES

Figure 3-1: Consumers’ Information Updating Process 71
Figure 3-2: The Geographical Locations of the Restaurants Included in Our Data 82
Figure 4-1: Example of Temporal Network.. 105
Figure 4-2: Aggregated User Activity Level in Half Hour of Day..................... 113
Figure 4-3: User Heterogeneity in Posting Activity and Active Times.............. 114
Figure 4-4: User Heterogeneity in Posting Activity and Active Times.............. 115
Figure 4-5: Network Structure of the Core Users (85 users).............................. 116

CHAPTER 1. INTRODUCTION

1.1. Background and Motivation

User-generated content (UGC), such as online product reviews, chat rooms, recommendation sites and wikis, has grown rapidly on the Internet as a result of the pervasiveness of online Web 2.0 technologies. Consumers share their purchase and consumption experiences of a wide assortment of products through product review websites, blogs, discussion forums, social shopping and social networking websites. According to Anderson (2006), this large-scale sharing of consumption experiences can help people learn about new products by bridging the chasm between unknown preferences and product awareness or needs. Chen and Xie (2008) also suggest that online reviews are helpful for consumers to identify the products that best match their idiosyncratic preferences. Surveys conducted by research companies provide evidence that online user reviews have become a pivotal source of product quality information for consumers (ChannelAdvisor 2010; ComScore 2007). It is expected that 155 million US Internet users will consume some form of UGC by 2013 (Verna 2009). As a new source of product information, online UGC supplements the traditional information channels (Chen and Xie 2008; Dellarocas 2003; Dellarocas 2006; Mayzlin 2006) and has greater influence on consumer choice than traditional marketing activities (Trusov et al. 2009).

From the firms' perspective, this large-scale sharing of consumption experiences is important for their market success because it has the potential to reduce consumers' uncertainty about the quality of a product or service before their purchase decisions and thus alleviate the information asymmetry between firms and consumers (Bass et al. 1972). This especially facilitates purchase decisions involving experience goods whose quality cannot be inspected before purchase. According to comScore (2007), 24 percent of Internet users seek for and read UGC prior to paying for a service that is delivered offline. The influence of UGC has attracted the attention of both practitioners and researchers. UGC has the potential to attract consumer visits, increase the time spent on a site, and create a sense of community among frequent shoppers (Kumar and Benbasat 2006). An increasing number of firms, including Amazon, Yelp, Dianping, and Epinions, are offering UGC services. These firms provide millions of UGC on diverse products and services on their websites and attract numerous visits daily. In order to effectively market with UGC on digital and social media, it is important for firms and marketers to discern how individual consumers use and respond to online UGC. For review websites and social media operators, an insight into how individual consumers view, perceive, and use online review information has crucial implications in terms of website design, information management strategies, and the use of information technologies as a means of extending reach and enhancing the richness of consumer reviews.

1.2. UGC Literature

With the increasing popularity of online UGC websites, a large body of empirical studies have documented a positive relationship between online UGC and firm performance such as product sales (Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Duan et al. 2008a; Duan et al. 2008b; Forman et al. 2008; Goh et al. 2013; Liu 2006; Lu et al. 2013; Moe and Trusov 2011b; Sonnier et al. 2011a; Sun 2012; Zhu and Zhang 2010) and stock returns (Luo 2009; McAlister et al. 2012; Tirunillai and Tellis 2012). These studies mainly focus on the numerical measures of online UGC, such as volume (i.e., the number of UGC or reviews), valence (i.e., average rating of UGC or reviews) and variance (i.e., variance of UGC ratings) in their analysis. The volume of online UGC indicates the popularity of a product. The motivation to consider the volume of online UGC is that when more consumers discuss online about a product, other consumers will be more likely to become aware of it (Dellarocas et al. 2007). The valence of online UGC or word-of-mouth carries important information about a product's quality and reflects the level of consumer satisfaction with it (Zhu and Zhang 2010). Variance of UGC is a measure that captures the heterogeneity in consumer opinions or the variability associated with the attributes of a product reviewed, and thus reflects the level of uncertainty on the product quality (Sun 2012).

Previous empirical studies generally suggested that the volume of online reviews was positively associated with product sales, but the relationship between the valence of reviews and product sales was mixed. For example, Chevalier and Mayzlin (2006) found that increases in the volume and valence of a book review can lead to an increase in book sales. However, Chen et al. (2004) found that the valence of online reviews was not related to product sales by using a similar data set from Amazon.com. Duan et al. (2008a) documented the importance of the number of online reviews in influencing movie box office sales, but the valence of online reviews was not influential. Liu (2006) found that while the volume of online word-of-mouth was positively associated with product sales, the relationship between the reviews' valence and sales was insignificant. Dellarocas et al. (2007) used a modified Bass diffusion model to study the role of online reviews in forecasting movie revenue and found that the valence of reviews was a better predictor than other metrics they considered. Chintagunta et al. (2010) found that the valence, but not the volume and variance, of reviews explained the opening day movie box office revenues. In addition, two follow-up studies examined the moderating effects of other factors on product sales, such as product and consumer characteristics (Zhu and Zhang 2010) and the matching of geographical locations between those of reviewers and consumers who read the reviews (Forman et al. 2008). At the individual consumer level, researchers also explored how online UGC affected individual consumers' choice decisions (Albuquerque et al. 2012; Goh et al. 2013; Zhao et al. 2013).

Online UGC have also been studied from other perspectives. For example, Godes and Mayzlin(2004) showed that the dispersion of conversations across online consumer communities was the main factor that influenced sales performance. Duan et al. (2008b) explored the positive feedback mechanism between word-of-mouth and retail sales. Mudambi and Schuff(2010) demonstrated that review extremity, review depth, and product type affected the helpfulness of online UGC. A few studies have however shown that online UGC could be biased due to several reasons, such as self-selection (Li and Hitt 2008), pricing effects (Li and Hitt 2010), and social dynamics (Godes and Silva 2012; Moe and Trusov 2011b; Wang et al. 2010).

Researchers also examined the roles of textual content in online UGC by adopting text mining techniques (Das and Chen 2007). These studies extended the

UGC literature by exploring the effect of text messages on firms' performance (Archak et al. 2011a; Ghose et al. 2012), which gave a more comprehensive view of online UGC's effects. They examined the influence of multiple sources of online communications (Gu et al. 2012) and the dynamics between online UGC and firms' market performance (McAlister et al. 2012; Sonnier et al. 2011a; Tirunillai and Tellis 2012). In addition, researchers combined text mining and other techniques such as semantic network analysis and graphic models to understand brand associative networks, monitor market structures, and extracted comparative relations between products from customer reviews (Decker and Trusov 2010; Netzer et al. 2012; Xu et al. 2011).

1.3. Research Overview and Contributions

This dissertation extends the literature as follows: (1) by investigating how individual consumers search, perceive and use online UGC information to explore new products; (2) by examining how consumers' experiential learning moderating the information role of online UGC; and (3) by examining the timing factors that affect the diffusion of UGC on social media platforms (SMP). Specifically, the dissertation has the following unique contributions to the literature on the individual consumer level impact of UGC and UGC diffusion. First, due to a lack of relevant data, previous studies has implicitly assumed that consumers search and browse online UGC related to a product or service before committing to their purchase decisions (Chevalier and Mayzlin 2006; Gu et al. 2012; Zhao et al. 2013). However, if one cannot ascertain that consumers *actually* search and browse online UGC before a purchase, the prior documented influence of online UGC on product sales may have been spurious, biased or non-causal in nature. In this study, we take advantage of individual consumers' information search records to explicitly examine how online UGC takes effect in the purchase decision process of individual consumers. Our uniquely rich data set in this research not only helps us to inspect the relationship between online UGC search and choice behaviours, it also provides crucial sources of identification for the causal effect of online UGC on individual purchases. We illustrate how researchers can make use of consumers' search data to explicitly model consumers' decision process in the light of online UGC usage.

Second, unlike previous studies which purely investigated the impact of online UGC on consumers' purchase outcomes in terms of discrete choice (Zhao et al. 2013) and expenditure (Albuquerque et al. 2012; Goh et al. 2013), this study goes beyond to examine how online UGC influences an individual consumer's variety seeking choice behaviour[1], or more specifically, the tendency to sample new products. Previous studies have pointed out that understanding variety seeking behaviours has important managerial implications in terms of product assortment, competitive positioning and pricing strategies (Kahn 1995; Sajeesh and Raju 2010; Seetharaman and Che 2009). Surprisingly, there is a lack of research in both Marketing and Information System literature to investigate how online UGC influences individual consumers' new product exploration and variety seeking behaviour. We posit that online UGC, by highlighting the variety of alternatives available in a market, can increase consumers' awareness and willingness to sample a product that they have not tried before.

[1]Generally, variety seeking refers to a phenomenon that consumers engage in varied behaviours, such as brand switching or multi-brand buying (McAlister and Pessemier 1982). In this paper, we focus on a specific aspect of variety seeking in terms of new product exploration (Kahn 1995).

Third, although previous studies suggests that online UGC can provide product quality information for consumers (Zhao et al. 2013), none of them have examined how the informational role of online UGC may change when consumers are able to learn product quality from their own consumption experiences. For frequently purchased products, there is informational value for the consumer to purchase a product because that consumer can repeatedly buy the product if he or she likes it (McFadden and Train 1996). Marketing researchers have suggested that consumer experiential learning is of great importance in the consumer choice process (Erdem and Keane 1996). As consumers can learn product quality from their own consumption experiences, the informational role of online UGC will decrease when consumers gain more experiences. Thus, an understanding on how consumers' experiential learning moderates the effect of UGC on consumer choice is of great importance for marketers in evaluating the impact of online UGC and thus provides useful guidance for firms when they run marketing campaigns on these new social media platforms.

Fourth, the dissertation contributes to the UGC diffusion literature by examining the timing effect of UGC diffusion on SMP. Traditional diffusion models conceptualize the diffusion process as being determined by the effects of innovation and imitation and ignore connection patterns between individuals (Bass 1969; Mahajan et al. 1990). Recent studies explicitly incorporate the interpersonal connections when examining word of mouth diffusion processes (Iyengar et al. 2011). Researchers show evidence of social contagion (or peer effects) in diverse contexts (Bandiera and Rasul 2006; Hill et al. 2006; Iyengar et al. 2011; Katona et al. 2011). Prior research also examines the role of local and global network structure of opinion leaders (Iyengar et al. 2011; Katona et al. 2011; Moynihan 2008; Nair et al. 2010; Yoganarasimhan 2012) and the characteristics of information content (Berger and Milkman 2012; Berger and Schwartz 2011; Berger et al. 2010; Zhang and Moe 2012) in the diffusion process. Surprisingly, these studies usually assume the information network is static and neglect the impact of human activity patterns across time in the information diffusion process (Iribarren and Moro 2009).

We next present three studies which investigate the impact of online UGC at the individual consumer level. Study One investigates the underlying process of how individual consumers perceive and use online UGC information to guide their new product exploration and purchase decisions. We propose that online UGC influences an individual consumer's new product exploration and purchase by (1) informing consumers of more choice alternatives in a market (information effect), (2) highlighting new choice alternatives that have a higher expected utility than that of their prior choices (experience effect), and (3) signalling the quality of competing choice alternatives (competition effect). Using a unique data set that consists of online reviews of restaurants on a popular consumer review website in China, consumers' information search and click stream records on the same website, and consumers' actual patronage data on restaurant dining transactions, we specify and estimate a two-stage structural discrete choice model to empirically evaluate the influence of online UGC on individual consumers' purchase decision in visiting restaurants. Our model assumes that consumers follow a two-stage hierarchical decision process. In the first stage, consumers decide whether to choose from a set of new (i.e., previously unvisited) restaurants or from a set of restaurants patronized before, conditional on their expectations about the utility they can get from each set of choices. In the second stage, conditional on their first stage decisions, consumers decide which specific restaurant to patronize. We specify random coefficients for our model parameters to capture consumers' heterogeneous responses to online UGC. Our model estimation approach accounts for observed and unobserved consumer heterogeneity, as well as for the potential endogeneity of consumer search using a control function approach (Petrin and Train 2010). Our findings show that consumers are more likely to sample a new restaurant (1) after being exposed to more online UGC of previously unvisited new restaurants, and (2) when online UGC of restaurants highlight new choice alternatives with a higher expected utility than that of consumers' previously patronized restaurants. Results show that information attributes from online UGC have significant influences on consumer choice among competing products. However, consumers are heterogeneous in terms of responses to UGC. We also find evidence that online UGC is more influential when consumers search for information to explore a new product. Specifically, consumers are more price sensitive and assign more positive weight on the volume of UGC when they explore new product alternatives.

Study Two examines how consumers' experiential learning moderates the informational role of online UGC on an individual consumer's purchase decision regarding frequently purchased products. We propose a structural model to capture consumer learning from both online UGC and consumption experiences. Adopting the Bayesian updating framework, we demonstrate how individual consumers perceive and interpret the information embedded in online UGC to update their quality perceptions of products. Our model assumes that consumers learn both the average product quality and the precision of UGC signals. We apply our model to the context of consumer dining choice by combining data from online reviews of restaurants and consumers' restaurant dining records. Our analysis leads to two important findings. First, consumers are able to learn about restaurant quality from both online UGC and their own consumption experiences regarding dining choice. There is a significant amount of consumer learning from the consumers' own consumption experiences, much more than from online UGC. Second, neglecting consumers' experiential learning can result in over-estimation of the impact of online UGC on consumers' restaurant choice. We demonstrate how our model can be used for firms' decisions on word-of-mouth marketing. Our policy simulation results suggest that the impact of online UGC on consumer decisions decreases with the number of consumers' consumption trips. Thus, online UGC promotions may be influential only for new products and it is possible that the impact would be of short duration.

Study Three examines the factors which affect the information diffusion on social media platforms (SMP). Users' attention is generally allocated in a rather inequitable manner on SMP. An important question for both researchers and practitioners is: how and why do the online popular contents get popular?

Previous studies have investigated this question from diverse perspectives. In this study, we propose that the time when the content is generated has a significant impact on its popularity. We investigate this timing effect of information diffusion on SMP by adopting a temporal networks modelling approach. Our research hypotheses focus on how users' active time periods may affect the spread of information at the dyadic level and how the temporal order of information diffusion may affect the popularity of online content at the global level. Using data from a popular micro-blog website, we find strong evidence that the timing when a piece of online content is posted has a significant effect on the popularity of the content.

CHAPTER 2. STUDY ONE

HOW DOES USER-GENERATED CONTENT INFLUENCE CONSUMERS' NEW PRODUCT EXPLORATION? AN EMPIRICAL ANALYSIS OF CONSUMER SEARCH AND CHOICE BEHAVIOR

2.1. Introduction

User-generated content (UGC), such as online product reviews, has grown rapidly on the Internet with the pervasiveness of online Web 2.0 technologies. Consumers share their purchase and consumption experiences of a wide assortment of products through product review websites, blogs, discussion forums, social shopping and social networking websites. According to Anderson (2006), these large-scale sharing of consumption experiences can help people learn about new products by bridging the chasm between unknown preferences and product awareness or needs. Chen and Xie (2008) also suggest online reviews are helpful for consumers to identify the products that best match their idiosyncratic preferences. Surveys conducted by research companies provide evidence that online user reviews have become a pivotal source of product quality information to consumers (ChannelAdvisor 2010; ComScore 2007).

In this paper, we examine the impact of online UGC on individual consumers' new product exploration in terms of both online search and variety seeking choice behaviors. A large number of empirical studies have investigated the influence of online UGC on aggregate product sales in different product categories (Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Clemons et al. 2006; Duan et al. 2008a; Forman et al. 2008; Godes and Mayzlin 2004; Gu et al. 2012; Liu 2006; Moe and Schweidel 2012; Sun 2012; Tirunillai and Tellis 2012;

Zhu and Zhang 2010). However, only a handful of studies explored how online UGC affects individual consumers' choice decision (Albuquerque et al. 2012; Goh et al. 2013; Zhao et al. 2013). In order to effectively market with UGC on digital and social media, it is important for marketers to know how individual consumers use and respond to online UGC. For review website and social media operators, the insight of how individual consumers view, perceive, and use online review information has crucial implications in terms of website design, information management strategies, and the use of information technologies as a means to extend reach and enhance richness of consumer reviews.

This study has three unique contributions to the literature on the individual level impact of UGC on consumers. First, due to a lack of relevant data, previous studies has implicitly assumed that consumers search and browse online UGC related to a product or service before committing to their purchase decisions (Chevalier and Mayzlin 2006; Gu et al. 2012; Zhao et al. 2013). However, if one cannot ascertain that consumers *actually* search and browse online UGC before a purchase, the prior documented influence of online UGC on product sales may have been spurious, biased or non-causal in nature. In this study, we take advantage of individual consumers' information search records to explicitly examine how online UGC takes effect in the purchase decision process of individual consumers. Our uniquely rich data set in this research not only helps us to inspect the relationship between online UGC search and choice behaviors, it also provides crucial sources of identification for the causal effect of online UGC on individual purchases. We illustrate how researchers can make use of consumers' search data to explicitly model consumers' decision process in the light of online UGC usage.

Second, unlike previous studies which purely investigated the impact of online UGC on consumers' purchase outcomes in terms of discrete choice (Zhao et al. 2013) and expenditure (Albuquerque et al. 2012; Goh et al. 2013), this study goes beyond to examine how online UGC influences an individual consumer's variety seeking choice behavior[2], or more specifically, the tendency to sample new products. Previous studies have pointed out that understanding variety seeking behaviors has important managerial implications in terms of product assortment, competitive positioning and pricing strategies (Kahn 1995; Sajeesh and Raju 2010; Seetharaman and Che 2009). In our context, we posit that online UGC, by highlighting the variety of alternatives available in a market, can increase consumers' awareness and willingness to sample a product that they have not tried before. Positive reviews or UGC that recommend a product, can also reduce psychological switching costs and thus encourage switching from consumers' prior choice to a recommended one (Li et al. 2011).

Third, we propose a structural two-stage discrete choice model based on a random utility framework (Ben-Akiva and Lerman 1985; Chintagunta 1999) that extends the UGC literature by demonstrating the role of UGC in consumers' hierarchical choice process of new product exploration and variety seeking behaviors. Specifically in this paper, we propose that online UGC plays different roles at different stages of the consumer choice process.

[2] Generally, variety seeking refers to a phenomenon that consumers engage in varied behaviors, such as brand switching or multi-brand buying (McAlister and Pessemier 1982). In this paper, we focus on a specific aspect of variety seeking in terms of new product exploration (Kahn 1995).

With these contributions in mind, we propose the following three research questions:

(1) How does consumers' online UGC search influence new product exploration behaviors?

(2) How do consumers' prior product consumption experiences affect their search or usage of online UGC to explore new products?

(3) To what extent does competition across online UGC of competing alternatives influence individual consumers' purchase decision, especially when they explore new products?

Adopting the insights from the variety seeking literature (Kahn 1995; McAlister and Pessemier 1982), Prospect theory (Kahneman and Tversky 1979) and the brand competition literature (Laroche et al. 1994), we propose that online UGC influences an individual consumer's new product exploration and purchase by (1) informing consumers of more choice alternatives in a market (*information effect*), (2) highlighting new choice alternatives that have a higher expected utility than that of their prior choices (*experience effect*), and (3) signaling the quality of competing choice alternatives (*competition effect*).

Using a unique data set that consists of online reviews of restaurants on a popular consumer review website in China, consumers' information search and clickstream records on the same website, and consumers' actual patronage data on restaurant dining transactions, we specify and estimate a two-stage structural discrete choice model to empirically evaluate the influence of online UGC on individual consumers' purchase decision in visiting restaurants. Our model assumes that consumers follow a two-stage hierarchical decision process. In the first stage, consumers decide whether to choose from a set of new (i.e., previously unvisited) restaurants or from a set of restaurants patronized before, conditional on their expectations about the utility they can get from each set of choices. In the second stage, conditional on their first stage decisions, consumers decide which specific restaurant to patronize. We specify random coefficients for our model parameters to capture consumers' heterogeneous responses to online UGC. Our model estimation approach accounts for observed and unobserved consumer heterogeneity, as well as for the potential endogeneity of consumer search using a control function approach (Petrin and Train 2010).

Our findings show that consumers are more likely to sample a new restaurant (1) after being exposed to more online UGC of previously unvisited new restaurants, and (2) when online UGC of restaurants highlight new choice alternatives with a higher expected utility than that of consumers' previously patronized restaurants. Results show that information attributes from online UGC have significant influences on consumer choice among competing products. However, consumers are heterogeneous in terms of responses to UGC. We also find evidence that online UGC is more influential when consumers search for information to explore a new product. Specifically, consumers are more price sensitive and assign more positive weight on the volume of UGC when they explore new product alternatives. These findings relating consumer new product exploration behaviors in the context of user-generated reviews provide new insights into how individual consumers perceive and use UGC information for consumption decisions, and have important implications for academic research and practice in the information systems and marketing fields.

2.2. Literature Review

2.2.1. Online UGC, Reviews and Word-of-mouth

Our current study is related to the literature that examines the impact of UGC on firm performance. With the increasing popularity of online UGC websites, a large body of empirical studies have documented a positive relationship between online UGC and firm performance such as product sales (Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Duan et al. 2008a; Duan et al. 2008b; Forman et al. 2008; Goh et al. 2013; Liu 2006; Lu et al. 2013; Moe and Trusov 2011b; Sonnier et al. 2011a; Sun 2012; Zhu and Zhang 2010) and stock returns (Luo 2009; McAlister et al. 2012; Tirunillai and Tellis 2012). These studies mainly focus on the numerical measures of online UGC, such as volume (i.e., the number of UGC or reviews), valence (i.e., average rating of UGC or reviews) and variance (i.e., variance of UGC ratings) in their analysis. The volume of online UGC indicates the popularity of a product. The motivation to consider the volume of online UGC is that when more consumers discuss online about a product, other consumers will be more likely to become aware of it (Dellarocas et al. 2007). The valence of online UGC or word-of-mouth carries important information about a product's quality and reflects the level of consumer satisfaction with it (Zhu and Zhang 2010). Variance of UGC is a measure that captures the heterogeneity in consumer opinions or the variability associated with the attributes reviewed of a product, and thus reflects the level of uncertainty on the product quality (Sun 2012).

Online UGC have also been studied from other perspectives. For example, Godes and Mayzlin (2004) showed that the dispersion of conversations across online consumer communities is the main factor that influences sales performance. Duan et al. (2008b) explored the positive feedback mechanism between word-of-mouth and retail sales. Mudambi and Schuff (2010) demonstrated that review extremity, review depth, and product type affect the helpfulness of online UGC. A few studies have however shown that online UGC can be biased due to several reasons, such as self-selection (Li and Hitt 2008), pricing effects (Li and Hitt 2010), and social dynamics (Godes and Silva 2012; Moe and Trusov 2011b; Wang et al. 2010). Researchers also examined the roles of textual content in online UGC by adopting text mining techniques (Das and Chen 2007). These studies extended the UGC literature by exploring the effect of text messages on firms' performance (Archak et al. 2011a; Ghose et al. 2012), which give a more comprehensive view of online UGC's effects. They examined the influence of multiple sources of online communications (Gu et al. 2012) and the dynamics between online UGC and firms' market performance (McAlister et al. 2012; Sonnier et al. 2011a; Tirunillai and Tellis 2012). In addition, researchers combined text mining and other techniques such as semantic network analysis and graphic models to understand brand associative networks, monitor market structures, and extract comparative relations between products from customer reviews (Decker and Trusov 2010; Netzer et al. 2012; Xu et al. 2011).

To the best of our knowledge, only a handful of empirical studies have explored the impact of online UGC on individual consumers' purchase decisions

(Albuquerque et al. 2012; Goh et al. 2013; Zhao et al. 2013). Albuquerque et al. (2012) developed a modeling approach that explains individual level choices of visiting a UGC platform, creating and purchasing content as a function of consumer characteristics and marketing activities. Goh et al. (2013) examined the relative impact of social media brand community contents from consumers and marketers on consumers' apparel purchase expenditures. Zhao et al. (2013) proposed a structural learning model to study the effect of online reviews on consumer purchases of an experiential product (i.e., books). However, all these prior-mentioned studies lacked individual consumer-specific UGC site visitation and search behavior data, and thus do not focus on how online UGC search influences new product exploration choice behaviors. To examine the impact of online UGC on consumer choice, a critical concern is that consumers may not have searched for information from online UGC prior to their purchases. If such is the case, the documented positive influence of online UGC on product sales may have been spurious (e.g., due to other unobserved influences), biased (e.g., consumers who browsed UGC related to a product may not have been the ones who bought the product) and non-causal (e.g., due to reverse causation or simultaneity) in nature.

2.2.2. Variety Seeking Behavior

Our paper is related to the variety seeking literature. Variety seeking has been defined and modeled from different perspectives by psychologists, consumer behaviorists, marketers, and economists (Givon 1984; Kahn 1995; McAlister and Pessemier 1982; Sajeesh and Raju 2010; Seetharaman and Che 2009; Simonson

1990). Generally, variety seeking refers to a phenomenon that consumers engage in varied behaviors, such as brand switching or multi-brand buying (McAlister and Pessemier 1982). In this study, we define and contextualize variety seeking as consumers' tendency to try or sample a new product that they have not purchased before. Variety seeking is pervasive because of the tendency of individuals to seek diversity in the choice of search and experience goods, or commodity and differentiated products or services in their daily life (Givon 1984). Variety seeking may arise over time, such as when consumers go to different restaurants from one dining occasion to the next or choose diverse places to take a vacation. Consumers can also seek variety within purchase occasions by choosing a portfolio of products from different firms or brands at one time (Simonson 1990). In sum however, we note that in the information systems literature, there has been surprisingly no research that examines variety seeking behavior as an outcome on part of consumers. While there have been some studies that examine the effect of online UGC on loyalty of consumers to stores or brands (Gauri et al. 2008), it is critical to note that in the marketing literature, consumers' variety seeking is not the flipside of loyalty, but is considered to potentially co-exist with inertia (habit persistence) within individuals (Bawa 1990; Roy et al. 1996).

As to why and when consumers seek variety, the literature has identified three main motivating factors of consumers' variety seeking behaviors: satiation, external situations, and preference uncertainty (Harlam and Lodish 1995; Kahn 1995; McAlister and Pessemier 1982; Simonson 1990). Satiation means that consumers seek variety because of their internal or personal desire for variety. For example, a consumer who drinks milk every morning may get weary of its taste after some time, and therefore switches to apple juice. Researchers modeled such variety seeking by assuming that consumers can derive utility from the change itself, irrespective of the brands he or she switched to or from (Givon 1984; Sajeesh and Raju 2010; Seetharaman and Che 2009). An alternative approach to measure such variety seeking assumed that consumers become satiated after exposure to some attributes and seek alternatives that offer some other attributes (McAlister 1982). Variety seeking triggered by an external situation refers to the scenario when consumers seek variety due to external constraints (McAlister and Pessemier 1982), such as multiple needs, multiple situations, and multiple uses, rather than an immediate internally derived need for variety. For example, consumers may seek to try out different restaurants across different occasions because of the multiple preferences of family members. Previous studies investigating such external situations in variety seeking have also explored the effect of price promotions (Kahn and Louie 1990; Kahn and Raju 1991) and retail environment (Menon and Kahn 1995). Lastly, variety seeking due to consumers' preference uncertainty typically implies that consumers seek variety so that they can have a portfolio of options as a hedge against future uncertainty or as a means to protect their continued interest in their favorite options (Harlam and Lodish 1995).

In this study, we focus on the following aspects of consumer variety seeking behavior. First, we investigate consumers' tendency to sample a new product that they have not purchased before, such that we can focus our efforts on examining how online UGC affects consumers' new product exploration. Second, our research context focuses on discrete choice situations where consumers choose one unit of a product (from a choice set of competing substitutes) at each purchase occasion but seek variety across purchase occasions over time. Third, unlike most existing literature, which treat variety seeking as an independent variable in their analysis, we model a consumer's new product exploration behavior as a decision variable and explicitly investigate how it can be influenced by online UGC, as an external triggering situation or stimuli (Anderson 2006; Chen and Xie 2008; Dellarocas 2003; Mayzlin 2006).

2.3. Research Hypotheses

Consumers often need to make purchase decisions under uncertainty because they usually lack information about product quality, seller reputation or other available product alternatives. To examine the influence of online UGC on individual consumers' new product exploration behaviors, we need to understand what kinds of information online UGC can provide for consumers and how consumers perceive and use information embedded in online UGC. In the following paragraphs, we focus on developing hypotheses based on the context of UGC's influence on consumers' variety seeking for frequently or repeat purchase experience goods (as opposed to one-time purchase goods such as books).

2.3.1. Information Effect: Consumer Awareness

Economics and marketing researchers have emphasized the crucial role of consumer information search on consumer choice behavior for years (Mehta et al. 2003; Nelson 1970; Stigler 1961). Researchers argue that online UGC are helpful for consumers to identify the products that best match their idiosyncratic preferences (Anderson 2006; Chen and Xie 2008). According to recent market surveys (ChannelAdvisor 2010; ComScore 2007), online UGC has become a pivotal source of product information to consumers. We argue that an information role of online UGC is to suggest or highlight other choice alternatives which consumers are not previously aware of (Anderson 2006; Chen and Xie 2008; Nelson 1970). This role is especially important when the products are highly differentiated or the market is highly competitive such that there are a large number of choice alternatives. In such cases, consumers may continually make purchase choices among products which they are already aware of (Nelson 1970). When other new choice alternatives are highlighted to them, consumers have the incentive to try these new products because satiated consumers of a repeatedly chosen good (especially a hedonic one) can derive utility from the change itself, irrespective of the alternative he or she switches to or from (Givon 1984; Seetharaman and Che 2009). Thus, we hypothesize that an individual consumer's higher extent of search of online UGC on new products will lead to a higher probability of new product exploration, i.e., variety seeking.

H1: A consumer is more likely to choose a new product when he or she searches more new product alternatives from online UGC.

2.3.2. Experience Effect: Consumer Prior Experiences

In addition to providing consumers with new choice alternatives, online UGC can provide detailed product quality information such that consumers are able to evaluate and compare different product alternatives (Chan et al. 2012b; Zhao et al. 2013). This role is especially important for experience goods which consumers cannot inspect product quality before their purchase or consumption. In a highly differentiated and frequently purchased product market, consumers usually have their own specific choice sets. Consumers may generally be loyal to a small set of choice alternatives and seek variety by switching from one product to another (Bawa 1990; Kahn et al. 1986). By searching online UGC, consumers can identify and more importantly, evaluate the expected utility or value of these new choice alternatives. Here, we examine how consumers' prior consumption experiences, i.e., the products consumers have purchased and consumed before, in relation to relative quality levels of new alternatives highlighted in online UGC, can influence the likelihood of new product exploration.

Prospect theory provides a relevant foundation to address how individual consumers' prior consumption experiences influence their new product exploration. According to Prospect theory, consumers have reference-dependent preferences and consumers are averse to losses (Kahneman and Tversky 1979). Reference dependence implies that a consumer's current consumption choice depends on his or her reference point. Loss aversion implies that the consumer is averse to negative departures from a reference point. We argue that consumers' prior consumption experiences establish consumers' referent points, i.e., the expectations about the maximum utility they can receive from their prior choice sets (KÖSzegi and Rabin 2006). Consumers may seek variety in consumption choices based on the benchmarks set by these reference points. Given that consumers are loss averse, especially relative to the referent benchmarks, they are thus more likely to try new alternatives which afford a higher level of utility (Kahneman and Tversky 1979; Loewenstein and Prelec 1993).

Online UGC provides the necessary quality information of new product alternatives and products which consumers are already familiar with. The uniform scaling of online UGC ratings makes the measurement and comparison of different products direct and convenient. Consumers can easily evaluate the relative attractiveness of products that they are unfamiliar with by comparing with the products they have purchased before. At the same time, when new alternatives highlighted by online UGC can provide higher expected utility, this generally signals a high benefit and low (psychological) cost of switching or seeking variety (Li et al. 2011). We call this the *experience effect* of online UGC. Thus, we hypothesize:

H2: A consumer is more likely to choose a new product when the new products he or she searches from online UGC can provide higher utility than that of prior choice alternatives.

Here, we emphasize that online UGC or reviews provide researchers an alternative method to measure consumers' experiences. Researchers usually measured consumers' consumption experiences by conducting surveys or interviews (Scott and Keiser 1984). Some empirical studies modeled consumers' prior experience as a time-invariant unobserved consumer heterogeneity component in fixed or random effects models or assumed that it follows a specific distribution (Erdem and Keane 1996). Online UGC provides a new information source which researchers can use to explicitly formulate variables that measure consumers' prior consumption experiences (Zhao et al. 2013).

2.3.3. *Competition Effect*

We name the last role of online UGC as the *competition effect*, which addresses the effect of online UGC of competing product alternatives on consumers' choice behavior. Previous research has found that consumers' judgments are relative in nature and they are affected by the context under which judgments are made (Laroche and Brisoux 1989; Laroche et al. 1994). In choosing amongst brands, consumers' choice of a certain brand is not only determined by the attributes of that brand, but also by the attributes of competing brands (Abe and Tanaka 1989; Kapoor and Heslop 2009; Laroche and Brisoux 1989; Laroche et al. 1994; Lynch et al. 1991). Applying the same logic to the influence of online UGC, we argue that consumers' judgment for one product is not only influenced by online UGC of the focal product but also by those of competing products in a choice set (Li et al. 2011).

Previous empirical studies of UGC generally found that the volume of online UGC is positively associated with product sales, but the relationship between the valence of reviews and sales is mixed. For example, Chevalier and Mayzlin (2006) found that increases in the volume and valence of book reviews can lead to an increase in book sales. However, Chen et al. (2004) found that the valence of online UGC is not related to sales by using a similar data set from Amazon.com. Duan et al. (2008a) documented the importance of the number of online UGC in influencing movie box office sales, but the valence of online UGC is not influential. Liu (2006) found that the volume but not the valence of online word-of-mouth is positively associated with product sales. Dellarocas et al. (2007) found that the valence of reviews is a better predictor of movie revenue than other metrics they considered. Chintagunta et al. (2010) found that the valence, but not the volume and variance, of reviews explains opening day movie box office revenues. Clemons et al. (2006) found that the variance of ratings and the strength of the most positive quartile of reviews have a significant impact on the sales growth of craft beers. Sun (2012) showed that a higher standard deviation of ratings on Amazon improves a book's relative sales rank when the average rating is lower. In summary, based on findings from these mostly aggregate level studies, we posit that at the individual consumer level, there is a significant influence of information attributes from online UGC on consumer choice (Huang et al. 2009; Zhao et al. 2013). In addition, we expect that consumer' responses to information from online UGC are heterogeneous such that some consumers are more sensitive to information from online UGC than others. Therefore, we hypothesize the following:

H3: Information attributes from online UGC have a significant influence on a consumer's choice decision among competing alternatives.

H4: Consumers show significant heterogeneity in terms of responses to information attributes from online UGC.

The literature on the economics of information suggests that consumers search for information prior to purchase in order to reduce their uncertainty about the decision (Nelson 1970; Stigler 1961). Greater uncertainty then should presumably lead to more extensive search (Punj and Staelin 1983; Urbany et al. 1989). Since consumers are likely to have more uncertainty on products which they are unfamiliar with or have not purchased before, we posit that online UGC will be more influential on consumers' purchase decision when they are considering the choice of a new product not sampled or purchased before.

H5: Online UGC has a more significant influence when a consumer is choosing from a set of new products, compared to when he or she is choosing from a set of products with prior purchase experiences.

2.4. Data Descriptions

2.4.1. Research Context and Data

We evaluate our research hypotheses in the context of consumers' restaurant dining choices. Such a context is appropriate for two reasons. First, the restaurant industry is typically quite competitive, especially within a major metropolitan city from which our data set is based on. Consumers usually seek varieties in restaurant dining choices in terms of food quality, restaurant ambience, service standard, cuisine type, and location. Therefore, it is a very apt context to investigate consumers' new product exploration behaviors. Second, restaurants and their associated cuisine offerings and services are essentially experience goods which consumers usually do not have full information about their quality before their first patronage. Online information channels, such as UGC and reviews, exert a substantial influence on consumers' choice of experience goods (Gu et al. 2012; Huang et al. 2009).

Our novel data set is compiled from four unique sources: online reviews of restaurants from a popular consumer reviews website in China, restaurant attributes information (e.g., location and promotions), consumers' web page browsing data on the website, and their restaurant dining transaction records. The overall timeline of our data set spans from 2003 to 2008. The restaurant reviews information spans from April 2003 to March 2008. Consumers' dining transaction records span from May 2005 to March 2008. Detailed information of restaurants' attributes is available from January 2006 to March 2008. Consumers' web page browsing data spans from January 2008 to March 2008.

We gathered the restaurant reviews data from Dianping.com (similar to Yelp.com), where consumers can share their experiences or reviews of restaurants and eateries in each major city. The restaurant reviews information includes consumers' ratings of restaurants in terms of overall quality, food taste, restaurant ambience, and service quality. The ratings scale ranges from 0 to 4, 0 being very bad and 4 being very good. We use the overall quality rating in our empirical analysis. In addition, reviewers can post information about the average price for each person, recommended dishes, and detailed qualitative comments for each restaurant. Restaurants are classified differentially by the review site in terms of geographical areas, price level, and cuisine type. Restaurants are located in 11 areas of the sampled city, divided into 5 different price levels, and categorized into offering 17 cuisine types. Other information on the restaurants includes that of promotions such as the availability of discount or promotional coupons and the promotional time period. Consumers' web browsing data includes the consumer's anonymized identity, accessed web page's URL, accessed restaurant's identifier, access date and time, as well as the IP address. Consumers' restaurant dining records were collated from information gathered using the review website's loyalty program member cards. The website distributes loyalty program member cards to their registered users. When consumers patronize a restaurant which has a joint partnership program with the review web site, they could get discounts and accumulate membership points by using the loyalty member card in each visit. Our consumer transactions data are thus sourced from restaurant customers who are also members of the review site's loyalty program. These dining transactions data include consumer's anonymized identity, restaurant's name, dining expenditure, and transaction date.

2.4.2. Consumer Search and Consideration Set

Consumers' web page browsing records play two crucially important roles in our study. First, this data can help us ascertain that consumers do search and use online UGC information before their purchases and thus provide a crucial source of identification for the parameters in our econometric model. If we cannot ascertain this, it will be hard for us to argue that consumers' purchase decisions are influenced by online UGC because information from UGC may merely reflect restaurants' quality at large. Second, this data can help us ascertain which specific web pages of restaurants consumers have searched and visited. This information can help us define a consumer's consideration set at each purchase occasion.

Researchers have pointed out the importance of consideration set in consumers' decision process because a consumer has limited information-processing abilities and thus he or she cannot make explicit utility comparisons across a large number of choice alternatives available on an online reviews site

(Andrews and Srinivasan 1995; Shocker et al. 1991). Prior studies have defined consideration set as a subset of alternatives that survive a screening process and receive serious consideration during the purchase occasion (Gilbride and Allenby 2004; Shocker et al. 1991). In our current study, both consumers' prior restaurant dining records and their UGC browsing data are used to define consumers' consideration sets. A consumer's consideration set at a specific trip consists of two groups of restaurants. One is a group of restaurants that the consumer has patronized before. These are the choice alternatives that a consumer has prior consumption experiences and thus can easily recall their quality. The other is a group of unvisited restaurants that consumers have searched before the transaction within a 7 days' time window. On Dianping.com, consumers can search and filter restaurants by their own screening rules. If they are interested in a specific restaurant, they can click onto the homepage of the restaurant to search for more detailed information[3]. Thus, we regard the new restaurants a consumer has searched before a transaction trip as the alternatives that the consumer has no experiences on but is interested in.

2.4.3. Consumer Trip Level Panel Data

Our final panel sample data includes 798 consumers' 3335 dining records in 215 restaurants in the period between December 2007 and March 2008. Consumers are included only if their web page browsing histories are observable. This is because we need to make sure the consumers in our sample do search for information from online UGC. To obtain the panel data set for our empirical analysis, we require each consumer to have at least 3 transactions in the sample

[3] The screenshots of Dianping.com are shown in the appendix.

time period. In addition, a consumer needs to have at least 1 transaction before December 2007 in order to be included in the final panel. This criterion is used to define each consumer's prior choice set and to decide whether a consumer chooses a new restaurant at his or her first trip in the sample period. The restaurants included in our final sample are those that were patronized by our sampled consumers from December 2007 to March 2008. The summary statistics for the online reviews associated with these restaurants are shown in Table 2-1.

Table 2-1: Summary Statistics of Restaurant Reviews

	Mean	SD	Min	Max
Number of ratings (in thousands)	0.45	0.49	0.01	3.79
Average rating of taste	1.83	0.24	1.08	2.50
Average rating of ambience	1.78	0.44	0.57	2.97
Average rating of service	1.64	0.42	0.84	3.46
Average price per person (in hundreds)	0.89	0.76	0.25	7.36
Variance of taste ratings	0.66	0.13	0.08	1.26
Variance of ambience ratings	0.57	0.16	0.22	1.24
Variance of service ratings	0.72	0.19	0.23	1.49
Variance of price per person	0.32	1.21	0.00	13.34
Average rating of overall quality	1.75	0.32	1.08	2.63
Variance of overall quality	0.65	0.13	0.35	1.12
Number of popular dish tags (in tens)	0.80	0.66	0.00	2.20
Number of restaurants	215			

Table 2-2: Summary Statistics of Consumer Search

D days window	Whether consumers search UGC		Number of UGC searches	
prior to visiting restaurant	Mean	SD	Mean	SD
$D = 1$	0.44	0.50	1.56	3.28
$D = 2$	0.61	0.49	2.58	4.37
$D = 3$	0.73	0.44	3.46	5.58
$D = 4$	0.81	0.39	4.18	6.15
$D = 5$	0.89	0.32	4.91	6.83
$D = 6$	0.94	0.24	5.57	7.57
$D = 7$	0.99	0.11	6.27	8.43
Number of transactions	3335			
Number of consumers	798			

2.4.4. *Consumer Search and Variety Seeking*

Table 2-2 presents the summary statistics of consumers' search of online UGC. We define a web page browsing record as a "search" activity only when (1) the restaurant a consumer browsed is included in our final sample[4], (2) the website browsing happened in a *D* days' time window before a specific transaction trip. As we can see from Table 2-2, consumers do search online UGC before their dining trips. However, the time window we choose will affect the extent to which we can identify consumers' searches. If we choose 1 day as the time window to define the scope of consumer search, we find that consumers searched online UGC in 44% of the transaction trips. If we use 7 day as the time window, we find that consumers searched online UGC in 99% of transaction trips. Table 2-2 also presents the number of consumer UGC searches before each transaction. Although it is likely that consumers are heterogeneous in terms of retrieving information from memory such that they may or may not be able to recall their search results which happened 7 days' before, we nevertheless use a 7 days' time window to define the scope of consumer search for two reasons. First, a 7 days' time window can help us to ascertain the extent of search for most of transactions (99%). Second, this time window can provide us a reasonable number of new choice alternatives in consumers' consideration sets (as shown in Table 2-3). We observe that 499 consumers chose a restaurant that they had not patronized before and had no online search records within a 7 days' time window in 757 transactions[5]. These transactions happened in 154 restaurants. For these observations, we consider the restaurant a consumer chose as a new restaurant and include it in the consumer's consideration set at that specific trip[6]. It is common

[4] Consumers can search any restaurant. However, we only have transactions data for restaurants that are included in our sample.
[5] We have 1830 such transactions when using a 1 day time window to define the scope of consumer search.
[6] In other words, we assume a consumer had searched the restaurant on the UGC site before the transaction if she finally chose it. We argue that consumers may have searched such a restaurant outside the 7 day window

in the literature to include the chosen alternative in a consumer's consideration set when researchers deal with the issue of large choice sets using a sampling approach (Lemp and Kockelman 2012). In our context, we do not adopt the random sampling approach used in some literature (McFadden 1978; Nerella and Bhat 2004) because we can observe consumers' information search records on the online UGC web site.

Table 2-3 shows the summary statistics of online UGC search and variety seeking choice behavior for consumers in our final sample. There are about 5 trips for each consumer on average and consumers searched online UGC information for 99% of all transactions[7]. On average, consumers accessed restaurants' web pages 6.27 times to search UGC information before a trip. 73% of these searches were for new restaurants (4.74 times for new restaurants and 1.53 times for old ones). Consumers searched 2.88 new restaurants and 0.64 old ones on average. Based on our definition of consumers' consideration set, we find that consumers on average considered 10.6 restaurants which include 3.1 new restaurants and 7.5 old ones in one purchase occasion. Consumers patronized a new restaurant in 55% of all transactions. There is an 84% chance that a consumer switched away to another restaurant across two restaurant trips.

---

or on another computer with a different IP address or without logging in with a registered user identity. However, we caution that we may potentially over-estimate the impact of online UGC if such a restaurant entered the consumer's choice set through offline factors or influences.

[7] The high percentage of online UGC search suggests that our sampled consumers do search and use online UGC before their purchase decisions. Since we only require consumers in our final sample to search online UGC at least once in the three months' time period, the high percentage of search implies that search behaviors are stable and consistent in our sample period.

Table 2-3: Consumer Search and Variety Seeking Statistics

Variables	Mean	SD	Min	Max
Trip search dummy (1 if search)	0.99	0.11	0.00	1.00
Number of searches at one trip	6.27	8.43	0.00	144.00
Number of searches for new restaurants	4.74	7.33	0.00	138.00
Number of searches for old restaurants	1.53	2.86	0.00	39.00
Percentage of searches for new restaurants	0.73	0.36	0.00	1.00
Number of new restaurants searched	2.88	3.84	0.00	80.00
Number of old restaurants searched	0.64	0.80	0.00	7.00
Percentage of new restaurants searched	0.75	0.34	0.00	1.00
Total number of restaurants in consideration set	10.61	6.73	3.00	83.00
Number of new restaurants in consideration set	3.10	3.84	0.00	80.00
Number of old restaurants in consideration set	7.50	5.50	1.00	47.00
New restaurant choice dummy (1 if choose new)	0.55	0.50	0.00	1.00
Restaurant switch dummy (1 if switch)	0.84	0.37	0.00	1.00
Number of trips	4.97	2.62	3.00	17.00
Number of transactions	3335			
Number of consumers	798			

2.5. Econometric Model Specification

Traditionally, it has been assumed that consumers evaluate a product alternative in terms of the utility to be derived from selecting that alternative and subsequently choose the alternative yielding the maximum utility. The underlying assumption that consumers spend time and efforts to evaluate a large number of alternatives is increasingly being questioned (Andrews and Srinivasan 1995; Fotheringham 1988; Gilbride and Allenby 2004; Liu and Arora 2011; Shocker et al. 1991). Researchers have proposed that consumers can use various decision rules to simplify complicated decision tasks (Gilbride and Allenby 2004). An alternative assumption is that consumers make decisions based on a hierarchical or sequential decision process whereby a subset of similar alternatives is selected from the universal set, and the final choice is chosen from the reduced set (Fotheringham 1988; Shocker et al. 1991). In this study, we propose a two-stage choice model by assuming that consumers follow a hierarchical decision process. At the first stage, a consumer decides whether to choose from a set of new restaurants that he or she has no prior consumption experience or from a set of restaurants patronized before. At the second stage, the consumer decides which specific restaurant to patronize. We specify our two-stage choice model below.

2.5.1. First-Stage Decision: Whether to Patronize a New Restaurant?

On a transaction trip at time *t*, consumer *i*'s utility of choosing a new product alternative is:

$$U_{it,c=new} = \alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new} + \varepsilon_{it}^{1} \tag{1}$$

where α_i is the consumer-specific fixed effect which captures the intrinsic preference for a new choice alternative. We assume α_i is normally distributed as $\alpha_i \sim \mathrm{N}(\bar{\alpha}, \sigma_\alpha^2)$. Z_{it} is a vector of control variables, including $InterPurchaseTime_{it}$, $NumOfPerson_{it}$, $NewRestSearchPercent_{it}$, $NewRestSearchPercent\_sq_{it}$, and $OldRestNum_{it}$[8]. Previous studies suggested that $InterPurchaseTime_{it}$ may increase consumers' variety seeking because of an increase in the desire for variety, changes in market information or household composition (Adamowicz and Swait 2013; Chintagunta 1999). $NumOfPerson_{it}$ is used to capture the social aspect of dining choices. $NewRestSearchPercent_{it}$ and its squared term are used to identify how consumers' online UGC search may affect consumers' new product exploration. $OldRestNum_{it}$ captures consumers' prior variety seeking tendency. IV_{it}^{new} is the "inclusive value" which measures the expected value of the maximum

[8] Refer to Table 2-4 for the definition and operationalization of the variables in our model.

Table 2-4: Description of Model Variables

Variables	Description
	Structural Model: First Stage Variables
$NewRest_{it}$	Dependent variable of first stage decision (= 1 if the consumer patronized a new restaurant; = 0 if the consumer chooses from a set of old restaurants)
$\bar{\alpha}$ (Mean of α_i)	Consumer's average intrinsic preference for or tendency of choosing a new product
σ_α (SD of α_i)	Consumer heterogeneity in terms of choosing a new product
$InterPurchaseTime_{it}$	Amount of time between consecutive purchase occasions
$NumOfPerson_{it}$	Estimated number of persons per trip (= transaction expenditure / estimated average price per person from online UGC)
$NewRestSearchPercent_{it}$	Percentage of searches for new restaurants across all searches
$NewRestSearchPercent\_sq_{it}$	Squared term of *NewRestSearchPercent*
$OldRestNum_{it}$	Total number of restaurants that a consumer has patronized before
$IV_{it}^{new}, IV_{it}^{old}$	Inclusive value which measures the expected maximum utility from a set of new or old product alternatives
	Structural Model: Second Stage Variables
$CuisineDummy_j$	16 dummy variables which indicate the cuisine type of a focal restaurant
$TagNum_j$	Number of recommended dish tags which has more than 20 recommendations
$SearchNum_{ijt}$	Number of searches for a focal restaurant before each transaction
$TripNum_{ijt}$	Number of prior trips for a focal restaurant before each transaction
$Promotion_{ijt}$	Whether a focal restaurant has a promotion (= 1 if yes; = 0 if no)
$UserRestDistance_{ij}$	Estimated distance between a consumer's location and a restaurant's address. We compute it by the following steps: 1. We have data on restaurants' detailed addresses. Based on this data, we get the latitude and longitude of each exact address. 2. We approximate consumers' geographical location based on the restaurants they patronized. We compute the weighted average latitude and longitude of restaurants that a consumer has patronized before the time period of the final data sample. The weight assigned to each restaurant is decided by the number of consumption trips. Below is the formula we used: $Latitude_i = \sum_J \frac{Number\ of\ trips\ for\ restaurant\ j}{Number\ of\ total\ trips} Latitude_j$ $Longitude_i = \sum_J \frac{Number\ of\ trips\ for\ restaurant\ j}{Number\ of\ total\ trips} Longitude_j$
$Volume_{ijt}$	Number of reviews of a focal restaurant
$QualityRating_{ijt}$	Average quality rating of a focal restaurant
$VarianceOfQualityRating_{ijt}$	Variance of quality rating of a focal restaurant
$Price_{ijt}$	Estimated average price per person for a focal restaurant
$VarianceOfPrice_{ijt}$	Variance of estimated average price per person for a focal restaurant
	Reduced-Form Model Variables
$NewRestSearchPercent_{it}$	Percentage of searches for new restaurants across all searches
$ExperienceVolume_{it}$	Average number of reviews of prior restaurant choices
$ExperienceQuality_{it}$	Average quality rating of prior restaurant choices
$ExperiencePrice_{it}$	Average estimated price per person of prior restaurant choices
$OldRestNum_{it}$	Total number of restaurants that a consumer has patronized before
$InterPurchaseTime_{it}$	Amount of time between consecutive purchase occasions
$NumOfPerson_{it}$	Estimated number of persons per trip

utility from a set of new alternatives[9]. ε_{it}^{1} is consumer *i*'s first stage idiosyncratic utility component which is unobservable to researchers. γ and λ are the model parameters to estimate.

Consumer *i*'s utility of choosing an old or previously selected product alternative is given by:

$$U_{it,c=old} = \lambda IV_{it}^{old} + \varepsilon_{it}^{1} \tag{2}$$

where IV_{it}^{old} is the "inclusive value" which measures the expected maximum utility from a set of old product alternatives. Assuming that ε_{it}^{1} follows an extreme value distribution, consumer *i*'s choice probability for the new products set and the old products set at time *t* will be:

$$P_{it,c=new} = \frac{e^{\alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new}}}{e^{\alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new}} + e^{\lambda IV_{it}^{old}}} \tag{3}$$

$$P_{it,c=old} = \frac{e^{\lambda IV_{it}^{old}}}{e^{\alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new}} + e^{\lambda IV_{it}^{old}}} \tag{4}$$

2.5.2. *Second-Stage Decision: Which Restaurant to Patronize?*

Conditional on the first stage decision *c*, we specify the utility $U_{ijt|c}$ of choosing alternative *j* as:

$$U_{ijt|c} = \varpi D_j + \beta_i R_{ijt} + \delta R_{ijt} * NewRest_{it} + \varphi X_{ijt} + \varepsilon_{ijt}^{2} \tag{5}$$

D_j is a set of dummy variables which captures the fixed effect of alternative *j*[10]. R_{ijt} is a vector of variables that measure the influence of UGC, including $Volume_{ijt}$, $QualityRating_{ijt}$, $VarianceOfQualityRating_{ijt}$, $Price_{ijt}$, and $VarianceOfPrice_{ijt}$. To capture consumers' different responses for online UGC,

[9] We mathematically define the inclusive value term after we specify the consumer's second stage utility function.

[10] Since there are too many choice alternatives (i.e., 215 restaurants) in our final sample, we estimate 16 cuisine type dummies instead of 215 restaurant dummies in our empirical application, for the practical reason of model parameters' parsimony.

we assume the coefficients for R_{ijt} to vary across consumers as $\beta_i \sim \mathrm{N}(\bar{\beta}, \sigma_i^2)$. $NewRest_{it}$ is a dummy variable which equals 1 if consumer i decides to choose from a set of new products at the first stage. The interaction terms $R_{ijt}*NewRest_{it}$ are used to measure the extent to which consumers' responses to online UGC are different when they choose from a set of new products, compared to when they choose from a set of old products. X_{ijt} is a vector of control variables, including $TagNum_j$, $SearchNum_{ijt}$, $TripNum_{ijt}$, $Promotion_{ijt}$, and $UserRestDistance_{ij}$. $TagNum_j$ controls for the time-invariant effect of popular recommended dishes of restaurant j. $SearchNum_{ijt}$ captures the effect of consumers' searches before their purchase decisions. $TripNum_{ijt}$ captures the effect of the consumer's loyalty or preference for alternative j. $Promotion_{ijt}$ measures the extent to which consumers' choices are influenced by restaurants' promotions on the UGC site. $UserRestDistance_{ij}$ is the distance between consumer i and restaurant j and thus accounts for the effect of transportation cost. ε_{it}^2 is consumer i's unobservable idiosyncratic utility at the second stage.

Based on consumer i's second stage utility function, we define the variable of "inclusive value" as $IV_{it}^c = \ln\left(\sum_c e^{\varpi D_j + \beta_i R_{ijt} + \delta R_{ijt} * NewRest_{it} + \varphi X_{ijt}}\right)$. This variable represents the expected maximum utility consumer i can get from category c (Ben-Akiva and Lerman 1985; Chintagunta 1999). In our specific research context, IV_{it}^{new} is the expected maximum utility which consumer i can get from the set of new products he or she has no prior consumption experience, and IV_{it}^{old} is the expected maximum utility that consumer i can get from the set of old products he or she has patronized before. This variable thus defines how consumers' first stage choice depends on the expected utility from the second stage choice.

Conditional on consumer i's first stage choice, if we assume that ε_{it}^{2} follows an extreme value distribution, the probability of consumer i choosing alternative j at time t is:

$$P_{ijt|c} = \frac{e^{\varpi D_j + \beta_i R_{ijt} + \delta R_{ijt} * NewRest_{it} + \varphi X_{ijt}}}{\sum_c e^{\varpi D_j + \beta_i R_{ijt} + \delta R_{ijt} * NewRest_{it} + \varphi X_{ijt}}} \tag{6}$$

Thus, the unconditional probability of consumer i choosing alternative j at time t is:

$$P_{ijt} = I_{it,\, j=new} P_{ijt|new} P_{new} + (1 - I_{it,\, j=new}) P_{ijt|old} P_{old} \tag{7}$$

$I_{it,\, j=new}$ is a dummy variable which equals 1 if alternative j is a new product for consumer i at time t.

Our model is consistent with the sequential multinomial logit model in the literature (Ben-Akiva 1973; McFadden et al. 1977). According to McFadden et al. (1977), a sufficient condition for a sequential model to be consistent with individual utility maximization is that the coefficient of inclusive value is between 0 and 1, i.e., $0 < \lambda < 1$[11]. Similar to the nested logit model, our model relaxes the IIA (independence of irrelevant alternatives) property by assuming that IIA holds within each subset of products (i.e., new products or old products) but does not hold in general for alternatives in different subsets. In other words, we assume that the substitution pattern between two old restaurants (or two new restaurants) is different from the substitution pattern between an old restaurant and a new restaurant. However, our model is different from the nested logit model because a

[11] Please refer to McFadden et al. (1977) for the theoretical foundations of our model.

choice alternative's membership in subset *c* from which the consumer selects a particular alternative is not predetermined *a priori*. For a nested logit model, once a subset *c* is chosen, the set of choice alternatives within *c* is static and known with certainty across the sample period (Train 2003). In our model, the new and old subsets of choice alternatives are usually different for different consumers and vary across purchase occasions.

2.6. Model Estimation and Findings

2.6.1. Identification and Estimation Methods

We use a limited information structural modeling and estimation approach to identify the effect of consumers' online UGC search on new product exploration. A critical issue in the model estimation is the potential endogeneity of consumers' new product searches (i.e., for variable $NewRestSearchPercent_{it}$) at the first stage decision because consumers who have the intention to seek variety are likely to search for new products from online UGC. To account for this source of potential endogeneity, we adopt the control function approach proposed by Petrin and Train (2010). Consumers' amount of prior contributions to the UGC site in a 7 days' window is used as an instrumental variable for the extent of new product search. We expect that a consumer's UGC contribution is correlated with his or her new product search. However, we do not expect the unmeasured factors which affect new product exploration to be correlated with prior UGC contributions. $NewRestSearchPercent_{it}$ is specified as linear in the instrument plus a separate error:

$$NewRestSearchPercent_{it} = \kappa ContriNum_{it} + \varepsilon_{it}^{3} \qquad (8)$$

We decompose the earlier specified ε_{it}^{1} in equation (1) as: $\varepsilon_{it}^{1} = \mu_{it}^{1} + \varepsilon_{it}^{1'}$. μ_{it}^{1} captures consumers' unobserved intention to seek variety, which is correlated with consumers' new product search. We re-specify μ_{it}^{1} and ε_{it}^{3} to be jointly distributed as bivariate normal and $\varepsilon_{it}^{1'}$ is iid extreme value. Then, consumer *i*'s utility of exploring a new product alternative with the control function is:

$$U_{it,c=new} = \alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new} + \theta_1 \varepsilon_{it}^{3} + e_i + \varepsilon_{it}^{1'} \qquad (9)$$

e_i is a consumer-specific error component generated by the control function and is normal with mean zero and constant variance. It cannot be separately identified from α_i. We derive the new probabilities that consumer *i* chooses the new product set and the old product set at time *t* as:

$$P_{it,c=new} = \frac{e^{\alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new} + \theta_1 \varepsilon_{it}^{3} + e_i}}{e^{\alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new} + \theta_1 \varepsilon_{it}^{3} + e_i} + e^{\lambda IV_{it}^{old}}} \qquad (10)$$

$$P_{it,c=old} = \frac{e^{\lambda IV_{it}^{old}}}{e^{\alpha_i + \gamma Z_{it} + \lambda IV_{it}^{new} + \theta_1 \varepsilon_{it}^{3} + e_i} + e^{\lambda IV_{it}^{old}}} \qquad (11)$$

Another issue is that we have two potentially endogenous variables ($NewRestSearchPercent_{it}$ and $NewRestSearchPercent\_sq_{it}$) while we only have one instrument. We deal with this issue by adopting the method suggested by Wooldridge (2002). We first regress $NewRestSearchPercent_{it}$ against the instrument (i.e., $ContriNum_{it}$, the number of prior UGC contributions). We compute the fitted value ($FitValue_{it}$) and residual (Rsd_1) based on the regression results. We then regress $NewRestSearchPercent\_sq_{it}$ against the squared term of the fitted value from the first regression (i.e., $FitValue\_sq_{it}$), following which we compute the residual (Rsd_2) from the second regression results. The two residual terms Rsd_1 and Rsd_2 enter in the two-stage choice model without any transformation. We conduct an F-test for the instrument in each regression. The F-test value was well over 10 in each case, indicating that our instruments are suitably good ones (Staiger and Stock 1997).

We finally estimate our two-stage model using the simulated maximum likelihood estimation technique. The model parameters for both stages are simultaneously estimated to capture (1) the interdependence of parameters, and (2) the interdependence of the two stage decisions, i.e., consumers' first stage choices are dependent on the expected utility from the second stage choices, while consumers' choice set at the second stage is decided by their first stage decisions.

2.6.2. Reduced-Form Model Analysis Results

We first use reduced-form binary Probit and Logit regressions to present some evidences that consumers' searches and prior consumption experiences have an impact on their new product exploration. Table 2-5 reports the results[12]. Models 1 and 2 show the estimation results of the panel level Probit and Logit models. Models 3 and 4 show the estimation results for which we addressed the endogeneity of consumer search using a control function approach. All these models show a significant positive relationship between consumer search and new product exploration. We conduct a *Wald* test of endogeneity of consumer search behavior and find a significant result (χ^2 = 8.15, and Prob. > χ^2 = 0.0043), which implies that consumers' search behaviors are endogeneous (Wooldridge 2002).

[12] We conduct many robustness checks too, where the detailed results are shown in the web appendix. First, we operationalize consumers' search behavior in many alternative ways, such as the percentage of new unique restaurants searched, total number of searches, numbers of searches for new and old restaurants. Second, we estimate these models by including the quadratic term of *NewRestSearchPercent*. However, all the coefficients for quadratic terms are not significant. Third, we also estimate these models by using other operationalizations of consumer's prior experiences (e.g., maximum and most recent volume and quality rating of reviews of prior choices). The coefficients for the alternative operationalizations of prior experience are all insignificant.

Table 2-5: Model Estimation Results from Reduced-Form Analysis

	Model 1: Panel Probit	Model 2 Panel Logit	Model 3 Panel Probit with IV	Model 4 Panel Logit with IV
Variables	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)
NewRestSearchPercent	1.27^{***}	2.14^{***}	5.18^{**}	8.60^{**}
	(0.078)	(0.14)	(1.68)	(2.83)
ExperienceVolume	0.13	0.21	0.20^{+}	0.33^{+}
	(0.11)	(0.19)	(0.12)	(0.20)
ExperienceQuality	-0.54^{*}	-0.91^{*}	-0.51^{*}	-0.86^{*}
	(0.24)	(0.41)	(0.24)	(0.41)
ExperiencePrice	0.70^{***}	1.17^{***}	1.03^{***}	1.72^{***}
	(0.21)	(0.35)	(0.25)	(0.42)
OldRestNum	-0.033^{***}	-0.055^{***}	-0.025^{***}	-0.042^{***}
	(0.0064)	(0.011)	(0.0072)	(0.012)
InterPurchaseTime	0.0098^{*}	0.016^{*}	-0.024	-0.039
	(0.0045)	(0.0077)	(0.015)	(0.025)
NumOfPerson	0.0083^{+}	0.015^{+}	0.0084^{*}	0.016^{+}
	(0.0043)	(0.0081)	(0.0043)	(0.0082)
Control function residual			-3.92^{*}	-6.48^{*}
			(1.68)	(2.83)
Constant	-0.23	-0.39	-3.35^{*}	-5.54^{*}
	(0.38)	(0.65)	(1.39)	(2.34)
Consumer fixed effect	-included-	-included-	-included-	-included-
Number of consumers	798	798	798	798
Number of observations	3335	3335	3335	3335
Log-likelihood	-2032.6	-2032.0	-2029.8	-2029.3
AIC	4083.2	4082.1	4079.6	4078.6
BIC	4138.2	4137.1	4140.8	4139.8

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For the consumer's experience-related variables, the coefficient for *ExperienceVolume* is insignificant, while those for *ExperienceQuality* and *ExperiencePrice* are significant and have the expected signs. This result implies that consumers' prior experiences (measured by average quality rating and average price of prior choices) have a significant influence on their new product exploration. We find that the number of restaurants that consumers have patronized before and the estimated number of persons per trip significantly affect consumers' new product exploration. However, the effect of *InterPurchaseTime* becomes insignificant after we control for the endogeneity of search by using the control function method.

2.6.3. Two-Stage Structural Model Analysis Results

To highlight the importance of incorporating consumers' hierarchical decision process in new product exploration, we estimate a one-stage choice model as the benchmark model which assumes consumers simultaneously evaluate all (both "new" and "old") choice alternatives and choose the alternative which yields the maximum utility. We specify consumer *i*'s utility function as:

$$U_{ijt} = \varpi D_j + \beta_i R_{ijt} + (\alpha_i + \gamma Z_{it} + \delta R_{ijt}) * I_{it,\, j=new} + \varphi X_{ijt} + \varepsilon_{ijt} \tag{12}$$

Assuming ε_{ijt} follows an extreme value distribution, the probability of consumer *i* choosing alternative *j* at time *t* will be:

$$P_{ijt} = \frac{e^{\varpi D_j + \beta_i R_{ijt} + (\alpha_i + \gamma Z_{it} + \delta R_{ijt}) * I_{it,\, j=new} + \varphi X_{ijt}}}{\sum e^{\varpi D_j + \beta_i R_{ijt} + (\alpha_i + \gamma Z_{it} + \delta R_{ijt}) * I_{it,\, j=new} + \varphi X_{ijt}}} \tag{13}$$

Table 2-6 presents the results of our benchmark model (Model 1). Model 2 in Table 2-7 shows the results from our two-stage decision model without accounting for the potential endogeneity of consumer search. The additional parameter for the inclusive value term in Model 2 examines the effect of the expected maximum utility that consumers can get when choosing from a set of new or old restaurants. Model 3 in Table 2-7 presents the estimation results of the full model which assumes consumer choice follows a two-stage hierarchical decision process and accounts for the endogeneity of consumer search. We have two more parameters in Model 3 which capture the effect of residual terms of the control function. According to the model fit statistics, including the log-likelihood function value, Akaike Information Criterion (AIC), and Bayesian Information

Criterion (BIC), we find that the two-stage choice models fit much better with our sample data than the one-stage benchmark model.

In terms of parameters estimation, we get generally consistent results from all three models. After we account for the endogeneity of consumer search at their first stage decision in Model 3, the magnitude of coefficients for both the linear and quadratic terms of consumer new restaurant search becomes smaller. Based on the estimated parameters from Model 3, we calculated the minimum point of the response curve to be at 0.566. According to our summary statistics on consumer search, we find that most of our data are located at the right part (or the increasing part) of the minimum point. The increasing slope (14.12x[13]) is lower than the one we get from Model 2 (15.54x), which implies that the effect of UGC search on a consumer's new product exploration becomes smaller after we control for the endogeneity of consumers' new restaurant search. This result is consistent with our expectation that consumers who want to try a new product are more

[13] x is the value of $NewRestSearchPercent_{it}$. It is required to be between 0.566 and 1 such that there is a positive slope.

Table 2-6: Model Estimation Results from One-Stage Choice Model

	Model 1	
Variables	Estimates	Std. Err.
$I_{it, j=new}$ * $\bar{\alpha}$ *(Mean of* α_i*)*	4.07***	0.431
$I_{it, j=new}$ * σ_α *(SD of* α_i*)*	0.90***	0.095
$I_{it, j=new}$ * *NewRestSearchPercent*	-11.89***	0.907
$I_{it, j=new}$ * *NewRestSearchPercent_sq*	9.62***	0.701
$I_{it, j=new}$ * *InterPurchaseTime*	0.002	0.009
$I_{it, j=new}$ * *NumOfPerson*	0.03**	0.010
$I_{it, j=new}$ * *OldRestNum*	0.05***	0.011
SearchNum	0.28***	0.014
TripNum	0.14***	0.009
Promotion	-0.19**	0.059
TagNum	0.28***	0.068
UserRestDistance	-0.17***	0.012
Volume (Mean)	-0.90***	0.115
Volume (SD)	0.61***	0.093
QualityRating (Mean)	0.32+	0.173
QualityRating (SD)	-0.16	0.689
VarianceOfQualityRating (Mean)	-0.34	0.492
VarianceOfQualityRating (SD)	2.91***	0.521
Price (Mean)	-0.71***	0.207
Price (SD)	1.03***	0.160
VarianceOfPrice (Mean)	-0.29+	0.161
VarianceOfPrice (SD)	-0.07	0.131
Volume * $I_{it, j=new}$	0.64***	0.095
QualityRating * $I_{it, j=new}$	-0.32	0.209
VarianceOfQualityRating * $I_{it, j=new}$	0.99	0.611
Price * $I_{it, j=new}$	-0.65**	0.218
VarianceOfPrice * $I_{it, j=new}$	0.36*	0.172
CuisineDummy	-included-	
Number of coefficients	43	
Number of consumers	798	
Number of observations	3335	
Log-likelihood	-5704.7	
AIC	11495.4	
BIC	11859.8	

Notes:

(1) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(2) "*Mean*" and "*SD*" denote the mean and standard deviation of the coefficient.

Table 2-7: Model Estimation Results from Two-Stage Choice Models

Variables	Model 2		Model 3 (Control Function)	
	Estimates	Std. Err.	Estimates	Std. Err.
	First Stage: Variety Seeking Decision			
NewRestSearchPercent	-9.48^{***}	0.793	-7.99^{***}	1.206
NewRestSearchPercent_sq	7.77^{***}	0.606	7.06^{***}	0.719
Control function residual (linear term)			-1.55^{+}	0.928
Control function residual (squared term)			0.75^{+}	0.396
IV^c	0.72^{***}	0.047	0.73^{***}	0.048
$\bar{\alpha}$ *(Mean of* α_i*)*	3.20^{***}	0.342	3.05^{***}	0.351
σ_α *(SD of* α_i*)*	-0.14	0.088	-0.14	0.087
InterPurchaseTime	0.02^{*}	0.008	-0.01	0.029
NumberOfPerson	0.02^{*}	0.008	0.02^{*}	0.008
OldRestNum	0.04^{***}	0.009	-0.01	0.049
	Second Stage: Product Alternative Decision			
SearchNum	0.24^{***}	0.015	0.24^{***}	0.015
TripNum	0.15^{***}	0.010	0.15^{***}	0.010
Promotion	-0.20^{***}	0.057	-0.20^{***}	0.057
TagNum	0.35^{***}	0.062	0.35^{***}	0.062
UserRestDistance	-0.17^{***}	0.012	-0.17^{***}	0.012
Volume (Mean)	-0.78^{***}	0.110	-0.78^{***}	0.110
Volume (SD)	0.41^{***}	0.121	0.41^{***}	0.122
QualityRating (Mean)	0.56^{***}	0.168	0.56^{***}	0.169
QualityRating (SD)	-0.30	0.250	-0.32	0.247
VarianceOfQualityRating (Mean)	0.63	0.493	0.67	0.496
VarianceOfQualityRating (SD)	-0.59	0.935	-0.77	0.924
Price (Mean)	-1.15^{***}	0.185	-1.17^{***}	0.185
Price (SD)	0.81^{***}	0.163	0.80^{***}	0.163
VarianceOfPrice (Mean)	0.08	0.173	0.10	0.171
VarianceOfPrice (SD)	0.14	0.130	0.13	0.128
*Volume * NewRest*	0.46^{***}	0.091	0.46^{***}	0.091
*QualityRating * NewRest*	-0.18	0.214	-0.19	0.215
*VarianceOfQualityRating * NewRest*	0.13	0.606	0.11	0.613
*Price * NewRest*	-0.48^{*}	0.226	-0.45^{*}	0.225
*VarianceOfPrice * NewRest*	0.09	0.197	0.08	0.196
CuisineDummy	-included-		-included-	
Number of coefficients	44		46	
Number of consumers	798		798	
Number of observations	3335		3335	
Log-likelihood	-5306.2		-5303.8	
AIC	10700.4		10699.6	
BIC	10969.3		10980.7	

Notes:

(1) $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

(2) "*Mean*" and "*SD*" denote the mean and standard deviation of the coefficient.

likely to search for new restaurants from online UGC. The coefficients for the residual terms of the control function are marginally significant, providing some evidence that consumer search behaviors are endogeneous. Both variables of *InterPurchaseTime* and *OldRestNum* are used to account for consumers' intrinsic variety seeking tendency at each specific trip. Their coefficients become insignificant after we include the control function residuals. This result also provides support that the instrumental variables and control functions we used can help us properly adjust the estimated parameters for the variables associated with consumer search.

2.6.4. Hypotheses Tests Results

We test our hypotheses based on the estimation results of Model 3. First, we get a significant U-shape relationship between consumers' new product search and their variety seeking decision. The U-shape relationship suggests that there is a threshold point in consumer search behavior. New product exploration is positively related to the extent of new product search when consumers' new product search percentage is greater than the threshold. Since consumers may have diverse incentives to search online UGC information, such as searching for promotion information or recommended dishes of a restaurant, a possible explanation is that only when consumers' new product search percentage is greater than the threshold, then can such search activities signify heightened incentives to explore a new product. Based on our summary statistic of consumers' new product search, the percentage of new restaurants search is larger than the minimum threshold point (i.e., 0.566) in 73.6% of transactions in our data sample. Thus, we argue that we generally find a positive relationship between consumers' new product exploration tendency and their information search for new products. Hypothesis H1 is thus supported.

The coefficient for the inclusive value is 0.73 and statistically significant. First, the parameter is between 0 and 1, suggesting that our two-stage choice model is consistent with individual utility maximization (McFadden et al. 1977). Second, the positive and significant coefficient suggests that a consumer is more likely to choose a new restaurant when the set of new restaurants he or she has searched can provide higher expected utility than the set of restaurants patronized previously. Thus, our hypothesis H2 is supported. In addition, the mean of consumer fixed effect α_i is positive and significant, suggesting that consumers on average prefer to try a new product. The standard deviation (SD) of α_i is not significant, implying that consumers are homogeneous in terms of preferring a new product. The coefficient for *NumOfPerson* is positive and significant, which implies that consumers are more likely to patronize a new restaurant when there are more persons in that trip. However, the coefficients for *InterPurchaseTime* and *OldRestNum* are not significant after we account for the endogeneity of search.

In terms of the impact of UGC information attributes on consumer choice, we find that the mean and SD of the coefficient for *Volume* are -0.78 (SE=0.110) and 0.41 (SE=0.122) respectively. The negative and significant mean of the coefficient for *Volume* suggests that consumers are on average less likely to patronize a restaurant which has a large number of reviews, controlling for everything else. This result contradicts with prior studies which suggest a positive relationship between the number of reviews and product sales (Chevalier and Mayzlin 2006; Liu 2006). A possible explanation is that a popularly reviewed restaurant is more likely to be crowded and thus consumers may switch to other choices. Alternatively, consumers are more likely to find negative reviews which may deter them from choosing a restaurant with a high volume of reviews (Chevalier and Mayzlin 2006). In addition, since consumers on average have a tendency to try a new restaurant, it is likely that a restaurant with a large number of reviews has already been patronized by the consumers in our sample, leading to a negative relationship between new product exploration and the volume of UGC. The large (relative to the mean of the coefficient) and significant SD of the coefficient for *Volume* suggests that consumers are substantially heterogeneous in terms of the responses to the number of online UGC.

The mean of the coefficient for *QualityRating* is positive and significant, implying that the average quality rating from UGC has a positive influence on individual consumer's variety seeking choices. However, the SD of the coefficient for *QualityRating* is not significant, suggesting that consumers are homogeneous in terms of the responses to the average quality ratings. This result is not surprising because the average quality ratings from UGC reflects the quality of a restaurant and consumers usually prefer a superior product no matter how different they are in term of preferences or backgrounds. The mean and SD of the coefficient for *VarianceOfQualityRating* are both insignificant, implying that consumers are on average risk-neutral to quality ratings and have no specific preference on restaurants with a higher or lower variance of quality ratings. The mean and SD of the coefficient for *Price* are -1.17 (SE=0.185) and 0.80 (SE=0.163) respectively and both are highly significant. This result suggests that on average, consumers are more likely to choose a restaurant with a lower price. However, consumers are heterogeneous in terms of price sensitivity such that some consumers are less price sensitive than others. The mean and SD of the coefficient for *VarianceOfPrice* are both insignificant, implying that consumers are on average risk-neutral to dining costs and have no specific preference on restaurants with a higher or lower variance of prices. In addition, the coefficient for *TagNum* is positive and significant, implying that consumers are more likely to patronize a restaurant that have more tags of popularly recommended dishes. In summary, we find significant influences of many information attributes of UGC on individual consumers' dining choices. Consumers are also significantly heterogeneous in terms of responses to some information attributes from online UGC. Thus, our hypotheses H3 and H4 are supported.

The coefficient for *Volume*NewRest* is positive and highly significant, implying that when consumers choose from a set of new products, the number of reviews has a higher positive effect on consumers' decisions. The negative and significant coefficient for *Price*NewRest* suggests that consumers are more price sensitive when exploring new products. Thus, our hypothesis H5 is supported. The coefficients for other interaction terms (*QualityRating*NewRest, VarianceOfQualityRating*NewRest,* and *VarianceOfPrice*NewRest*) are not significant. This is not surprising because we generally do not find evidence of consumer heterogeneity in the main effects of these information attributes of online UGC.

In terms of the control variables, the coefficient for *SearchNum* is positive and highly significant, indicating that the more times an individual consumer searches a restaurant, the more likely he or she will choose the restaurant. The coefficient for *TripNum* is positive and significant, indicating that consumers are more likely to patronize a restaurant that they have more prior experiences when choosing from a set of old restaurants. This result is consistent with prior consumer loyalty studies (Guadagni and Little 1983). Surprisingly, we find a negative and significant coefficient for *Promotion,* which implies that consumers are less likely to patronize a restaurant that offers promotions on the UGC site. Since only a small proportion (29.8%) of restaurants offered promotions during our data sample period, consumers may have neglected such promotions from UGC. Another possible reason is that restaurants which did not have online promotions on the UGC site may had offered offline promotions, which may had counteracted against the influence of online UGC promotions. The coefficient for *UserRestDistance* is negative and significant, suggesting that consumers are less likely to choose a restaurant which is farther away from their own locations.

2.7. Discussion and Conclusion

In this study, we investigate the underlying process of how individual consumers perceive and use online UGC information to guide their new product exploration and purchase decisions. Specifically, we empirically analyze the relationship between online user-generated reviews and consumers' new product exploration by assembling a novel data set of individual consumers' restaurant patronage transactions, consumers' information search records, and online reviews of restaurants. We propose that consumers' purchase decisions follow a two-stage process and online UGC plays important roles in the whole process. At the first stage, consumers decide whether to explore a new product or choose from a set of products that they have purchased before. We propose that online UGC affects consumers' variety seeking decision via the *information effect* and *experience effect*. At the second stage, consumers decide which specific product to purchase. We propose that online UGC affects consumers' purchase decisions via the *competition effect*. The *information effect* suggests that consumers are more likely to explore a new product when they are exposed to online UGC information because online UGC increases the awareness of more choice alternatives. The *experience effect* implies that the influence of online UGC on consumers' new product exploration behaviors depends on consumers' prior consumption experiences. A consumer, upon exposure to relevant online UGC, is posited to seek more variety when the UGC signals relatively superior choice alternatives when compared to the consumer's prior choices. The *competition effect* posits that a consumer's perception of a product depends not only on reviews of a focal product but also those of rival products. Online UGC helps consumers to make the final purchase decisions by providing information of competing products. This is crucial when consumers choose from a set of new products for which they have no prior consumption experiences.

These previously undocumented findings on consumer new product exploration behaviors in the context of user-generated reviews have important implications for academic research in the information systems and marketing fields, as well as for practice in terms of content marketing and designing product recommendation systems in e-commerce websites. This study contributes to the academic literature in following aspects. First, we illustrate how researchers can make use of consumers' search data to explicitly model consumers' decision process in the light of online UGC. Previous studies that examined the impact of online UGC on individual consumer decisions usually make the implicit assumption that consumers search online UGC before their purchase decisions (Chevalier and Mayzlin 2006; Goh et al. 2013; Zhao et al. 2013). However, none of these studies have data on individual consumer's search behavior. Taking advantage of consumers' search records, we verify consumers' UGC search before their purchase decisions, operationalize relevant variables of consumers' UGC searches, and define consumers' consideration sets in their purchase incidences. The detailed search records of consumers can help us to evaluate the influence of online UGC in a more accurate and revealing manner.

Second, we examine how individual consumers perceive and use UGC information to guide their new product exploration and purchase decisions. Our results show that online UGC affects consumers' new product exploration and purchase decisions by (1) informing consumers more choice alternatives in a market (*information effect*), (2) highlighting new choice alternatives that have a higher expected utility than that of their prior choices (*experience effect*), and (3) signaling the quality of competing choice alternatives (*competition effect*). The information effect suggests that online UGC platforms are valuable for variety seeking consumers by drawing their attention to new choice alternatives. This supplements the existing literature which proposes that online UGC can benefit consumers by signaling the quality of products (Zhao et al. 2013). To the best of our knowledge, this is the first study that empirically examines the value of online UGC in terms of helping consumers to identify new products that match with their specific preferences. The experience effect implies that the influence of online UGC on individuals' new product exploration decision is dependent on a consumer's prior consumption experiences. Although a new product may tempt consumer switches across products or brands, the experience effect reveals that a product with a high expected utility is more likely to encourage variety seeking when the consumer has relatively inferior prior product consumption experiences as measured by the UGC information content. The competition effect suggests that individual consumers rely on UGC information to evaluate the quality of competing choice alternatives and make their purchase decisions, which is consistent with the prior literature (Lu et al. 2013; Zhao et al. 2013). Our results show that consumers are heterogeneous in response to online UGC. We find evidence that online UGC is more influential when consumers explore new products than when they choose from their prior choice sets. Specifically, we find that the number of reviews has a higher positive influence on consumers' choice decisions and consumers are more sensitive to price when they explore new products.

Third, our two-stage choice model extends the UGC literature by demonstrating the role of UGC in consumers' hierarchical choice process. We propose that online UGC plays different roles at different stages of the consumer choice process. In our specific research context, online UGC draws awareness to new choice alternatives (*information effect*) and helps consumers to evaluate the expected value of these new alternatives relative to their prior choices (*experience effect*) at the first stage. At the second stage, online UGC helps consumers to make the final decision by signaling the quality of products in the chosen product subset. Model fit statistics imply that our two-stage model fits the data much better than the one-stage choice model. Researchers have pointed out that consumers' purchase decisions follow a hierarchical or sequential process when they need to choose from a large number of alternatives (Fotheringham 1988; Gilbride and Allenby 2004; Liu and Arora 2011). Thus, our current result can be applied to other competitive product markets in which a large number of producers compete with each other to satisfy the wants and needs of a large number of consumers.

Fourth, our study shows that as a result of extrinsic exposure to UGC information, consumers are more likely to explore new products in their purchase decisions. Previous studies have identified that satiation, preference uncertainty, and external situations are three motivating factors of consumers' variety seeking (Kahn 1995; McAlister and Pessemier 1982). This study thus pinpoints online UGC as an external trigger of consumer variety seeking behaviors. In terms of such external factors, companies can strategically alter relevant marketing-mix variables to affect consumers' variety seeking behaviors, for example, by using price promotions and retail environment manipulations (Kahn 1995; Lu et al.

2013). Our findings imply that consumers' online UGC search behavior can be used as a good predictor of variety seeking tendency, and importantly, user-generated reviews can be a potent extrinsic tool for content marketing purposes to influence choice switching.

Our study has a number of important practical implications for retailers who are interested in pursuing content marketing strategies in various social media platforms. First, the information effect in our study suggests that online UGC is valuable for variety seeking consumers by simply providing and highlighting new choice alternatives. The competition effect also highlights that the number of online UGC has a higher positive impact on consumer choice when consumers explore a new product. Therefore, marketers should strategically stimulate consumers to generate more word of mouth information, especially for marketers who want to promote their new products.

Second, our findings in terms of the experience effect of online UGC imply that in order to influence consumers' choice or variety seeking, it is necessary and important for marketers to take consumers' prior consumption experiences into consideration (e.g., by assimilating information from past reviews, CRM databases and customer satisfaction surveys). The experience effect implies that consumers have the tendency to switch to products which can offer a higher expected utility than that of their prior choices. This asymmetric choice switching tendency between highly rated restaurants and lowly rated restaurants has important implications on market competition. Given consumers' exposures to online UGC, highly rated restaurants are more likely to capture incremental market share from lowly rated ones. As a result, positive online word of mouth not only increases a firm's customer base but can also mitigate against customer defections.

Third, our results have practical implications for managers and designers of product recommendation systems on e-commerce websites. In order to increase such websites' informativeness for consumers, it is beneficial to account for consumers' specific purchase or browsing history in personalizing recommendations. Our results suggest that consumers are usually interested in products that can offer superior or better experiences than those of their prior choices, or in products that are rated relatively better among a group of alternatives. Thus, product recommendation website designers should take individual consumers' consumption experience into consideration when designing recommendation systems. For example, it will be directly more effective to recommend consumers a product with a higher quality rating of online UGC than that of their prior choices[14]. In addition, when recommending a product, it is instructive to show how this product is relatively rated in the market. It is critical for designers to devote their efforts to facilitate and enhance consumers' product exploration and evaluation through the experience and competition effects when consumers routinely utilize information from online UGC.

This study has several limitations, some of which can serve as fruitful areas for future research. First, our empirical model omits consumers' information

[14] In location-based recommendations implemented on *Apple iOS* 6's integration with *Yelp* and *OpenTable*, a mobile phone can presumably track the locations a user has been in various markets, such that the restaurant review or reservation app can intelligently recommend options that are superior to those searched or patronized previously.

search decisions such as how many products to search and which specific product to search. We treat an individual consumer's UGC site browsing history data as exogenously given[15] and make use of it to define a consumer's consideration set at each purchase occasion. Future search can extend the current study by explicitly modeling consumers' decision of information search. This decision can help us to understand how consumers narrow down their consideration sets in the information search stage. Second, our current model does not incorporate consumers' product quality learning behavior. In our current context, it is a complex research issue to investigate a consumer's learning behavior because a consumer can learn product quality from both online UGC and their own consumption experiences. We systematically examine such consumer learning behavior in a separate study (Wang et al. 2013). Third, we do not account for the qualitative influence of review texts and comments on individual consumers' choice in this study. Previous studies have documented the influence of these qualitative comments on consumer behaviors (Archak et al. 2011a; Ghose et al. 2012; Netzer et al. 2012; Tirunillai and Tellis 2012). Future research can examine the effect of the UGC and review texts on consumers' variety seeking behaviors, although we have to qualify that this is highly challenging given the linguistic challenges involved in the text mining of the Chinese language. Fourth, as in most previous studies which investigate the influence of online UGC, the influence of alternative unobserved sources of information, such as information from other UGC platforms and offline word of mouth, cannot be ruled out. As such, while

[15] It should be noted that we do account for the potential endogeneity of consumers' new product search on their new product exploration behaviors.

we acknowledge that a potential source of endogeneity bias is likely from omitted variables, we do include a comprehensive set of control variables and robustness checks[16] in our empirical analyses.

[16] Refer to the appendix for the results of robustness checks conducted (which showcase the consistency of our findings).

CHAPTER 3. STUDY TWO

LEARNING FROM SELF AND THE CROWD: THE INFORMATIONAL ROLE OF USER-GENERATED CONTENT FOR FREQUENTLY PURCHASED PRODUCTS

3.1. Introduction

Online information channels, such as product reviews, chat rooms, recommendation sites and wikis, have rapidly gained popularity on the Internet and are increasingly available for a wide range of products and services. These online channels have become important sources of information for consumers (Chen and Xie 2008; Dellarocas 2003; Dellarocas 2006; Mayzlin 2006). On social media platforms, the body of the information that consumers generate is popularly known as user-generated content (UGC). It is expected that 155 million US Internet users will access some form of UGC by 2013 (Verna 2009). This large-scale sharing of consumption experiences is important for the marketing success of firms because it has the potential to reduce consumers' uncertainty about the quality of a product or service before their purchase decisions and thus alleviate the information asymmetry between firms and consumers (Akerlof 1970). This especially facilitates purchase decisions involving experience goods whose quality cannot be inspected prior to the purchase. According to comScore (2007), 24 percent of Internet users seek for and read UGC prior to paying for a service that is delivered offline. Studies also suggest that UGC has a greater influence on consumer choice than firms' traditional marketing activities (Trusov et al. 2009).

The influence of UGC has attracted considerable attention from both practitioners and researchers. Online UGC has the potential to attract consumer visits, increase the time spent on a site, and create a sense of community among frequent shoppers (Kumar and Benbasat 2006). An increasing number of firms, including Amazon, Yelp, Dianping, and Epinions, are offering UGC services. These firms provide millions of UGC on diverse products and services on their websites and attract numerous visits daily. Researchers have examined the impact of UGC at both the aggregated product level (Chevalier and Mayzlin 2006; McAlister et al. 2012; Sonnier et al. 2011; Tirunillai and Tellis 2012) and the individual consumer level (Albuquerque et al. 2012; Chan et al. 2012; Goh et al. 2013; Zhao et al. 2013).

This study contributes to the UGC literature with respect to two dimensions. First, we propose a structural model to capture consumer learning from both online UGC and consumption experiences. Adopting the Bayesian updating framework, we demonstrate how individual consumers perceive and interpret the information embedded in online UGC to update their quality perceptions of products. This result has important implications for both marketers and website managers. In order for effective marketing with UGC on digital and social media, it is important for marketers to discern how individual consumers use and respond to online UGC. For review websites and social media operators, an insight into how individual consumers view, perceive, and use online review information has crucial implications in terms of website design, information management strategies, and the use of information technologies as a means of extending reach and enhancing the richness of consumer reviews

Second, we investigate how consumers' experiential leaning can moderate the informational role of online UGC for frequently purchased products. In fact, for frequently purchased products, there is informational value for a consumer intending to purchase a product because that consumer can repeatedly buy the product if he or she likes it (McFadden and Train 1996). Marketing researchers have suggested that consumer experiential learning is of great importance in the consumer choice process (Erdem and Keane 1996). As consumers can learn product quality from their own consumption experiences, the informational role of online UGC will decrease when consumers gain more experiences. Thus, an understanding on how consumers' experiential learning moderates the effect of UGC on consumer choice is of great importance for marketers in evaluating the impact of online UGC and thus provides useful guidance for firms when they run marketing campaigns on these new social media platforms. In addition, it is important to take into account consumers' experiential leaning when examining the impact of UGC bias. Previous studies have demonstrated that online UGC may fail to reflect the true quality of a product due to manipulation by firms (Dellarocas 2006; Mayzlin 2006), consumers' self-selection (Li and Hitt 2008), social dynamics (Godes and Silva 2012; Wang et al. 2010), and price effect (Li and Hitt 2010). The potential bias of online UGC has crucial implications in terms of a firm's profits, pricing strategy, and consumer surplus (Li and Hitt 2008; Li and Hitt 2010; Moe and Trusov 2011). There could be substantial value for UGC websites to invest in the prevention and elimination of this bias (Li and Hitt 2008). Since consumers are able to learn product quality through their own consumption experiences in the frequently purchased product category, we expect that the potential bias of online UGC will be alleviated by consumers' experiential learning.

Thus, to deepen our understanding of the informational roles of UGC for frequently purchased products, we specifically propose the following research questions:

(1) How does an individual consumer interpret the product information embedded in online UGC to guide his or her purchase decisions?

(2) Regarding frequently purchased products, when consumers can learn product quality from their consumption experiences, how will consumers' experiential learning moderate the informational role of UGC?

We calibrate our proposed model on a unique data set consisting of restaurant reviews and consumers' restaurant dining records. Our final panel data set comprises 4724 dining records of 539 consumers in 19 restaurants. Our empirical analysis leads to two important findings. First, consumers can learn restaurant quality from both online UGC and their own consumption experiences in dining choices. There is a significant amount of learning from consumers' own consumption experiences, much more than from online UGC. Second, the neglecting of consumers' experiential learning can lead to over-estimation of the impact of online UGC on consumer restaurant choice. We demonstrate how our model can be used for firms' decisions on word-of-mouth marketing. Our policy simulation results suggest that the impact of online UGC on consumer choice decreases with the number of consumers' consumption trips. Thus, online UGC promotions may be influential only for new products and the impact may be of short duration.

The rest of this study is organized as follows. In the next section, we briefly review the relevant studies and discuss how we extend the existing literature. In Section 3, we specify our econometric model which captures consumer learning from both online UGC and consumption experiences. We describe our data in Section 4. In Section 5, we discuss the estimation and identification issues. We then present our estimation results in Section 6. In Section 7, we conclude with directions for future research.

3.2. Literature Review

This study fits into two main streams of literature. The first stream of literature is related to the impact of UGC on firm performance, and the second stream of literature relates to Bayesian learning in marketing and economics studies. First, we briefly review these two streams of literatures and then discuss how this study differs from and extends the literature.

3.2.1. UGC and Firms' Marketplace Performance

The popularity of UGC websites facilitates measurements of online social communications for marketers and researchers (Chen and Xie 2008; Dellarocas 2003; Mayzlin 2006). Researchers have explored the impact of UGC on firms' market performance in terms of product sales (Archak et al. 2011; Chevalier and Mayzlin 2006; Clemons et al. 2006; Dellarocas et al. 2007; Duan et al. 2008a; Duan et al. 2008b; Forman et al. 2008; Godes and Mayzlin 2004; Gu et al. 2012; Liu 2006; Moe and Trusov 2011; Sonnier et al. 2011; Zhu and Zhang 2010) or stock returns (Luo 2009; McAlister et al. 2012; Tirunillai and Tellis 2012).

Previous studies have mainly examined two metrics of UGC: volume and valence. Volume refers to the number of UGC ratings, while valence refers to the average of UGC ratings. The argument in favor of volume is that when more consumers discuss a product, other consumers will be more likely to become aware of it (Dellarocas et al. 2007). The argument in favor of valence is that word-of-mouth communication carries important information about a product's quality and may reflect the level of consumer satisfaction (Zhu and Zhang 2010). Previous studies have generally suggested that the volume is positively associated with product sales, but the relationship between the valence and product sales has been met with mixed responses (Chen et al. 2004; Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Duan et al. 2008a; Liu 2006).

Follow-up studies have also examined the moderating effects of other factors on the relationship between UGC and product sales These factors include product and consumer characteristics (Zhu and Zhang 2010) and the matching of geographical locations between consumers who write about UGC and consumers who read about UGC (Forman et al. 2008). In addition, researchers have also examined the impact of the variance of UGC ratings on product sales. Clemons et al. (2006) found that the variance of UGC ratings and the strength of the most positive quartile of UGC have a significant impact on the growth of craft beers. Sun (2012) showed that a higher standard deviation of UGC ratings on Amazon improves a book's relative sales ranking when the average rating is lower.

Regarding unstructured text messages, researchers adopted the text mining technique to extract the sentiments embedded in the messages (Das and Chen

2007). These studies extended the UGC literature by exploring the impact of text messages on firm performance (Archak et al. 2011), which provided a more comprehensive view of online communications. In addition, researchers also extended the literature by examining the influence of multiple sources of online communications (Gu et al. 2012; McAlister et al. 2012; Sonnier et al. 2011) and the dynamics between online UGC and firms' market performance (McAlister et al. 2012; Sonnier et al. 2011; Tirunillai and Tellis 2012).

This study complements the literature on UGC with regard to the two following dimensions. First, while most of previous studies have documented the relationship between UGC and firms' market performance at the aggregated level, only a small number of studies have explored how online UGC influences individual consumers' choice decisions (Albuquerque et al. 2012; Goh et al. 2013; Zhao et al. 2013). We propose a structural model to capture consumer learning from both online UGC and consumption experiences. Adopting the Bayesian updating framework, we demonstrate how individual consumers perceive and interpret the information embedded in online UGC to update their quality perceptions of products. Second, this study focuses on frequently purchased product categories involving the purchase of the same product several times. This is unlike one-time purchase products such as movies and books which usually have their unique product life cycles and follow predictable exponential patterns (Moe and Trusov 2011), There is an informational value for the consumer in consuming frequently purchased products. In this case, consumers' experiential learning would moderate the role of UGC on their future choices.

3.2.2. Quality Learning and Consumer Choice

Previous researchers have examined the relationship between information search and consumer choice behavior (Nelson 1970; Stigler 1961). Information is a valuable resource for consumers in guiding their purchase decisions. However, product quality information is usually difficult to acquire because of the intangible nature of quality, especially regarding experience goods. Thus, consumers need to learn about product quality via word-of-mouth communication, marketing communications, or their personal consumption experiences (Banerjee and Fudenberg 2004; Ellison and Fudenberg 1995; McFadden and Train 1996; Nelson 1974). In marketing literature, Erdem and Keane (1996) produced the first paper that empirically studied the effect of learning from advertising and consumption experiences. They found evidence that advertising and consumption experiences reduce uncertainty and generate significant learning. Subsequent studies extended their work by applying the notion of consumer learning in studying consideration set formation under conditions of price uncertainty and consumer search (Mehta et al. 2003), consumer learning of both service quality and usage (Iyengar et al. 2007), consumers' cross category learning (Erdem 1998), and physicians' prescription decisions (Chintagunta et al. 2009; Coscelli and Shum 2004).

There are two studies which adopt a similar approach to model how consumers learn from online UGC (Chan et al. 2012; Zhao et al. 2013). Compared with these two studies, our current study differs in the following aspects. First, we model how consumers learn product quality from both online UGC and consumers' consumption experiences and examine the interaction effects of these two learning processes. By focusing on frequently purchased products, we study how the informational role of online reviews will change when consumers can learn about the product quality through their consumption experiences. Chan et al. (2012) investigated consumer learning from online UGC, however, they failed to control for consumers' experiential learning, which is a very important aspect -for frequently purchased products (McFadden and Train 1996). Zhao et al. (2013) also modeled consumer leaning from both online UGC and consumers' consumption experiences. However, because their investigation was on books, which belong to the one-time purchase product category, they assumed that consumers learn about the average quality of a product category rather than about the focal product from their prior consumption experiences. Since the quality of a book may have no relationship with the quality of books in the same category, it is hard to argue that consumers' experiential learning can be helpful in this context. This might explain their finding that consumers learn more from online UGC than from their own consumption experiences.

3.3. The Econometric Model

3.3.1. Research Scenario and Utility Specification

Consider a situation in which consumer *i* needs to make a purchase decision from a frequently purchased product category with *j=1, 2,…, J* alternatives. We assume that consumer *i* faces uncertainty about both the intrinsic quality and price of the choice alternatives. In other words, instead of being aware of the true quality and price of alternative *j*, consumer *i* holds only *subjective beliefs* that are captured by her prior information sets.

We assume there is a UGC website where consumer *i* can search for and read the UGC of the *J* alternatives. We assume that consumers can learn about the average quality q_j' and the average price p_j of alternative *j* by reading UGC. Because consumer *i* can repeatedly purchase product *j*, he or she can also learn the average quality q_j and the price p_j of alternative *j* through his or her own consumption experiences. We further assume that both the online UGC and the consumption experiences of consumer *i*'can only provide noisy signals, which indicates that consumer *i* cannot discern the exact quality of q_j, and p_j. Consumer *i* can only make choice decisions based on his or her perceptions about q_j, and p_j. Figure 3-1 illustrates how consumers update their information sets in their decision process.

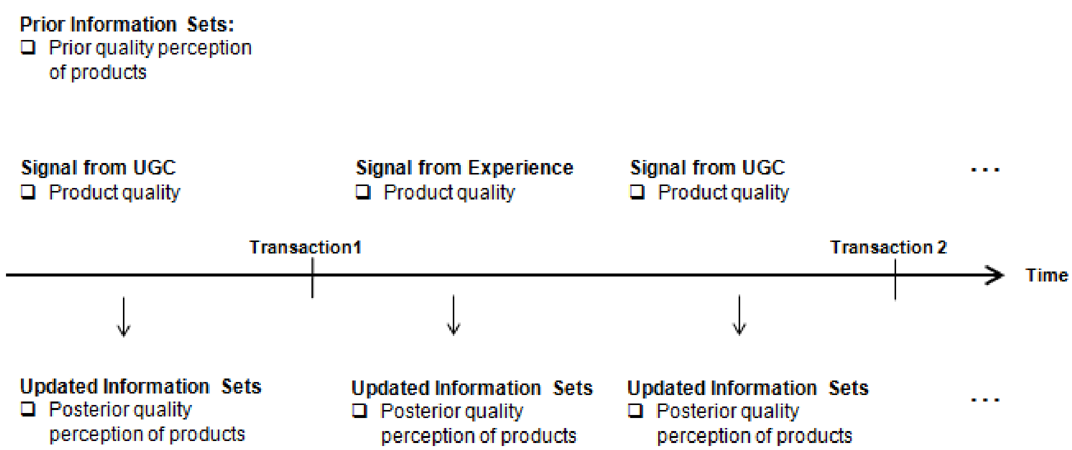


Figure 3-1: Consumers' Information Updating Process

We assume that the utility of consumer *i* from alternative *j* at purchase occasion *t* is a quadratic function of his or her quality belief $\tilde{q}_{jt}$ and cost belief $\tilde{p}_{jt}$. This function form allows for a flexible specification with respect to consumers' risk attitudes. We specify the utility function as:

$$U_{ijt} = \alpha_i \tilde{q}_{ijt} - \alpha_i r_q \tilde{q}_{ijt}^2 + \beta_i \tilde{p}_{ijt} - \beta_i r_p \tilde{p}_{ijt}^2 + \boldsymbol{\gamma} \mathbf{X_{ijt}} + \varepsilon_{ijt} \tag{1}$$

where r is the risk coefficient; consumer i is risk neutral if r is zero, and risk averse if it is positive, and risk seeking if it is negative; $\mathbf{X_{ijt}}$ is a vector of observed characteristics of alternative j when consumer i purchases alternative j at time t, which includes transportation costs, number of UGCs, and coupon promotion; α_i and β_i capture the specific preferences of consumer i for quality and price; and ε_{ijt} is the random error term that varies with consumer i, alternative j, and time t.

To capture consumer heterogeneity, we assume that intrinsic preferences for quality and price vary across consumers as follows: $\alpha_i \sim N(\overline{\alpha}, \sigma_\alpha^2)$ and $\beta_i \sim N(\overline{\beta}, \sigma_\beta^2)$. $\overline{\alpha}$ and $\overline{\beta}$ capture consumers' average sensitivity to quality and price. σ_α^2 and σ_β^2 measure the level of consumer heterogeneity in terms of quality and price preferences.

3.3.2. Quality Signals from Consumption Experience

Consumer i learns the true quality[17] of alternative j from his or her own consumption experiences based on the Bayesian updating framework. We assume that the learning is not perfect and consumption experiences only provide noisy signals of the quality of alternative j. Let E_{ijt} denote the quality signal associated with the consumption experience of consumer i regarding alternative j at purchase occasion t.

$$E_{ijt} = q_j + \xi_{ijt} \tag{2}$$

The random variable ξ_{ijt} denotes the noise associated with consumption experience.

[17] We assume consumers' cost learning follows the identical Bayesian updating structure. Hereafter we only model consumers' quality learning for convenience

We assume that ξ_{ijt} is independent and identically distributed (i.i.d.) according to a normal distribution $\xi_{ijt} \sim N(0, \sigma^2_{\xi j})$. Previous studies suggest that the error term ξ_{ijt} can be decomposed into two parts: the inherent product precision (such as some random events, or the changing of cooks) and consumers' idiosyncratic quality perceptions (Erdem and Keane 1996; Roberts and Urban 1988). Unfortunately, unless a controlled experiment is conducted, these two random components cannot be separately identified. Following the Bayesian learning literature (Erdem and Keane 1996), the variance $\sigma^2_{\xi j}$ is assumed to be common knowledge.

3.3.3. Quality Signals from Online UGC

Let R_{ijt} denote the rating of consumer *i* on alternative *j* after his or her consumption occasion *t*. We assume that R_{ijt} reflects the consumption experience E_{ijt} *of* consumer *i* (i.e., mean of R_{ijt} is same with E_{ijt}) but has a different variance with E_{ijt} because we expect that there is some information loss when consumer *i* writes down his or her consumption experience.

$$R_{ijt} = q_j + \eta_{ijt} \tag{3}$$

where η_{ijt} is independent and identically distributed (i.i.d.) according to a normal distribution $\eta_{ijt} \sim N(0, \sigma^2_{\eta j})$. Unlike the variance of experience signal, we assume that consumer *i* is uncertain of the variance of η_{ijt}. There are several uncertainties in the variance of ratings of UGC which consumers cannot tell. First, due to consumer heterogeneity, different consumers usually have their own specific preferences which affect their ratings. Consumers are uncertain about to what extent this heterogeneity will affect the variance of ratings. Second, although the website has spent effort to prevent fake reviews, it is still possible that firms may try to manipulate the online ratings. Consumers are uncertain about to what extent firms can influence online ratings. Third, the potential social dynamics of online ratings can also increase consumer uncertainty of the informational value of online UGC. Here we assume $\xi_{ijt}=\lambda\eta_{ijt}$, where λ indicates the intrinsic helpfulness of the online UGC relative to consumers' consumption experiences[18]. Thus we have $\sigma^2_{\xi j}=\lambda^2\sigma^2_{\eta j}$. If an UGC creates more noisy signals than a consumption experience, λ should have a value of less than 1.

To enhance the ease of online UGC usage, UGC websites usually aggregate UGC by counting the number of ratings and computing the average of ratings from heterogeneous reviewers. In addition, they also show the distribution of overall ratings. Consumer *i* reads the UGC of alternative *j* at purchase occasion *t* and updates his or her quality perception on alternative *j*. We assume that consumers update their quality perceptions based on the summary statistics of UGC which are aggregated by UGC websites. In other words, we assume that consumers do not rely on any specific rating but rather on the average of ratings which integrates all other consumers' evaluations. We believe this assumption is reasonable because there are usually a large number of ratings for a product and it is impossible for consumers to learn from each specific reviewer's ratings.

We assume that the average of ratings of alternative *j* at time *t* is the main quality signal:

[18] This is a very important assumption of our paper. This assumption implies that the variance of experience signals is proportional to the variance of UGC signals which can increase the tractability our model. Because we assume the variance of consumers' experience signals is common knowledge, consumers' learning of the variance of UGC signals actually reflects their beliefs on λ. This assumption can be relaxed by allowing λ to vary with choice alternatives.

$$
\begin{aligned}
&\bar{R}_{jt} = \frac{1}{N_{jt}} \sum_{t=1}^{N_{jt}} R_{ijt} = \frac{1}{N_{jt}} \sum_{t=1}^{N_{jt}} (q_j + \eta_{ijt}) = q_j + \frac{1}{N_{jt}} \sum_{t=1}^{N_{jt}} \eta_{ijt} \\
&\bar{R}_{jt} \sim N(q_j, \frac{\sigma_{\eta j}^2}{N})
\end{aligned} \tag{4}
$$

where N_{jt} is the number of ratings for alternative j at time t; $\bar{R}_{jt}$ is the average of ratings of these N_{jt} reviews for alternative j at time t.; the term $\frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \eta_{ijt}$ represents aggregated random errors. Since the number of UGC ratings will increase over time (i.e., more consumers with heterogeneous preferences will post their personal evaluations on alternative j), the term $\frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \eta_{ijt}$ will approach zero with the increase in the number of ratings. As a result, the average of ratings will become a more and more accurate signal of products' average quality.

3.3.4. Bayesian Learning from Online UGC

In the above sections, we show how we model the signals consumers receive from their experiences and online UGC. We next show how consumers adjust their product quality perceptions based on these signals using the Bayesian updating framework. As we mentioned before, we assume that consumers are uncertain about both the average product quality (q_j) and the precision of the online UGC signals ($\sigma_{\eta j}^2$) for alternative j. Online UGC provides imperfect information about the average product quality. $\sigma_{\eta j}^2$ is the variance of the signal that captures the noise that is associated with online UGC ratings.

We assume that at time 0, consumer i has initial prior opinions about the average quality (q_j) and the precision (i.e., the reciprocal of variance $\sigma_{\eta j}^2$ in statistics) of online UGC signals for alternative *j*, which jointly follow the normal scaled inverse-chi square distribution[19]:

$$\tilde{q}_j \mid \tilde{\sigma}^2_{\eta j} \sim N(q_{j0}, \tilde{\sigma}^2_{\eta j} / K_{j0}),\ \tilde{\sigma}^2_{\eta j} \sim Inv-\chi^2(\nu_{j0}, \sigma^2_{\eta j0}) \tag{5}$$

where q_{j0}, K_{j0}, ν_{j0}, and $\sigma^2_{\eta j0}$ are parameters of the prior joint distribution. After reading N_{jt} consumers' ratings of alternative *j* from the UGC website, consumer *i* updates his or her beliefs of these hyper-parameters at time *t* according to the Bayes rule (De Groot 1970) as follows:

$$\begin{aligned}
q_{jt} &= \frac{K_{j0}}{K_{j0} + N_{jt}} q_{j0} + \frac{N_{jt}}{K_{j0} + N_{jt}} \bar{R}_{jt} \\
K_{jt} &= K_{j0} + N_{jt} \\
\nu_{jt} &= \nu_{j0} + N_{jt} \\
\nu_{jt}\sigma^2_{\eta jt} &= \nu_{j0}\sigma^2_{\eta j0} + (N_{jt} - 1)S_{jt}{}^2 + \frac{K_{j0}N_{jt}}{K_{j0} + N_{jt}}(\bar{R}_{jt} - q_{j0})^2
\end{aligned} \tag{6}$$

Thus, we get the posterior of the quality information of alternative *j* as follows:

$$\begin{aligned}
&\tilde{q}_j \mid \tilde{\sigma}^2_{\eta j}, UGC_{jt} \sim N(q_{jt}, \tilde{\sigma}^2_{\eta j} / K_{jt}) \\
&\tilde{\sigma}^2_{\eta j} \mid UGC_{jt} \sim Inv-\chi^2(\nu_{jt}, \sigma^2_{\eta jt}) \\
&E(\tilde{\sigma}^2_{\eta j} \mid UGC_{jt}) = \frac{\nu_{jt}\sigma^2_{\eta jt}}{\nu_{jt} - 2} \text{ for } \nu_{jt} > 2 \\
&Var(\tilde{\sigma}^2_{\eta j} \mid UGC_{jt}) = \frac{2\nu^2_{jt}\sigma^4_{\eta jt}}{(\nu_{jt} - 2)^2(\nu_{jt} - 4)} \text{ for } \nu_{jt} > 4
\end{aligned} \tag{7}$$

The Bayesian updating process shows that the number of ratings affects the weight of consumers' quality updating from online UGC. The posterior quality perception of consumer *i* for alternative *j* is q_{jt}. As we can see from equation (5), a higher number of UGC ratings lead to more quality updating from online UGC. The posterior precision perception of online UGC of consumer *i* for alternative *j*

[19] We suppress subscript *i* here for notational convenience.

is $\frac{\nu_{jt}\sigma^2_{\eta jt}}{\nu_{jt}-2}$. Based on equation (5), we find that both the number of UGC ratings and the variance of UGC ratings play an important role in the process of updating of consumer *i*'s precision perception of online UGC signals. Consumer *i*'s perception of the UGC signal precision decreases with the variance of UGC ratings and increases with the number of UGC ratings.

3.3.5. Bayesian Learning from Consumption Experiences

Consumers' personal consumption experiences can also provide quality signals on products. However, unlike online UGC which can be searched for before a purchase decision, experience signals can only be observable after consumption. At time *t-1*, we assume that the quality perception of consumer *i* of alternative *j* is $\tilde{q}_j \sim N(q_{j,t-1}, \sigma^2_{qj,t-1})$ where $\sigma^2_{qj,t-1} = \frac{\nu_{jt-1}\sigma^2_{\eta j,t-1}}{(\nu_{jt-1}-2)K_{jt-1}}$. At time *t*, consumer *i* purchased alternative *j* and received a quality signal E_{ijt}. According to the Bayes rule (De Groot 1970), we have the posterior $\tilde{q}_j \mid E_{ijt} \sim N(q_{jt}, \sigma^2_{q_{ijt}})$, where:

$$q_{jt} = \frac{\frac{q_{ij,t-1}}{\sigma^2_{qj,t-1}} + \frac{D_{ijt} \bullet E_{ijt}}{\sigma^2_{\eta j}}}{\frac{1}{\sigma^2_{qj,t-1}} + \frac{D_{ijt}}{\sigma^2_{\eta j}}} \text{ and } \sigma^2_{qjt} = \frac{1}{\frac{1}{\sigma^2_{qj,t-1}} + \frac{D_{ijt}}{\sigma^2_{\eta j}}} \qquad (8)$$

D_{ijt} is a dummy variable and has a value of one if consumer *i* chose alternative *j* at time *t*.

3.3.6. Choice Probability

Given the utility function of consumer *i* (1), his or her expected utility associated with alternative *j* is:

$$
\begin{aligned}
E[U_{ijt} \mid E_{ijt}, UGC_{jt}] &= \alpha_i \tilde{q}_{jt} - \alpha_i r_q \tilde{q}_{jt}^2 - \alpha_i r_q Var(\tilde{q}_{jt}) + \beta_i \tilde{p}_{jt} \\
&\qquad - \beta_i r_p \tilde{p}_{jt}^2 - \beta_i r_p Var(\tilde{p}_{jt}) + \boldsymbol{\gamma}\mathbf{X_{ijt}} + \varepsilon_{ijt} \\
&= V_{ijt} + \varepsilon_{ijt}
\end{aligned} \tag{9}
$$

Assuming that the error term ε_{ijt} follows the type-I extreme value distribution, the probability of consumer *i* choosing alternative *j* at time *t* is:

$$
P_{ijt} \mid E_{ijt}, UGC_{jt} = \frac{\exp(V_{ijt})}{\sum_{k=0}^{J} \exp(V_{ijt})} \tag{10}
$$

The following are the intuitions behind the learning model. Before choosing a product, consumer *i* can search for product information of all choice alternatives from online UGC websites. The UGC helps consumers *i* to update his or her quality perceptions of all choice alternatives. The average of UGC ratings indicates the quality of alternatives and is used to update consumer *i*'s quality perception. The number of UGC ratings affects the extent to which consumer will update their product quality perceptions based on the average of UGC ratings. The number of UGC ratings, the variance of UGC ratings, and the average of UGC ratings play important roles in the updating process of consumer *i*'s precision perception. Consumer *i*'s precision perception of the UGC signal decreases with variance of UGC ratings and the difference between the average of UGC ratings and consumers' prior quality perception, but however, this perception increases with the number of UGC ratings. Consumer *i* chooses an alternative *j* based on his or her quality perceptions of product alternatives. After his or her consumption, consumer *i* receives a noisy quality signal of product *j*. Consumer *i* consequently updates his or her quality perception of alternative *j*.

3.4. Data Description

The data for this study is obtained from Dianping.com, a popular user generated reviews site in China. Dianping covers over 2300 cities in China, with more than 1 million businesses featured on its website. It provides reviews for consumer-service oriented businesses, such as restaurants, shopping, beauty and cosmetics products, hotels, sports activities, car services, life services and so on. Our study focuses on consumers' restaurant visit decisions. Restaurant choice is a suitable context for our study for several reasons. First, restaurant choice is categorized as experience goods and the quality cannot be inspected before consumption. Thus, other consumers' evaluation of a restaurant can provide valuable information for the focal consumer. Second, restaurants can be frequently patronized by a consumer. Consumers can learn the quality of the restaurants from both online UGC and their own consumption experiences. Consumers' experiential learning may moderate the informational role of online UGC and the extent to which consumers rely on UGC to guide their purchase decisions. Therefore, we investigate how consumers learn restaurant quality from online UGC and how their experiential learning moderates the information role of online UGC in this research.

Our dataset focuses on a major city in China and is composed of three sections: online reviews of restaurants, information on restaurant attributes, and consumers' restaurant dining records. The overall timeline of our data set - stretched from 2003 to 2008. The information on restaurant reviews was collected from April 2003 to March 2008. Consumers' dining transaction records were collected from May 2005 to March 2008. Detailed information of restaurant promotions was collected from January 2006 to March 2008.

Table 3-1: Consumer Switching Frequency

Variables	Mean	S.D.
SwitchRest[20]	0.60	0.49
SwitchCuisine	0.50	0.50
SwitchLocation	0.44	0.50
SwitchPrice	0.37	0.48
Number of Trips		304109

Information on restaurant reviews includes consumers' ratings in terms of overall quality, tastiness of the food, restaurant ambience, and service quality. The ratings are measured on a scale of 0 to 4, with 0 being 'very bad', and 4 being 'very good'. In addition, reviewers could post information of the estimated average cost per person, recommended dishes, and detailed qualitative comments on each restaurant. Restaurants are classified according to review sites in terms of geographical locations, price levels, and cuisine types. The restaurants are located in 11 districts of the sample city, graded according to 5 different price levels, and categorized into 17 cuisine types. Other information on the restaurants includes the availability of coupon promotions and whether restaurants have bought keywords for search advertising in order to get a more prominent sponsor link.

Table 3-2: Consumer Switching Patterns

	Total	SwitchCuisine	SwitchLocation	SwitchPrice
SwitchRest	182,295	150,495	132,382	111,847
		82.6%	72.6%	61.4%

Consumers' restaurant dining records were gathered using the review website's loyalty program member cards. The website distributes loyalty member

[20] *SwitchRest* is the dummy variable that indicates whether consumers switch to another restaurant. It is 1 when consumers do switch and 0 otherwise. Other variables are defined in the similar way. *SwitchCuisine* indicates whether consumers switch to a new cuisine type. *SwitchLlocation* indicates whether consumers switch to a new geographical location. *SwitchPrice* indicates whether consumers switch to a new price level.

cards to their registered customers. When consumers patronize a restaurant which has a joint partnership program with the review site, they accumulate membership points and receive a discount by using the loyalty member card at each visit. Our data on consumer transactions are thus sourced from consumers who are members of the review site's loyalty program. Such data on dining transactions, while keeping consumers' identity anonymous, includes each restaurant's name, as well as consumers' dining expenditure, and transaction dates.

Table 3-3: Summary Statistics for Restaurant Reviews

Rest. No.	Location	Cuisine	Volume	Taste		Ambience		Service		Price	
				mean	*s.d.*	*mean*	*s.d.*	*mean*	*s.d.*	*mean*	*s.d.*
1	babaiban	shanghai	217	1.82	0.77	1.66	0.73	1.59	0.77	79.42	42.84
2	babaiban	shanghai	234	1.73	0.78	2.06	0.84	1.57	0.83	146.49	78.55
3	babaiban	hunan	976	2.2	0.87	1.51	0.69	1.36	0.74	56.97	20.43
4	lujiazui	shanghai	1113	1.91	0.81	2.31	0.9	1.76	0.86	109.62	55.53
5	lujiazui	japanese	258	1.7	0.8	1.36	0.62	1.4	0.76	86.02	35.93
6	lujiazui	western	185	1.93	0.77	2.09	0.84	1.89	0.85	134.14	85.63
7	babaiban	shanghai	760	1.74	0.8	1.88	0.8	1.61	0.84	77.35	32.49
8	yuanshen	hunan	121	1.93	0.93	1.44	0.74	1.37	0.81	55.04	22.99
9	babaiban	western	824	2.18	0.94	2.47	0.93	2.62	1.08	209.28	45.91
10	babaiban	sichuan	436	1.81	0.88	1.46	0.68	1.49	0.85	71.62	26.02
11	babaiban	hotpot	324	2.02	0.93	2.26	0.77	1.99	0.84	90.37	32.32
12	lujiazui	shanghai	303	1.78	0.81	1.94	0.78	1.83	0.94	81.93	70.92
13	babaiban	sichuan	745	1.92	0.85	2.27	0.79	1.9	0.91	102.92	49.38
14	babaiban	sichuan	807	1.69	0.77	1.75	0.75	1.42	0.78	62.08	21.83
15	lujiazui	sichuan	995	1.89	0.85	2.22	0.8	1.55	0.84	88.49	101.23
16	tangqiao	hunan	243	1.77	0.87	1.46	0.71	1.16	0.8	51.15	19.91
17	lujiazui	japanese	215	1.54	0.87	1.6	0.74	1.51	0.87	112.97	54.49
18	jinqiao	shanghai	122	1.79	0.7	2.07	0.71	1.56	0.76	84.91	43.1
19	babaiban	hotpot	1621	2.19	0.79	2.53	0.84	2.1	0.94	84.58	77.29

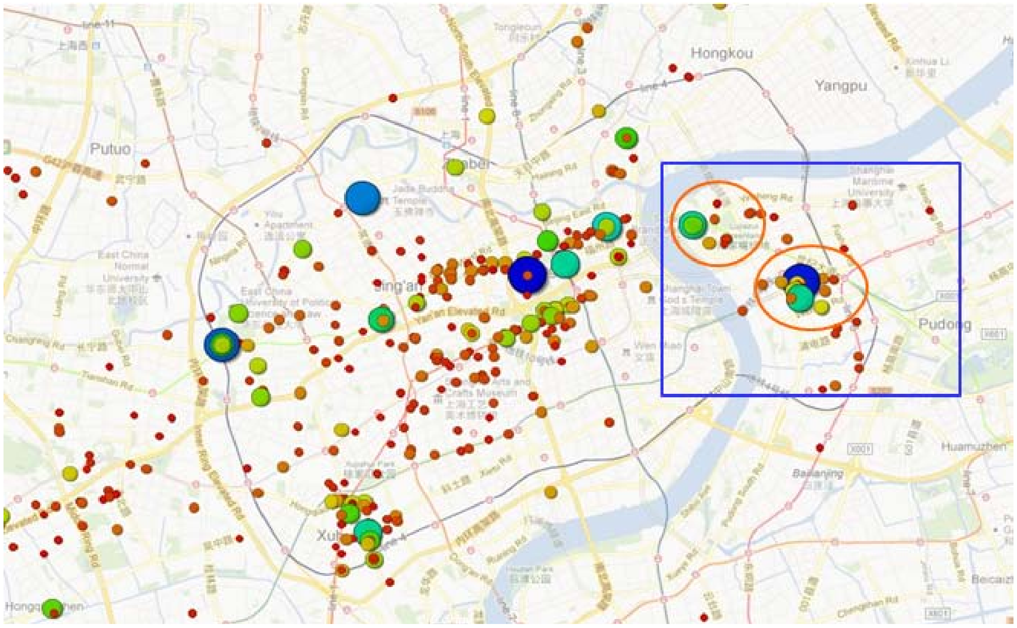


Figure 3-2: The Geographical Locations of the Restaurants Included in Our Data
Note. The color and size of each dot represent the revenue of individual restaurants.

Our data set include transactions of 439 restaurants. Researchers have pointed out that consumers may not consider all available choice alternatives in their decision process (Shocker et al. 1991). Previous studies have also showed that consideration sets have an important influence on consumer choice (Andrews and Srinivasan 1995; Roberts and Lattin 1991). In addition, when the choice set faced by an individual becomes very large, computational limitation makes estimation with the full choice set intractable (Bordley 2013; Lemp and Kockelman 2012). Thus, to study how consumers learn about restaurant quality from online UGC to guide their purchase decisions, we narrowed down consumers' choice set by focusing on restaurants located in one specific district of the city[21]. Figure 3-2 shows the locations of our sample restaurants. The chosen district (i.e., the PuDong district) is separated from other districts by a river. We

[21] The alternative method we used to define consumer choice set was to sample restaurants by focusing on one specific cuisine type: sichuan cuisine. For this alternative sample, we included sichuan cuisine restaurants from multiple geographical locations.

believe this criterion is valid for two reasons. First, the restaurant industry is horizontally differentiated by location and restaurant managers usually strategically choose their restaurants' locations to segment the market (Auty 1992). Second, researchers have suggested that in the consumers' hierarchical choice process, consumers' choice-switching behavior will manifest the competition structure of the market (Grover and Dillon 1985). The inter-partition switching is lowest at the highest level in the hierarchy and highest at the lowest level. Table 3-1 shows consumers' switching frequencies on their purchase occasions. Table 3-2 demonstrates the switching patterns of consumer choice in our entire transaction data set. As it can be seen from Table 3-1, beyond restaurant switching, consumers switch the most in terms of cuisine type, followed by location and price level in our data set. This trend is further elaborated in Table 3-2. We found that consumers' restaurant switches were mainly accompanied by switches in cuisine types, followed by switches in location, and price levels.

Because restaurants in our sample might join the partnership program of the review site at different times, we needed to ensure that all the restaurants in our final sample had joined the partnership program. Thus, we sampled the restaurants whose first transactions occurred before 1 Jan, 2007 and whose last transactions transpired after 1 Jan, 2008. In addition, we also excluded restaurants whose cuisine types were not comparable to others categories, such as "bread, dessert, simple snacks". The final sample consisted of 19 restaurants. Table 3-3 provides further information on these restaurants.

3.5. Estimation

Our data set included observations of other consumers' quality evaluations of restaurants (i.e., user-generated restaurant reviews), which could be used to model consumer quality learning. Instead of modeling consumer learning from each specific review (i.e., by weighting reviews differently according to the characteristics of the reviews and reviewers), we assumed that consumers updated their quality perceptions by using the summary statistics of reviews (number of ratings, average ratings, and variance of ratings).

However, a significant problem we encountered in the estimation process was that consumer i observed the signal E_{ijt} while we as econometricians usually did not. Our product review data provided only a small number of observations of consumers' ratings for the chosen restaurants after their consumption trips. Regarding those trips, as we failed to observe consumers' ratings for the chosen restaurants, we instead simulated consumers' experience signals by following the literature on learning (Erdem 1998; Erdem and Keane 1996).

Compared with prior learning studies, we had data of restaurant reviews which were the ratings of other consumers' consumption experiences. The review data helped us in simulating consumers' experience signals. Our model assumed that UGC ratings reflect consumers' consumption experiences and that the variance of UGC ratings is proportional to the variance of consumers' evaluations of their consumption experiences (λ is the proportional constant). We supposed that consumer i updated perceptions about the mean and variance of the UGC signals of restaurant j were E_j and V_j. Thus, we drew consumers' experience signals from a distribution with the mean E_j and variance λV_j. We stress that consumer *i* updated his or her perceptions about the precision of experience signals by adopting the updated distribution of UGC signals.

We next discuss the identification of the parameters in our model. Since we observed UGC signals and simulated consumers' experience signals based on UGC data, we were able to make inferences on a product's true quality (or price) based on the average of UGC ratings. The number of UGC ratings, the average of UGC ratings, and the observed choice jointly helped us to identify the quality (or price) weight. The risk coefficient for quality (or price) measures the consumer's sensitivity to quality (or price) uncertainty. As shown in Equation (6), this uncertainty depends on the number of the consumer's prior trips to a specific restaurant, the number of UGC ratings, and the variance of UGC ratings for the restaurant. Thus, the risk coefficient for quality (or price) is identified by the variance of consumers' prior trips and these UGC metrics. Given our panel data, we could easily identify consumers' unobserved heterogeneity after we identified the quality (or price) weight and the risk coefficient.

3.6. Results and Managerial Implications

3.6.1. Reduced-Form Analysis

We first showed the relationship between online UGC data and consumers' restaurant choices by using the multinomial logit model. We specified random coefficients to capture consumers' heterogeneity in terms of their sensitivities to online UGC metrics and price.

Table 3-4: Summary Statistic for Restaurant Reviews

Variables	**Pudong District Sample**				**Sichuan Cuisine Sample**			
	Model (1)		Model (2)		Model (3)		Model (4)	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Number of ratings	0.13^{***}		0.11^{***}		0.35^{***}		0.23^{***}	
Taste	3.15^{***}	4.06^{***}			-0.9^{***}	4.93^{***}		
Ambience	5.52^{***}	3.64^{***}			3.03^{***}	2.46^{***}		
Service	-4.5^{***}	1.36^{***}			-1.4^{***}	0.89^{***}		
Variance of Taste	-0.02	6.99^{***}			-3.6^{***}	3.35^{***}		
Variance of Ambience	-6.4^{***}	5.84^{***}			1.22^{***}	0.08		
Variance of Service	6.24^{***}	5.96^{***}			2.41^{***}	1.36^{***}		
Overall Quality			6.11^{***}	5.09^{***}			2.53^{***}	1.88^{***}
Variance of Overall Quality			-2.8^{***}	5.09^{***}			-0.8^{***}	2.23^{***}
Average Price	-2.2^{***}	2.92^{***}	-4.8^{***}	6.65^{***}	-4.1^{***}	10.6^{***}	-3.0^{***}	9.7^{***}
Variance of Price	-1.3^{***}	1.61^{***}	-0.02	1.86^{***}	-1.0^{***}	1.1^{***}	-3.5^{***}	3.6^{***}
N	96235		96235		178668		178668	
AIC	15154.5		16413.0		24765.1		29450.2	
BIC	15315.5		16498.3		24936.7		29541.1	
Log lik.	-7560.2		-8197.5		-12365.6		-14716.1	
Chi-squared	4916.6		4512.2		6258.5		3555.6	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3-4 records the reduced-form estimation results. Models 1 and 2 are estimation results based on sampled restaurants in the Pudong district. According to the estimation for the means of the coefficients in Model 1, we obtained the following results: (1) the number of UGC ratings has significant positive effects on consumer choice probability which suggests that consumers on average prefer a popular restaurant; (2) both taste rating and ambience rating on average have significant positive effects on consumer choice, indicating that consumers are more likely choose a restaurant with a higher rating in terms of taste and ambience; (3) the mean of price coefficient is negative and significant, implying that consumers on average are more likely to choose a low priced restaurant; (4) service rating has a surprisingly significant negative effect; and (5) the coefficients of the variances in restaurant ratings suggest consumers' risk attitudes. As can be seen, the coefficient for variances of taste ratings is negative and insignificant. The coefficient for variance of ambience ratings is negative and significant. The coefficient for variance of service ratings is positive and significant. The coefficient for variance of price is negative and significant. These results imply that consumers are risk-averse to taste, ambience, and price but risk-seeking with regard to service. The estimation for standard deviations of the coefficients in Model 1 captures consumers' heterogeneity in terms of sensitivity to online UGC information. As can be seen, all the coefficients are statistically significant and the magnitude is large, which indicate that consumers show high levels of heterogeneity in terms of their responses to online UGC.

In Model 2, we estimated our model by using restaurants' overall quality ratings instead of ratings for the three attributes of taste, ambience, and service. The mean of coefficient for overall quality is positive and significant, indicating that the overall quality has a significant positive effect on consumer restaurant choice. The mean of coefficient for variance of overall quality is negative and significant, implying that consumers are risk-averse to quality uncertainty. Models 3 and 4 show the estimation results based on the restaurants of the "Sichuan Cuisine" category. The results are generally consistent with Models 1 and 2, which show the robustness of our definition of consumer choice sets.

3.6.2. Structural Model Results

Tables 3-5 and 3-6 show our estimation results from the proposed structural learning model (Model 3) and two comparative models (Models 1 and 2). In Model 1, we assumed that consumers rely on UGC to guide their purchase decisions but do not incorporate a learning structure. In Model 2, we assumed that consumers make use of UGC to update their quality perceptions of choice alternatives following the Bayesian updating framework. Comparing these three models will help us to understand the importance of including the learning structure and controlling for consumers' experiential learning.

According to the model fit statistics, including the log-likelihood function value, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), we found that: (1) Model 2 fits the data better than Model 1, suggesting that it is important to add the learning structure to examine the influence of online UGC; and (2) Model 3 fits the data much better than Model 2, implying that incorporating consumer experiential learning can significantly improve the explanation power of our model. Furthermore, there is a greater improvement in model fit from Model 2 to Model 3 than from Model 1 to Model 2. This implies that consumers learn more from their own consumption experiences than from online UGC information.

Table 3-5: Multinomial Logit Model Results

Model 1: Without Learning		
	Estimates	*Std. Err.*
	Mean of coefficients	
Number of ratings	0.15^{**}	0.008
Quality	3.54^{**}	0.187
Variance of quality	-4.42^{**}	0.532
Price	-3.59^{**}	0.176
Variance of price	0.18^{**}	0.065
	Standard deviation of coefficients	
Quality	1.52^{**}	0.205
Variance of quality	5.45^{**}	0.497
Price	3.85^{**}	0.194
Variance of price	0.05	0.183
Number of observations	4724	
Number of consumers	539	
Log-likelihood	-7960.5	
AIC	15939	
BIC	15954	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3-6: Estimation Results from Structural Learning Models

	Model 2: Learning from UGC		Model 3: Learning from UGC and Exp.	
	Estimates	*Std. Err.*	*Estimates*	*Std. Err.*
	Mean of coefficients			
Number of ratings	0.15**	0.008	0.19**	0.007
Quality (α)	18.18**	1.776	1.79**	0.152
Risk factor for quality (r_q)	0.20**	0.005	0.04**	0.003
Price (β)	-5.34**	0.201	-0.81**	0.086
Risk factor for price (r_p)	0.18**	0.060	0.49**	0.045
Helpfulness of UGC (λ)	-	-	0.12**	0.038
	Standard deviation of coefficients			
Quality (α)	0.57**	0.154	2.53**	0.191
Risk factor for quality (r_q)	11.60**	3.388	0.06**	0.010
Price (β)	3.99**	0.213	0.80**	0.123
Risk factor for price (r_p)	0.50**	0.103	0.11	0.061
Number of observations	4724		4724	
Number of consumers	539		539	
Log-likelihood	-7703.5		-6144.9	
AIC	15425		12310	
BIC	15440		12327	

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

In terms of parameter estimation, we found that most of the coefficients for UGC ratings are significant, suggesting that online UGC plays an important role in the consumer decision making process. The estimated parameters are qualitatively consistent in our three models. Model 3 has one more parameter λ than Models 1 and 2. This parameter captures the helpfulness of online UGC relative to consumers' consumption experiences. Our estimated value of λ is 0.12 (SE=0.038) and statistically significant. The fact that λ is less than 1 suggests that online UGC provides more noisy signals than consumers' own consumption experiences. Furthermore, the low value of λ (relative to 1) implies that the product quality information from one consumption experience is comparable with about 8 UGC ratings.

The coefficient for the number of UGC ratings is positive and significant in three models. This result suggests that consumers are more likely to purchase a product with a greater number of ratings. The mean of the coefficient for quality is positive and significant in three models. However, the magnitude varies substantially. Specifically, we find that the parameter estimates from Models 1 and 2 which ignore learning from consumers' own consumption experiences are biased upwards. The standard deviation (SD) of the coefficient for quality is statistically significant and large (relative to the mean of the coefficient) in terms of magnitude. This suggests that consumers are heterogeneous in terms of their sensitivity to perceived product quality. The coefficient for variance of quality in Model 1 is negative and significant, implying that consumers are risk averse and thus less likely to choose a product when other consumers' ratings on the product are heterogeneous. This is consistent with the estimates for risk attitude for quality in Models 2 and 3. The mean of risk parameter r_q is positive and significant in both Model 2 and Model 3, suggesting that consumers are risk averse to quality uncertainty. In addition, the SD of risk parameter r_q is significant which implies that consumers are heterogeneous in their risk aversion to quality.

The mean of price coefficient is negative and significant in all three models, suggesting that consumers are less likely to purchase a product with a higher price. The SD of price coefficient is significant and large relative to the mean, suggesting that consumers are heterogeneous in their price sensitivities. In terms of consumer risk attitude to price, we found different results from Model 1 and Models 2 and 3. The coefficient for variance of price in Model 1 is positive and significant, implying that consumers are more likely to choose a product when other consumers' estimated price differs. In Models 2 and 3, the mean of risk parameter r_p is positive and significant, suggesting that consumers are risk seeking regarding price uncertainty. The SD of risk parameter r_p is not significant in Model 3, which implies that consumers are homogeneous in their risk attitudes to price.

In summary, we present the following findings of this study. First, consumers can learn product quality from both online UGC information and their own consumption experiences for frequently purchased products. There is a significant amount of consumer learning from consumers' own consumption experiences, much more than from online UGC. Second, it is important to control for consumers' experiential learning when examining the impact of online UGC on consumer choice regarding frequently purchased products. Failure to do so may result in over-estimation of the impact of online UGC.

3.6.3. Counterfactual Simulation

Our development of a structural model has the advantage of enabling us to conduct policy simulation to examine the impact of firms' marketing policies and reviewer behaviors on consumer choice and the market share of a product. We presented one counterfactual simulation here to demonstrate the managerial relevance of this study. Previous studies have demonstrated that online UGC may not be able to reflect the true quality of a product due to firms' manipulation (Dellarocas 2006; Mayzlin 2006), consumers' self-selection (Li and Hitt 2008), social dynamics (Godes and Silva 2012), and price effect (Li and Hitt 2010). The potential bias of online UGC has crucial implications for a firm's profits, pricing strategy, and consumer surplus (Li and Hitt 2008; Li and Hitt 2010; Moe and Trusov 2011). Researchers have also pointed out that there could be substantial value in review systems investing to prevent and eliminate this bias (Li and Hitt 2008). Since consumers can learn product quality through their own consumption experiences in the frequently purchased product category, we expect that the impact of online UGC on consumer choice will change with changes in consumers' consumption experiences. In other words, the potential bias of online UGC will be alleviated by consumers' experiential learning.

The objective of this counterfactual simulation is to examine how and to what extent consumers' experiential learning can affect the information role of online UGC. Specifically, if a self-selection bias exists in consumers' UGC contributions, we investigated to discern whether consumers' experiential learning can moderate their sensitivity to online UGC. We assumed that two scenarios existed. In the first scenario, we assumed that all restaurants' average ratings were exactly equal to their true qualities. In the second scenario, we assumed there was a representative restaurant *j*[22] whose average UGC rating was higher than its true quality while the average ratings of other restaurants remained equal to their true qualities. We further assumed that there was a representative consumer *i* for whom we set his or her responses to online UGC at the mean level of our estimations. We calculated the difference of the probability of consumer *i* of choosing restaurant *j* between the first scenario and the second scenario and examined how the different changes came about with the consumption trips of consumer *i* in restaurant *j*.

[22] We specifically choose the restaurant whose average rating is the median of all restaurants in our sample as the representative one.

Table 3-7 presents the results of the simulations after we set the average UGC rating of restaurant *j* one unit (or half unit) higher than *its* true quality. We present our key findings. First, we found that the probability of consumer *i* choosing restaurant *j* increased with the increase the average UGC ratings. This suggests that firms can benefit from higher UGC ratings. Second, the increase in the choice probability of consumer *i* in choosing restaurant *j* decreased with the number of consumption trips of consumer *i* to restaurant *j*. This suggests that when consumer *i* was able to learn about the quality of restaurant *j* through his or her personal consumption experiences, the impact of UGC ratings was found to decrease. This result is consistent with our expectation that we would generally over-estimate the impact of online UGC if we ignored consumers' experiential learning. We stress here that firms should be more cautious when promoting frequently purchased products via online UGC websites. Since consumers' experiential learning plays a significant role, online UGC promotions may be influential only for new products and the impact can only be of a short duration. Third, we further found that the level of over-estimation first increased and then decreased over the number of trips to restaurant *j* made by consumer *i*. This result is consistent with the mechanism behind Bayesian updating because the experiential learning of consumer *i* could both adjust the quality perception of consumer *i* on restaurant *j* and reduce the quality uncertainty of consumer *i* on restaurant *j*. Because consumer *i* is risk averse, lower quality uncertainty on restaurant *j* can increase the probability of consumer *i* choosing restaurant *j*.

Table 3-7: Simulation Results

Increase of UGC rating	Number of consumption experiences	Increase of consumer choice probability	Level of Over-estimation
ΔRating = 1	N=0	Δ7.53%	
ΔRating = 1	N=1	Δ6.85%	0.68%
ΔRating = 1	N=2	Δ6.45%	1.08%
ΔRating = 1	N=5	Δ6.23%	1.3%
ΔRating = 1	N=10	Δ6.53%	1%
ΔRating = 1	N=100	Δ6.78%	0.75%
ΔRating = 0.5	N=0	Δ3.47%	
ΔRating = 0.5	N=1	Δ3.29%	0.18%
ΔRating = 0.5	N=2	Δ3.17%	0.3%
ΔRating = 0.5	N=5	Δ3.28%	0.19%
ΔRating = 0.5	N=10	Δ3.44%	0.03%
ΔRating = 0.5	N=100	Δ3.47%	0%

3.7. Conclusion

Researchers have examined the impact of UGC at both the aggregated product level (Chevalier and Mayzlin 2006; McAlister et al. 2012; Sonnier et al. 2011; Tirunillai and Tellis 2012) and the individual consumer level (Albuquerque et al. 2012; Goh et al. 2013; Zhao et al. 2013). This study extends the UGC literature by investigating how consumers' experiential leaning can moderate the informational role of online UGC for frequently purchased products. In the case of frequently purchased products, there is informational value for the consumer to purchase a product because that consumer can continue to buy the same product if he or she likes it. We propose a structural learning model to capture how consumers learn product quality from both online UGC and their own consumption experiences. Our model assumes that consumers learn about both the average product quality and the precision of UGC signals. We apply our model to the context of consumer dining choice by combining data from online reviews of restaurants and consumers' restaurant dining records.

Our analysis leads to two important findings. First, consumers can learn about restaurant quality from both online UGC and their own consumption experiences in dining choice. There is a significant amount of consumer learning from consumers' own consumption experiences, much more than from online UGC. Second, neglecting consumers' experiential learning can lead to over-estimation of the impact of online UGC on consumer restaurant choice. We demonstrate how our model can be used for firms' decisions on word-of-mouth marketing. Our policy simulation results suggest that the impact of online UGC on consumer decision decreases with the number of consumers' consumption trips. Thus, online UGC promotions may be influential only for new products and the impact can only be of short duration.

This study has several limitations which suggest opportunities for future research. First, our model assumes that consumers search for and read online UGC before making purchase decisions. However, this may not be true. Future research can extend this study by incorporating consumers' information search decisions. Second, our model assumes that consumers are myopic and maximize their utility on each purchase occasion. It will be interesting to study consumers' forward-looking behaviors when examining the informational role of online UGC. Third, our model only captures consumers' experiential learning on product quality. A possible extension is to model consumers' experiential learning on the helpfulness of a UGC site. It will be interesting to study how the helpfulness of a UGC site affects an individual consumer's site visits and information search. Fourth, as in most previous studies which investigate the influence of online

UGC, the influence of alternative unobserved sources of information, such as information from other UGC sites and offline word-of-mouth communication, cannot be ruled out.

CHAPTER 4. STUDY THREE

EXAMING THE TIMING EFFECT OF INFORMATION DIFFUSION ON SOCIAL MEDIA PLATFORMS: A TEMPORAL NETWORK APPROACH

4.1. Introduction

Word of mouth has long been recognized as an important mechanism by which information can reach large populations (Bass 1969; Katz and Lazarsfeld 1955; Rogers 1995). The widespread adoptions of social media platforms (hereafter referred to as SMP) provide great opportunities for practitioners and researchers to collect data on information diffusion in social networks. Sharing online content is an integral part of modern life in digital domains. Consumers talk about new running shoes, complain about bad hotel stays, and share information about which restaurant to patronize in diverse SMP such as blogs, micro-blogs, wikis, social bookmarking and social network sites. Allsop et al. (2007) find that 59% of people report that they frequently share online content with others. According to Harris (2010), someone tweets a link to a *New York Times* story once every four seconds.

Although such social transmissions have important impacts on consumers and brands, less is known about why certain pieces of online content are more popular than others. Research shows that online users’ attention is allocated in a rather asymmetric way. Most online content gets only some views or shares, whereas a few others receive the most attention and spread widely throughout the blogosphere (Szabo and Huberman 2010). Companies often create online marketing campaigns or encourage consumer-generated content on SMP in the hope that people will share this content with others. However, some of these efforts succeed while others fail. For content providers, web hosts, and advertisers, an important question is: how and why do the popular contents online get popular?

Previous studies have investigated this question from diverse perspectives. Traditional diffusion models conceptualize the diffusion process as being determined by the effects of innovation and imitation and ignore connection patterns between individuals (Bass 1969; Mahajan et al. 1990). Recent studies explicitly incorporate the interpersonal connections when examining word of mouth diffusion processes (Iyengar et al. 2011). Researchers show evidence of social contagion (or peer effects) in diverse contexts (Bandiera and Rasul 2006; Hill et al. 2006; Iyengar et al. 2011; Katona et al. 2011). Prior research also examines the role of local and global network structure of opinion leaders (Iyengar et al. 2011; Katona et al. 2011; Moynihan 2008; Nair et al. 2010; Yoganarasimhan 2012) and the characteristics of information content (Berger and Milkman 2012; Berger and Schwartz 2011; Berger et al. 2010; Zhang and Moe 2012) in the diffusion process. Surprisingly, these studies usually assume the information network is static and neglect the impact of human activity patterns across time in the information diffusion process (Iribarren and Moro 2009). In the social media context, we posit that a user's usage pattern matters in two ways. First, users are heterogeneous in terms of when they are the most "active" on SMP (Warren 2010). We define a user as being "active" when the user logs in and generates or shares online content with other users. On a daily basis, users may be more active in some time periods than in other time periods on average and different users may have different active time periods. Thus, to increase the probability of success for social media marketing, advertising or public relations campaigns, it is important for firms to understand when consumers are likely to be online or active. Second, user interactions on SMP follow a temporal order. For example, suppose that user B follows or is a friend of user A, while user C follows or is a friend of user B on a SMP. Then, information from A cannot reach C if the communication between B and C happens before the communication between A and B. Thus, the temporal dimension should be taken into account when examining information diffusion on SMP (Lee et al. 2010; Tang et al. 2010a).

To address these research gaps, we examine the temporal effects of information diffusion on SMP by investigating the role of users' active time and temporal order of information transmission. To be specific, we empirically examine the following research questions:

(1) How does the temporal heterogeneity in users' active time periods on a SMP affect the extent of social contagion or spread of information at the dyadic level?

(2) How does the temporal order of information transmission affect the popularity of online content on a SMP?

We adopt the temporal networks modeling approach to investigate our research questions. Compared to the static networks approach in most prior research, the temporal networks approach incorporates information about when things happen from the dynamical system to the network (Holme and Saramäki 2011). This approach can help us to examine the effects of users' active time periods and the temporal order of information diffusion on the popularity of online content on SMP. In addition, since the incidence of information overload on SMP may lead to a dearth of user attention for each specific information content (Falkinger 2008; Zandt 2004), we also examine the moderating role of information overload in the diffusion process.

Using data from a popular micro-blog website, we empirically tested our hypotheses. We found that: (1) users are more likely to share a content which is more recent relative to their active time periods; (2) the more followings a user has, the less likely he or she will re-post his or her followings' posts; (3) a content generator's temporal reachablity has a significant positive impact on the popularity of the content; (4) a content is more likely to get popular when the conent is generated during the active time periods of a SMP.

Our study here generates the following contributions. First, to the best of our knowledge, this is the first study to examine the temporal effect of information diffusion on SMP in the information systems field. We provide new insights in terms of how users' active time periods may affect the social contagion or spread of information at the dyadic level, how the temporal order of information diffusion may affect the popularity of online content and the role of opinion leaders in information transmission. Second, our results would potentially shed light on how to design and implement more effective and successful viral marketing, advertising or public relations campaigns. We aim to provide insights in terms of when the most effective time slots to encourage the generation and propagation of contents are, and who can be the likely candidates for opinion leaders on social media at different time periods.

4.2. Related Literature

A large body of research in marketing has examined the diffusion of new products (Mahajan et al. 1990; Meade and Islam 2006). Traditional diffusion models conceptualize the diffusion process as being determined by two effects: innovation and imitation (Bass 1969). Researchers have extended the Bass framework to consider the diffusion across multiple consumer segments, each with its own unique adoption behavior (Garber et al. 2004; Gatignon et al. 1989; Kumar and Krishnan 2002; Putsis et al. 1997; Van den Bulte and Joshi 2007). These studies show that adoption rates can vary dramatically across markets and consumer segments. Although the Bass model and its generalizations are popular in practice and research fields, these models assume that every consumer is connected with every other consumer, and estimate a uniform interpersonal influence on this fully connected network. Given the central role of social communication in diffusion processes, it is therefore crucial to incorporate the fine-grained structure of interpersonal connections in diffusion models (Mahajan et al. 1990).

Various technological innovations in recent years have made it possible to collect data on interpersonal relationships and communications between consumers. Researchers have developed models to investigate the impact of network characteristics in the diffusion process in empirical studies. Most recent studies focused on detecting the existence of peer effects (or social contagion) (Aral and Walker 2011; Bapna and Umyarov 2011) and on identifying the opinion leaders (Yoganarasimhan 2012). Studies that examined peer effects seek to understand whether friendship ties affect consumers' choices (Iyengar et al. 2011). The critical challenge to identify peer effects is endogeneity problems such as endogenous group formation and peers' exposure to similar unobserved environmental factors (Hartmann et al. 2008). Researchers have proposed new methods to address these endogeneity problems (Bramoulle et al. 2009; Brock and Durlauf 2007) and some recent studies have found evidence of peer effects. For example, Hill, et al. (2006) used telecommunication data to show that customers who communicated with a customer of a particular service have an increased likelihood of adopting that service. Bandiera and Rasul (2006) documented how farmers' decisions to adopt a new crop relate to the adoption choices of their family and friends.

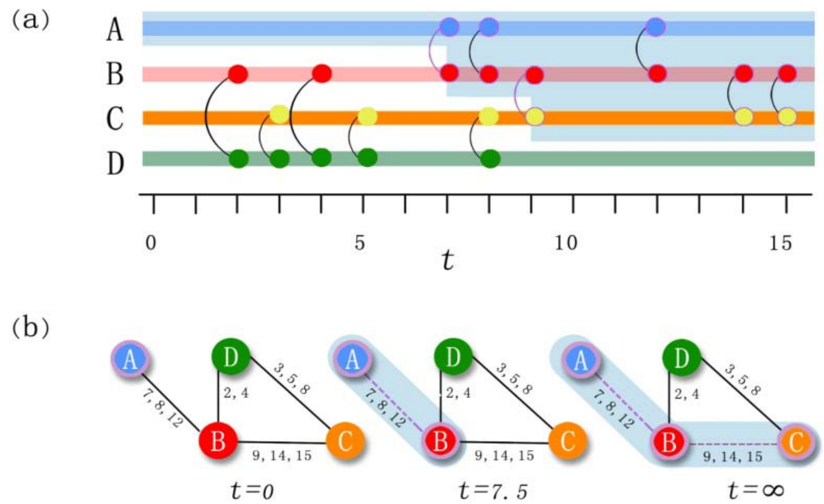
The literature on opinion leaders defines opinion leaders as a small minority that exerts a strong influence on the opinions and decisions of the majority (Iyengar et al. 2011; Katz and Lazarsfeld 1955). The idea of opinion leaders is attractive to marketing managers because of firms' strategy to increase word of mouth transmission through these opinion leaders (Goldenberg et al. 2009; Zhang and Moe 2012). Some empirical evidences have documented the influence of opinion leaders and provide guidance on how to identify opinion leaders. For example, Nair et al. (2010) studied physician prescription behavior and found that opinion leaders in a physician's reference group may have a significant influence on the physician's prescription behavior. Iyengar et al. (2011) found evidence of contagion operating over network ties after controlling for marketing efforts and arbitrary system wide changes. Goldenberg et al. (2009) and Katona et al. (2011) recommended that individuals with a certain network structure such as high number or density of connections are influential in diffusion. In the context of YouTube, Yoganarasimhan (2012) and Susarla et al. (2012) empirically demonstrated that the size and structure of an author's local network are significant drivers of the popularity of videos seeded by a user, even after controlling for observed and unobserved video characteristics, unobserved author characteristics, and endogenous network formation.

Previous research also examined the role of content characteristics on word of mouth transmission (Berger and Milkman 2012; Berger and Schwartz 2011; Zhang and Moe 2012). For example, online contents need to be surprising, interesting, or practically useful to be talked about (Berger and Milkman 2012; Berger and Schwartz 2011). In addition, content that is associated with positive awe, negative emotions, or physical arousal motivates people to share more online word of mouth (Berger 2011; Berger and Milkman 2012). Content that generates negative publicity may also increase adoption rates for products that have low awareness due to the increased visibility and higher word of mouth transmission that negativity brings about (Berger et al. 2010). Finally, products that are cued more by environment or are publicly visible may still generate ongoing word of mouth transmission (Berger and Schwartz 2011).

In contrast to prior work in the information systems and marketing fields highlighted above, this research addresses gaps and extends work in these areas in the following ways. First, different from the Bass model framework, we explicitly incorporate the structure of interpersonal connections in the information diffusion process. Second, we extend the literature of peer effects and opinion leaders by explicitly addressing the temporal order of information diffusion. Third, we adopt the temporal networks modeling approach to analyze how the timing of online content generation may affect social contagion or spread of information at the dyadic level and content popularity at the global level. We also investigate how information overload on SMP may moderate the temporal effect of online contents diffusion.

4.3. Theoretical Background

We investigate our research questions based on the theory of temporal networks. In temporal networks, the times when edges are active are an explicit element of the network representation. Figure 4-1 shows an example of a temporal network. Most recent network studies have neglected the time dimension by aggregating the contacts between nodes to edges, even in cases when detailed information on the temporal sequences of contacts or interactions is available. However, the time ordering can be very crucial and the timing of connections and their correlations do have effects that go beyond what can be captured by static networks. For example, the edges between nodes of temporal networks need not be transitive. In static networks, whether directed or not, if A is directly connected to B and B is directly connected to C, then A is indirectly connected to C via a path over B. However, in temporal networks, if the edge from A to B is active only at a later point in time than the edge from B to C, then A and C are disconnected, as nothing can propagate from A via B to C. Figure 4-1 illustrates this point more clearly.



Notes: This figure is used to show that the effect that arises from the time ordering of contacts is crucial in the diffusion process and cannot be captured by static networks. In (a), the temporal dimension is explicitly shown for the communication incidence between nodes A, B, C, and D. In (b), the numbers on the edges indicate the times of the contacts. Assume that, for example, a disease starts spreading at node A and spreads further as soon as a contact happens. According to the history of contact incidence, D will not be infected, as is shown in the $t = \infty$ picture. However, if the spreading started at node D, the entire set of nodes would eventually be infected.

Figure 4-1: Example of Temporal Network
(Adapted from Holme and Saramäki 2011)

Many systems can be modeled as temporal networks, such as the flow of information via email messages, mobile telephone calls, and social media interactions. As Holme and Saramäki (2011) suggest, if the system is temporally and topologically connected in a way that affects the dynamics of interest, then temporal networks may be an optimal theoretical framework. Researchers have shown evidence that users with many followers are not always the best information spreader when the temporal order of information adoption is taken into account in the context of Twitter (Lee et al. 2010). Therefore, we believe that the temporal networks approach is applicable here to address our research questions.

4.4. Research Hypotheses

We have two points to emphasize before we propose our hypotheses. First, the structure of static networks can be characterized by a number of measurements, which are based on connections between neighboring nodes (e.g., the clustering coefficient) or between larger sets of nodes (e.g., the betweenness centrality measure). When the time dimension is included in the network structure, we emphasize that these measures will need a re-evaluation. We explain below how we will revise these measurements when proposing our research hypotheses. Second, since we are investigating information diffusion processes on SMP, we will mainly focus on the consumers' content sharing behavior, through users' activities on social media including both content generation and content sharing.

4.4.1. Basic Assumptions

We first state two basic premises of this study. First, we assume that users usually spend limited time on a daily basis on SMP and thus there are some time periods of a day when users are more active compared to other times (Warren 2010). Second, we assume that users are heterogeneous in their active time periods to consume, generate, and share online contents on SMP. Because online contents are generated and shared by these users in different active time periods, this difference in temporal heterogeneity matters substantially in the information diffusion process (Lee et al. 2010). We will present evidence to support our assumptions when elaborating the data used for the empirical analysis.

A1: On average, social media users are more active in some time periods of a day as compared to other time periods.

A2: Social media users exhibit temporal heterogeneity in active time periods on SMP.

4.4.2. Dyadic Level Timing Effect

At the dyadic level, we posit that whether online content will be shared by a user depends on when the content is exposed to the specific user. When a piece of online content (i.e., the focal content) is generated during a user's inactive time period on social media, this content may not catch the attention of the user when he or she becomes active on a specific SMP. This is because other more recent content may have been generated before the user becomes active, while social media contents (e.g., on Facebook or Twitter) are typically organized and presented in a reverse chronological order. In addition, the large amount of user-generated contents may lead to potential information overload (Cheng et al. 2010; TechCrunch.com 2010). Therefore, the user may neither be interested nor inclined to spend effort to search for the focal information content and thus he or she may not be exposed to it. In contrast, when the online content reaches a user at his or her active opportune time periods on SMP, the probability of the online content being shared by this user will be higher if the content garners the attention or interest of the user.

H1: Online content is more likely to be shared by a user on a SMP when the content is generated or shared by other connected users during the specific user's active time periods on the SMP.

The above arguments suggest that the temporal effect of information sharing will be affected by the extent of information overload for a specific user, i.e., how many other contents are generated by other social media users during the period of time between when the focal content is generated and when the user becomes active on a specific SMP. When there is a high level of information overload, the user's interest and attention on each piece of content are more likely to be dispersed and thus he or she is less likely to share the focal content (Falkinger 2008; Zandt 2004).

H2: The level of information overload moderates the temporal effect of dyadic level information sharing on a SMP, such that the temporal effect of dyadic level information sharing is higher (lower) when there is a higher (lower) level of information overload.

At the global network level, we examine how temporal attributes may affect the popularity (i.e., total number of re-posts or comments that a piece of content receives) and the transmission velocity (i.e., the number of re-posts and number of comments that a piece of content receives per unit time) of online content on SMP. We first define some useful concepts in temporal networks and then propose our research hypotheses below.

4.4.3. Timing Effect on the Popularity of Online Contents

Paths that connect nodes represent the pathways constraining the dynamics of any process taking place on a network. In temporal networks, paths are usually defined as sequences of contacts with non-decreasing times that connect sets of nodes (Pan and Saramäki 2011). In other words, paths must be constrained to sequences of link activations that follow one another in time. The time-respecting paths define which nodes can be reached from which other nodes within some observation window [t_0, T]. The set of nodes that can be reached by time-respecting paths from nodes i is called the set of influence of i (Holme and Saramäki 2011). Previous studies have pointed out the importance of nodes' temporal reachability in diffusion (Holme 2005; Moody 2002). In the social media context, we propose that the reachability of a content generator or the first content sharer has a positive effect on the popularity of the online content on a SMP.

H3: Online content on a SMP will be more popular if the content generator or sharers have a higher level of temporal reachability (i.e., are more temporally connected with other users) when the specific content is generated or shared.

The above hypothesis focuses on the temporal order of information diffusion on a SMP. Another important point is about when the online contents are more viral in general. As indicated by a study on Facebook, content postings during mornings on a brand's fan page are 39.7% more effective in terms of user engagement than those published in the afternoons (Warren 2010). Following our prior assumptions, i.e., users are on average more active in some time periods compared to other time periods, we thus characterize a SMP's active time periods from these users' active durations. Therefore, we posit that online contents generated during these time periods will become more popular.

H4: Online content on a SMP will be more popular if the content is generated or shared during the active time periods of the SMP.

4.5. Research Context

We collected data from a popular micro-blog website in China (Weibo.com) to test our hypotheses. Weibo.com has more than 300 million registered users as of February 2012 (Bloomberg 2012). According to iResearch's report (2011), Sina's Weibo had 56.5% of China's micro-blogging market based on active users and 86.6% based on browsing time over competitors. In Weibo.com, the relationship of "following" between users can be unidirectional; a user does not have to "follow" those who "follow" him[23]. A user receives all the messages from those he or she "follows", and this unique mechanism of following and subscription makes Weibo.com a social medium of information diffusion.

Weibo.com allows data collection of user data through its own set of APIs. We develop our own scripts to collect individual user level data from the site. We specify the network boundary by the following steps. First, we randomly select one initial user, and select his/her followers and followings that are located in one specific area[24] as the first layer users (i.e., core users) of our sampling network. Second, based on the first layer users, we select all their followers and followings as our second layer users (i.e., the snowball sampling approach). Third, we select all their followers and followings of the second layer users, which consist of our third layer users. If we take into account the directions of ties, we actually sampled a network with 5 layers. Our sampling method is similar with the

[23]We adopt the notations from Twitter. We define followers as users who follow the focal user and followings as users who are followed by the focal user.
[24] This criterion is used to narrow down the size of the first layer users.

“expanding selection” approach which is outlined by Doreian and Woodard (1992).

Our data set consist of three types of information: users’ personal information, network connections, and daily activities on Weibo.com. Personal information includes username, gender, location, occupation, number of followers, number of followings, number of posts, and any other information voluntarily filled in by the user. Network information includes the list of followers and the list of followings of a focal user. Users’ daily activities include each piece of online content which is generated or shared by the focal user, the time when the content is generated or re-posted, number of re-posts, number of comments of a focal content, the user ID of the commenter, the time when the focal content is commented, and the detailed content of the comment.

4.6. Econometric Model Specifications

We specify our econometric models in this section. For the dyadic level timing effect analysis, we examine how a user *i*’s active time period affects *i*’s probability to share a piece of content generated or shared by user *i*’s followings. The binary dependent variable $Share_{ikjt}$ indicates whether user *i* shared the content *k* which is generated by user *j* at time *t*. This sharing decision is modeled using a binary logit model:

$$\text{logit}\{\Pr(Share_{ikjt}=1)\}=\alpha_1 Latency_{ij}+\alpha_2 IO_{it}+\alpha_3 Latency_{ij}\times IO_{it}+\boldsymbol{\lambda}\mathbf{Z_i}+\boldsymbol{\theta}\mathbf{J}+\boldsymbol{\delta}\mathbf{V_k}+\ _{\mathbf{ij}} \quad (1)$$

where $Latency_{ij}$ is the fastest time-respecting path between user *i* and *j*, i.e., the time elapsed from the time content *k* is generated by user *j* to the time user *i* becomes active; IO_{it} is the level of information overload of user *i* at time *t*, which is measured as the number of followings of use *i*. $\mathbf{Z_i}$ is a vector of control variables for user i, including user i's daily usage volume, preference, occupation, static network structure, etc.; $\mathbf{J_k}$ is a vector of characteristics of content k, such as sentiment, novelty, and practical usefulness of the content, etc.; $\mathbf{V_{ij}}$ is a vector of variables which capture the relationship between user i and user j, such as whether they are in the same location and the tie strength between them. According to H1 and H2, we expect that α_1 and α_3 are negative.

We measure the popularity of a piece of online content as the overall number of re-posts and overall number of comments this content receives on a SMP (till the time of data collection). To examine the popularity of content k which is generated by user j at time t, we specify the model as follows:

$$Pop_{kjt} = \beta_1 Reach_{jt} + \beta_2 Activity_t + \boldsymbol{\lambda} \mathbf{Z_j} + \boldsymbol{\theta} \mathbf{J_k} + \varepsilon_{kjt} \qquad (2)$$

where Pop_{kjt} is the popularity of content k which is generated by user j at time t; $Reach_{jt}$ is the temporal reachability of user j at time t; $Activity_t$ is the general activity level of the SMP at time t; $\mathbf{J_k}$ is a vector of characteristics of content k, such as whether content k is original generated or re-posted by user j, number of characters, novelty, etc.; $\mathbf{Z_j}$ is a vector of control variables for user j, including user j's daily usage volume, static network structure, etc.; ε_{kjt} is the residual error term. Based on H3 and H4, we expect that β_1 and β_2 are positive.

4.7. Data Description

In this section, we describe our data and show the estimation results of our study. Because of the large amount of micro-blog data from Weibo.com, we sampled a small group of users to conduct the analysis. We generated our data sample based on the following steps. First, we randomly selected one user from the population as the focal user. We selected all the followers of the focal user.

The focal user and his followers formed the core users included in our analysis. Second, we selected all the followers of the core users as the periphery users. Thus, the final network included two layers of users: core users (85) and the periphery users (9828).

The data was collected at the end of May 2012. For the sampled users, we have the list of their followers, their personal profiles, and their detailed posts in 4 weeks (from 15th April 2012 to 13th May 2012). Based on this data set, we first present some evidence that support our basic assumptions of this study.

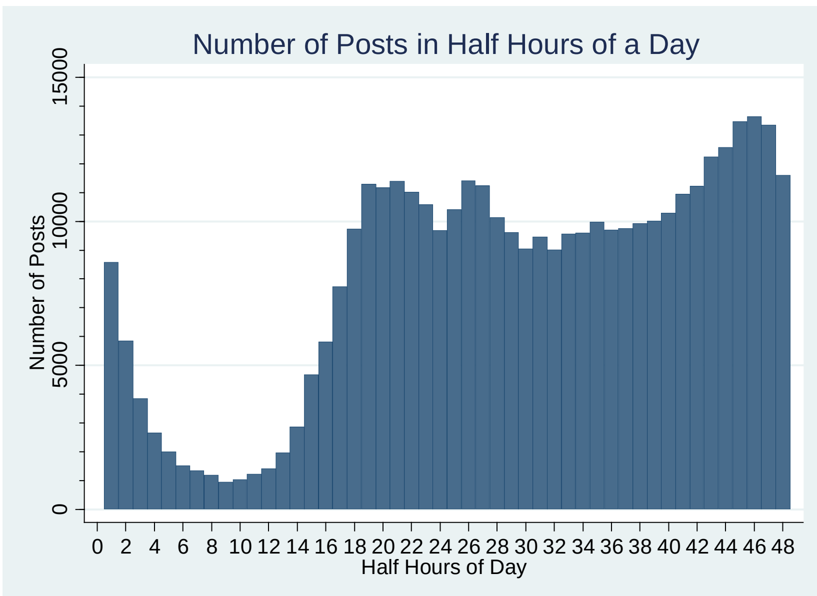


Figure 4-2: Aggregated User Activity Level in Half Hour of Day

In our data set, we divide a day's duration into 48 time slots and each slot lasts for half an hour. Figure 4-2 shows the aggregated number of posts in each half hour slot from our sampled users. Generally, there are a large number of posts (or re-posts) in the day time, while the number of posts (or re-posts) is small in the wee hours of the morning after midnight. Specifically, from midnight to 1 a.m., the number of posts drops sharply. From 7 a.m. onward, the number of posts increases rapidly. At around 9.30 a.m., it reaches the first peak and lasts for more than 2 hours. At noon time, the number of posts drops slightly in a short period. It reaches the second peak at around 1 p.m. and lasts for about 1 hour. From 2 p.m. to 8 p.m., the number of posts maintains at a stable level. At around 8 p.m., users' posting activity starts to increase and reaches the highest peak at 11 p.m. In general, posting activity levels in Figure 4-2 support our assumption A1. Figure 4-3 shows the daily activities of 4 randomly chosen users. These users show substantial heterogeneity in terms of the number of posts across different active time periods. Figure 4-3 thus gives supporting evidence for our assumption A2.

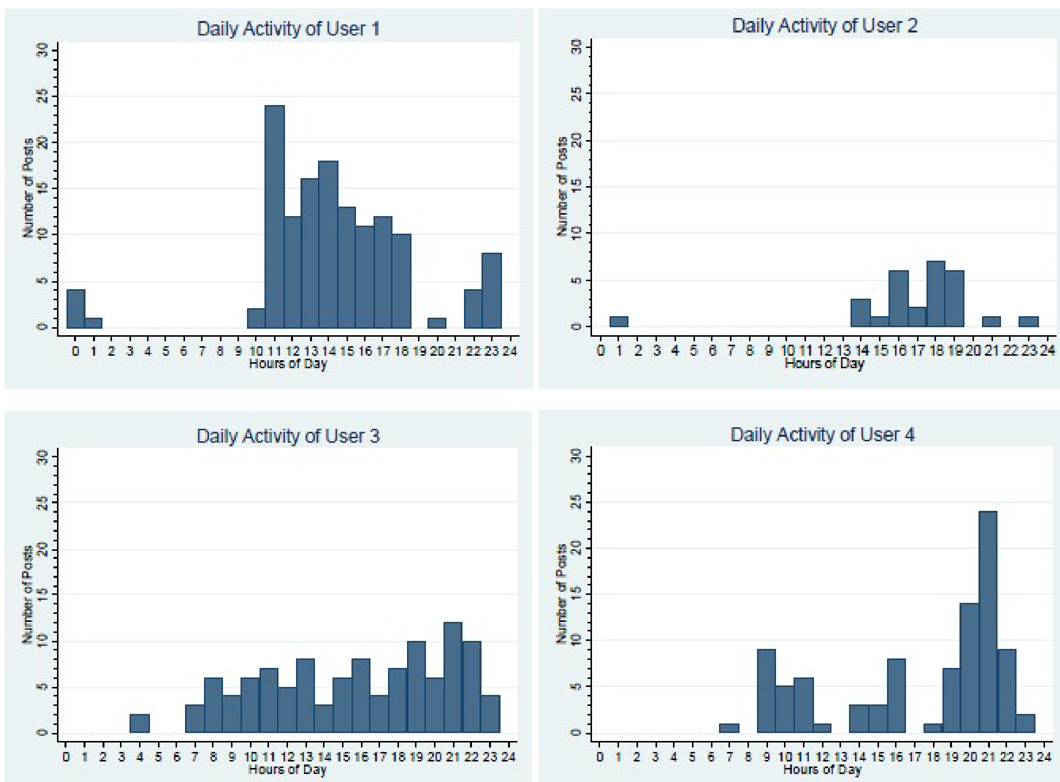


Figure 4-3: User Heterogeneity in Posting Activity and Active Times

Table 4-1 shows the profile information of our sampled users (9913). 53 percent of the users are male. 3 percent of users are verified users (i.e., Weibo.com verified the validity of a user's personal identity). On average, the sampled users have about 581 days' of usage experiences (till the day of data collection), are followed by 810 other users, follows 402 other users, have 95 mutual friends, post (or re-post) 628 pieces of content, and bookmark 68 favorite posts.

Table 4-1: User Profile Information

	mean	sd	min	max
Male	0.53	0.50	0.00	1.00
Duration (days)	581.22	228.08	115.00	1108.00
Verified	0.03	0.17	0.00	1.00
Number of followers	810.12	9195.90	0.00	529448.00
Number of followings	402.42	503.60	1.00	2000.00
Number of mutual friends	94.64	177.78	0.00	1990.00
Number of posts (or re-posts)	628.15	1444.24	0.00	87428.00
Number of favorite posts (or re-posts)	67.54	381.03	0.00	16374.00
N	9913			

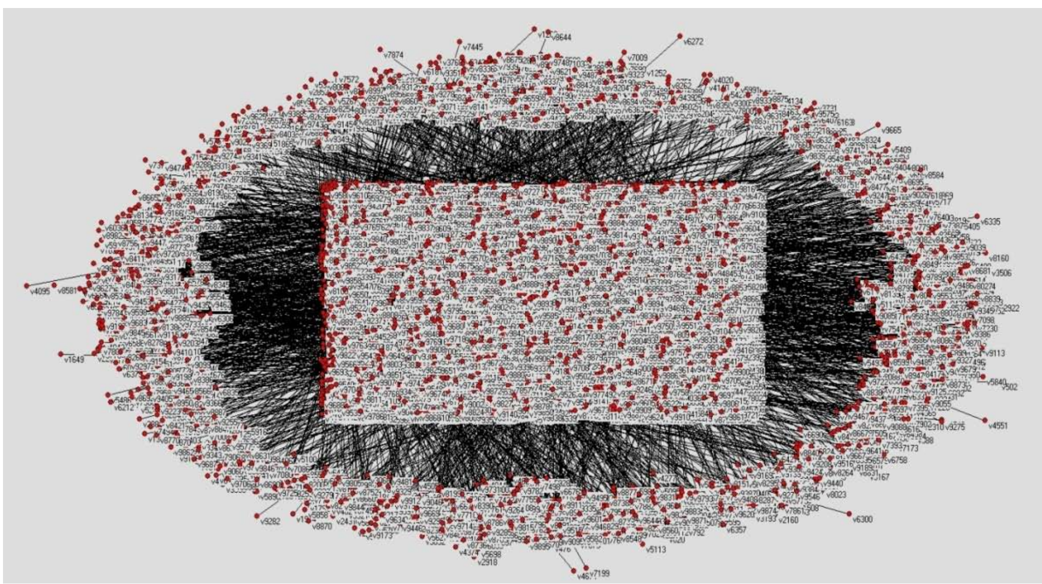


Figure 4-4: User Heterogeneity in Posting Activity and Active Times

Figure 4-4 shows the network structure of our sampled users (9913 users). We can easily identify the two layers, i.e., the core users and the periphery users from the network structure. Figure 4-5 demonstrates the sub-network of the 85 core users. The vertex in the middle is the seed user who was randomly chosen.

We can identify two groups of densely connected users from his/her followers. One is in the bottom-right of Figure 4-5 and the other is in the top-right of Figure 4-5. Our analysis will focus on the posts of these 85 core users.

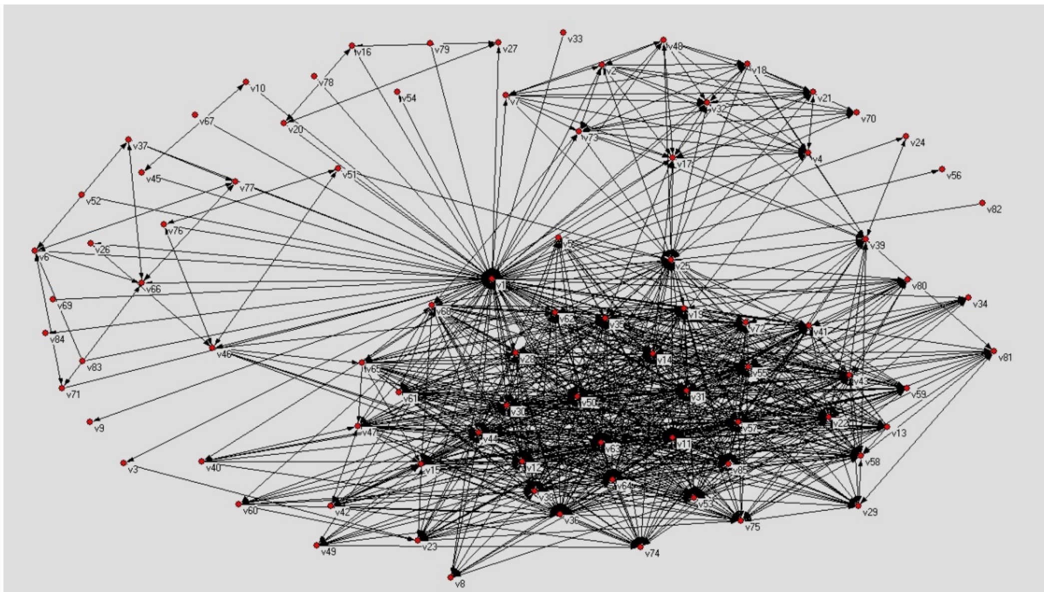


Figure 4-5: Network Structure of the Core Users (85 users)

4.8. Estimation Results

We then show the model estimation results of the econometric model in equation (1) and (2) to validate H1 to H4. We estimate equation (1) to validate H1 and H2. To examine users' re-posts behavior at the dyadic level, we mainly look at how the original posts from the core users are re-posted by their followers. Across the 4 weeks of our data sample, we identify 953 original posts from 66 core users. 199 of these original posts, which are from 39 core users, are re-posted by these 39 users' followers. Our unit of analysis is the "post-follower" pairs. For each observation, the dependent variable is a binary dummy which signals whether user *i* re-posts content *k* that is generated by user *j*. It should be noted that we require user *i* is the follower of user *j*. Latency is measured as the time elapsed from the time content *k* is generated by user *j* to the time use *i* becomes active. We operationalize information overload as the number of followings of use *i*. In addition, we include controls of user *i*, such as number of followers, number of mutual friends, number of posts, and his/her usage experience (till the day of data collect). A location dummy variable is also included which indicates whether user *i* are in the same geographical location with user *j*. The descriptive statistics for variables in equation (1) is shown in Table 4-2.

Table 4-2: Descriptive Statistics for Model 1

	mean	sd	min	max
Share	0.00	0.038	0.00	1.00
Latency	6.91	1.789	-4.09	10.57
IO (Number of followings ('000))	0.40	0.508	0.00	2.00
Latency*IO	2.63	3.160	-8.07	20.80
Number of followers ('000)	0.83	8.899	0.00	529.45
Number of mutual friends ('000)	0.09	0.185	0.00	1.99
Number of posts ('0000)	0.06	0.172	0.00	8.74
Duration (days)	5.61	2.376	1.15	11.06
Location	0.27	0.441	0.00	1.00
N	220527			

Table 4-3 reports the estimation results for equation (1). The coefficient of latency is negative and highly significant, which implies that the long time elapsed, the less likely a piece of content is re-posted by a follower of the content generator. Thus, our hypothesis H1 is supported. The coefficient of information overload (or number of followings) is negative and significant. This result indicates that the more followings a user has, the less likely he or she will re-post his or her followings' posts. However, the coefficient for the interaction term of latency and information overload is positive and insignificant, which suggests that our hypothesis H2 is not supported. In addition, we find that the coefficient of duration is positive and significant. The coefficient of location dummy is positive and significant, which implies that users are more likely to re-post a content which is generated by a user from the same geographical location.

Table 4-3: Estimation Results for Model 1

	Share
Latency	-0.65***
	(0.031)
IO (Number of followings)	-1.77***
	(0.43)
Latency*IO	0.083
	(0.077)
Number of followers	-0.062
	(0.050)
Number of mutual friends	0.53
	(0.56)
Number of posts	0.18
	(0.17)
Duration	0.10***
	(0.028)
Location	0.52***
	(0.11)
Constant	-2.83***
	(0.25)
N	131484
LL	-1883.98
AIC	3786.0

Standard errors in parentheses;
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We estimate equation (2) to test H3 and H4. The 85 core users have 3234 posts in total across four weeks. We measure the dependent variable of popularity of a piece of content in terms of both the number of re-posts (i.e., sharings) and the number of comments. Posts are classified into two categories: original content and re-posted content. We measure the site's general activity level in half-hour intervals as the total number of active users within a half-hour duration (based on our four weeks' observations of 9913 users). User *j*'s temporal reachability when a piece of content *k* is generated or re-posted is measured as the number of user *j*'s active followers. We consider user *j*'s follower *i* as active if user *i* posts or re-posts any information within a 15 minutes interval after the content *k* has been generated or reposted by user *j*. In addition, we also include control variables for users' static network measures, such as out-degree and closeness centrality. Table 4-4 shows the summary statistics for variables included in the empirical model for equation (2).

Table 4-4: Descriptive Statistics for Model 2

	mean	sd	min	max
Number of re-posts	0.38	0.999	0.00	31.00
Number of comments	1.80	3.581	0.00	51.00
Activity (total number of users in half-hour intervals ('000)	2.95	0.562	0.47	3.50
Reach (number of active followers)	3.52	3.395	0.00	22.00
Closeness centrality	0.35	0.072	0.01	0.50
Out-degree('00)	2.45	1.808	0.18	7.39
Re-post dummy	0.71	0.456	0.00	1.00
N	3234			

Table 4-5 shows the OLS results for our equation (2). The coefficient of *Reach* (i.e., users' temporal reachability) is positive and significant when we use the number of reposts as the measure of content popularity. When we use the number of comments as the measure of popularity, the coefficient of *Active*, the site's general activity level is positive and statistically significant. These results provide evidence that users' temporal reachability and a website's temporal activity level do have significant effects on content popularity. The coefficients of users' static network measures are generally positive but only significant when we use the number of comments as the measure of popularity. In addition, we find that re-posted contents are less likely to be re-posted and commented by others. We are currently generating more model covariates as control variables in order to test the robustness of our results. In addition, we will include additional data and incorporate other modeling techniques to address endogeneity issues which we will discuss in the following section.

Table 4-5: OLS Results for Model 2

	(1) Number of Re-posts	(2) Number of Re-posts	(3) Number of Comments	(4) Number of Comments
Activity	0.012	0.015	0.24^{*}	0.32^{**}
	(0.032)	(0.033)	(0.11)	(0.11)
Reach	0.031^{***}	0.029^{***}	0.067^{***}	0.012
	(0.0053)	(0.0084)	(0.018)	(0.028)
Closeness Centrality		-0.18		4.80^{***}
		(0.25)		(0.81)
Out-degree		0.0050		0.13^{**}
		(0.015)		(0.051)
Re-post dummy	-0.17^{***}	-0.17^{***}	-3.11^{***}	-3.13^{***}
	(0.038)	(0.038)	(0.13)	(0.13)
Constant	0.35^{***}	0.40^{**}	3.05^{***}	1.02^{*}
	(0.096)	(0.13)	(0.32)	(0.43)
N	3234	3234	3234	3234
R^2	0.017	0.017	0.159	0.172
AIC	9121.0	9124.4	16875.5	16829.6

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.9. Discussion and Conclusion

This study contributes to the information diffusion, social networks and social media literature by examining the timing effect of information diffusion on SMP. Adopting a temporal networks modeling approach, we develop research hypotheses by focusing on how users' active time periods may affect the social contagion or spread of information at the dyadic level and how the temporal order of information diffusion may affect the popularity of online content on SMP. Using data from a popular micro-blog website, we empirically tested our hypotheses. We found that: (1) users are more likely to share a content which is more recent relative to their active time periods; (2) the more followings a user has, the less likely he or she will re-post his or her followings' posts; (3) a content generator's temporal reachablity has a significant positive impact on the popularity of the content; (4) a content is more likely to get popular when the content is generated during the active time periods of a SMP.

Our study generates the following contributions. First, to the best of our knowledge, this is the first study to examine the temporal effect of information diffusion on SMP in the information systems field. Prior studies have examined the role of local and global network structure of opinion leaders (Iyengar et al. 2011; Katona et al. 2011; Moynihan 2008; Nair et al. 2010; Yoganarasimhan 2012) and the characteristics of information content (Berger and Milkman 2012; Berger and Schwartz 2011; Berger et al. 2010; Zhang and Moe 2012) in the diffusion process. We extend the diffusion literature by providing new insights in terms of how users' active time periods could affect the social contagion or spread of information at the dyadic level and how the temporal order of information diffusion could affect the popularity of online content at the global network level.

Second, our results shed light on how to design and implement more effective and successful viral marketing, advertising or public relations campaigns on SMP. Specifically, our results suggest that firms need to monitor the general active time periods of SMP users and make great efforts to generate social media content on those time periods. For example, the content is more likely to be shared from 9 a.m. to 11 a.m. in our context. In addition, it is also important for managers to analyze the temporal reachability of their networks on SMP to effectively implement viral marketing campaigns. When the viral marketing is targeted at opinion leaders, the temporal reachability of such leaders can be analyzed and acted upon.

Our study has several limitations which suggest directions for future research. First, our study mainly examines the popularity of online content on SMP, while neglecting the process of information diffusion. It will be an interesting question to study how the timing effect influences the velocity of information diffusion on SMP. Second, another potential extension is to study how the timing effect influences the role of opinion leaders or social hubs. It is possible that opinion leaders or social hubs with some characteristics are more influential than others at different time points. Third, our analysis relies on the assumption that the observed network at the time of the study is static. However, the networks on SMP usually change over time. It is also important to explore how the network dynamics may affect our results, especially the potential endogeneity concerns. Despite these limitations, our results clearly show the importance of timing effect in the information diffusion process. We hope that our study will stimulate further interest in this challenging and interesting research area.

CHAPTER 5. CONCLUSION

The objective of this dissertation is to examine the influence of online user-generated content (UGC) on individual consumers' behavior. Applying a wide variety of theories and techniques drawn from economics, marketing, information systems, and psychology, we propose and validate empirically the mechanisms through which UGC influences individual consumers' new product exploration, quality learning, and information sharing behavior. This dissertation extends the literature by (1) investigating how individual consumers search, perceive and use online UGC information when exploring new products; (2) examining how consumers' experiential learning moderating the information role of online UGC; and (3) examining the timing factors that affect the UGC diffusion on social media platforms (SMP).

As a new information source of products, online UGC plays important informative roles in the consumer decision process. It can potentially reduce consumers' quality uncertainty about product or service before their purchases and thus alleviate the information asymmetry between firms and consumers. The first informative role of online UGC we examined is how online UGC can help consumers learn about new products and identify the products that best match their idiosyncratic preferences. Study One investigates the underlying process how individual consumers perceive and use online UGC information to guide their new product exploration as well as purchase decisions. We find that online UGC influences an individual consumer's new product exploration and purchase decision by (1) informing consumers of more choice alternatives in a market, (2) highlighting new choice alternatives that have a higher expected utility than that of their prior choices, and (3) signaling the quality of competing choice alternatives. Our analysis also suggests that consumers follow a two-stage decision process when searching new products from UGC. In the first stage, consumers decide whether to explore a new product. In the second stage, consumers decide which specific product to choose.

The second informative role of online UGC is providing product quality information for consumers. Consumers can make use of online UGC to update their quality perceptions of products. Based on their updated quality perception, consumers decide which product to purchase. Study Two models consumers quality learning from online UGC by following the Bayesian updating framework. Focusing on frequently purchased product category, we extend the literature by examining how consumers' experiential learning moderates the informational role of online UGC on an individual consumer's purchase decision. Our structural econometric model can capture consumer learning from both online UGC and consumption experiences. Our model assumes that consumers learn both the average product quality and the precision of UGC signals. We show that there is a significant amount of consumer learning from consumers' own consumption experiences, much more than from online UGC for frequently purchased products. Neglecting consumers' experiential learning can over-estimate the impact of online UGC on consumer restaurant choice. The impact of online UGC on consumer decisions decreases with the number of consumers' consumption trips. Thus, online UGC promotions may be influential only for new products and the impact can last only for a short time period.

The above two studies mainly look at how individual consumers perceive, interpret, and make use of online UGC to guide their purchase decisions. Study Three investigates how firms can strategically influence consumer's sharing behavior. We propose that the timing of when the content is generated has a significant impact on its popularity. We investigate this timing effect of information diffusion on SMP by adopting a temporal networks modeling approach. We found that: (1) users are more likely to share a recent content along the history of active periods in the data; (2) the more followings a user has, the less likely he or she will re-post his or her followings' posts; (3) a content generator's temporal reachablity has a significant positive impact on the popularity of the content; (4) a content is more likely to become popular when the content is generated during the active time periods of a SMP.

Overall, these undocumented findings on consumer new product exploration, quality learning and information sharing behaviors in the context of user-generated content enhance our understanding of the influence of current social media platforms. Our results also have important implications for practice in terms of content marketing and designing product recommendation systems in e-commerce websites.

First, firms and marketers can actually benefit from online UGC and thus should strategically stimulate consumers to generate more word of mouth information, especially for marketers who want to promote their new products. In order to influence consumers' choice of new product, it is necessary and important for marketers to take consumers' prior consumption experiences into consideration. Consumers have the tendency to switch to products which can offer a higher expected utility than that of their prior choices. This asymmetric choice switching tendency between highly rated products and lowly rated products has important implications on market competition. Given consumers' exposures to online UGC, highly rated products are more likely to capture incremental market share from lowly rated ones. As a result, positive online word of mouth not only increases a firm's customer base but can also mitigate against customer defections. However, we caution that firm should control for consumers' experiential learning when examining the impact of online UGC on consumer choice of frequently purchased products. Because consumers can learn about product quality from both online UGC and their own consumption experiences in frequently purchase product category, neglecting consumers' experiential learning can lead to over-estimation of the impact of online UGC on consumer choice. In addition, our results suggest that firms should strategically choose the right time when promoting their product via online UGC. This timing effect is crucial for implementing more effective and successful viral marketing, advertising or public relations campaigns on SMP. It will be helpful for firms to monitor the general active time periods of SMP users and make great efforts to generate social media content on those time periods. It is also important for managers to analyze the temporal reachability of their networks on SMP. When the viral marketing is targeted at opinion leaders, the temporal reachability of such leaders can be analyzed and acted upon.

Second, our results have practical implications for managers and designers of product recommendation systems on e-commerce websites. In order to increase such websites' informativeness for consumers, it is beneficial to account for consumers' specific purchase or browsing history in personalizing recommendations. Our results suggest that consumers are usually interested in products that can offer superior or better experiences than those of their prior choices, or in products that are rated relatively better among a group of alternatives. Thus, product recommendation website designers should take individual consumers' consumption experience into consideration when designing recommendation systems. For example, it will be directly more effective to recommend consumers a product with a higher quality rating of online UGC than that of their prior choices. In addition, when recommending a product, it is instructive to show how this product is relatively rated in the market. In frequently purchased product category, it is especially important to take individual consumers' consumption experience into consideration when recommending products to consumers. Since consumers can learn about product quality from their own consumption experiences, it will be less efficient to keep recommending the same product to a consumer.

BIBLIOGRAPHY

Abe, S., and M. Tanaka. 1989. Is Brand Evaluation Independent of Other Brands? *Advances in Consumer Research*. **16**(1) 439-442.

Adamowicz, W.L., and J.D. Swait. 2013. Are Food Choices Really Habitual? Integrating Habits, Variety-seeking, and Compensatory Choice in a Utility-maximizing Framework. *American Journal of Agricultural Economics*. **95**(1) 17-41.

Albuquerque, P., P. Pavlidis, U. Chatow, K.-Y. Chen, and Z. Jamal. 2012. Evaluating Promotional Activities in an Online Two-Sided Market of User-Generated Content. *Marketing Science*. **31**(3) 406-432.

Allsop, D.T., B.R. Bassett, and J.A. Hoskins. 2007. Word-of-Mouth Research: Principles and Applications. *Journal of Advertising Research*. **47**(4) 398-411.

Anderson, C. (ed.) *The Long Tail: why the future of business is selling less of more* Hyperion, New York, 2006.

Andrews, R.L., and T.C. Srinivasan. 1995. Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data. *Journal of Marketing Research*. **32**(1) 30-41.

Aral, S., and D. Walker. 2011. Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *Manage. Sci.* **57**(9) 1623-1639.

Archak, N., A. Ghose, and P.G. Ipeirotis. 2011a. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*. **57**(8) 1485-1509.

Archak, N., A. Ghose, and P.G. Ipeirotis. 2011b. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*.

Auty, S. 1992. Consumer Choice and Segmentation in the Restaurant Industry. *The Service Industries Journal*. **12**(3) 324-339.

Bandiera, O., and I. Rasul. 2006. Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal*. **116**(514) 869-902.

Banerjee, A., and D. Fudenberg. 2004. Word-of-mouth learning. *Games and Economic Behavior*. **46**(1) 1-22.

Bapna, R., and A. Umyarov "Are Paid Subscriptions on Music Social Networks Contagious? A Randomized Field Experiment," in: *Workshop on Information Systems and Economics 2011*, Shanghai, China, 2011.

Bass, F.M. 1969. A New Product Growth for Model Consumer Durables. *Management Science*. **15**(5) 215-227.

Bass, F.M., E.A. Pessemier, and D.R. Lehmann. 1972. An Experimental Study of Relationships Between Attitudes, Brand Preference, and Choice. *Behavioral Science*. **17**(6) 532-541.

Bawa, K. 1990. Modeling Inertia and Variety Seeking Tendencies in Brand Choice Behavior. *Marketing Science*. **9**(3) 263-278.

Ben-Akiva, M.E. "The Structure of Travel Demand Models," MIT, 1973.

Ben-Akiva, M.E., and S.R. Lerman (1985), "Discrete Choice Analysis: Theory and Application to Predict Travel Demand," MIT Press.

Berger, J. 2011. Arousal Increases Social Transmission of Information. *Psychological Science*. **22**(7) 891-893.

Berger, J., and K.L. Milkman. 2012. What Makes Online Content Viral?? *Journal of Marketing Research (JMR)*. **49**(2) 192-205.

Berger, J., and E.M. Schwartz. 2011. What Drives Immediate and Ongoing Word of Mouth? *Journal of Marketing Research*. **48**(5) 869-880.

Berger, J., A.T. Sorensen, and S.J. Rasmussen. 2010. Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. *Marketing Science*. **29**(5) 815-827.

Bloomberg "Sina's Weibo Outlook Buoys Internet Stock Gains: China Overnight," 2012.

Bramoulle, Y., H. Djebbari, and B. Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics*. **150**(1) 41-55.

Brock, W.A., and S.N. Durlauf. 2007. Identification of binary choice models with social interactions. *Journal of Econometrics*. **140**(1) 52-75.

Celen, B., and S. Kariv. 2004. Observational Learning Under Imperfect Information. *Games and Economic Behavior*. **47**(1) 72-86.

Chan, T.Y., C. Wu, H. Che, and X. Lu "Consumer Learning and the Informational Value from Online Reviews: An Empirical Study of Dianping.com," 2012a.

Chan, T.Y., C. Wu, H. Che, and X. Lu. 2012b. Consumer Learning and the Informational Value from Online Reviews: An Empirical Study of Dianping.com. *working paper*.

ChannelAdvisor. 2010, Through the Eyes of the Consumer: 2010 Consumer Shopping Habits Survey. Retrieved from http://go.channeladvisor.com/rs/channeladvisor/images/us-wp-consumer-survey-2010.pdf

Chen, P.-Y., S.-y. Wu, and J. Yoon. 2004. The Impact of Online Recommendations and Consumer Feedback on Sales. *2004 Twenty-Fifth International Conference on Information Systems* 711-724.

Chen, Y., and J. Xie. 2008. Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. *Management Science*. **54**(3) 477-491.

Cheng, J., A. Sun, and D. Zeng. 2010, Information Overload and Viral Marketing: Countermeasures and Strategies

Advances in Social Computing. S.-K. Chai, J. Salerno and P. Mabry (Eds.)Springer Berlin / Heidelberg, pp. 108-117.

Chevalier, J.A., and D. Mayzlin. 2006. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*. **43**(3) 345-354.

Chintagunta, P.K. 1999. Variety Seeking, Purchase Timing, and the "Lightning Bolt" Brand Choice Model. *Management Science*. **45**(4) 486-498.

Chintagunta, P.K., S. Gopinath, and S. Venkataraman. 2010. The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*. **29**(5) 944-957.

Clemons, E.K., G.G. Gao, and L.M. Hitt. 2006. When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry. *Journal of Management Information Systems*. **23**(2) 149-171.

ComScore "Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior," in: *November 29*, 2007.

Constine, J. "Study: Facebook Pages Shouldn't Post More Than 1x Every 3 Hours," 2012.

Coscelli, A., and M. Shum. 2004. An empirical model of learning and patient spillovers in new drug entry. *Journal of Econometrics*. **122**(2) 213-246.

Currim, I.S. 1982. Predictive Testing of Consumer Choice Models Not Subject to Independence of Irrelevant Alternatives. *Journal of Marketing Research*. **19**(2) 208-222.

Dabholkar, P.A. 2006. Factors Influencing Consumer Choice of a "Rating Web Site": An Experimental Investigation of an Online Interactive Decision Aid. *Journal of Marketing Theory and Practice*. **14**(4) 259-273.

Das, S.R., and M.Y. Chen. 2007. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*. **53**(9) 1375-1388.

De Groot, M.H. (1970), "Optimal Statistical Decisions," New York, McGraw-Hill.

Decker, R., and M. Trusov. 2010. Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*. **27**(4) 293-307.

Dellarocas, C. 2003. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*. **49**(10) 1407-1424.

Dellarocas, C. 2006. Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms. *Management Science*. **52**(10) 1577-1593.

Dellarocas, C., X. Zhang, and N.F. Awad. 2007. Exploring the Value of Online Product Reviews in Forecasting Sales: the Case of Motion Pictures. *Journal of Interactive Marketing*. **21**(4) 23-45.

Doreian, P., and K.L. Woodard. 1992. Fixed list versus snowball selection of social networks. *Social Science Research*. **21**(2) 216-233.

Duan, W., B. Gu, and A.B. Whinston. 2008a. Do Online Reviews Matter? -- An Empirical Investigation of Panel Data. *Decision Support Systems*. **45**(4) 1007-1016.

Duan, W., B. Gu, and A.B. Whinston. 2008b. The Dynamics of Online Word-of-Mouth and Product Sales--An Empirical Investigation of the Movie Industry. *Journal of Retailing*. **84**(2) 233-242.

Ellison, G., and D. Fudenberg. 1995. Word-of-Mouth Communication and Social Learning. *Quarterly Journal of Economics*. **110**(1) 93.

Erdem, T., and M.P. Keane. 1996. Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science*. **15**(1) 1-20.

Falkinger, J. 2008. Limited Attention as a Scarce Resource in Information-Rich Economies. *The Economic Journal*. **118**(532) 1596-1620.

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

Fotheringham, A.S. 1988. Note—Consumer Store Choice and Choice Set Definition. *Marketing Science*. **7**(3) 299-310.

Garber, T., J. Goldenberg, B. Libai, and E. Muller. 2004. From Density to Destiny: Using Spatial Dimension of Sales Data for Early Prediction of New Product Success. *Marketing Science*. **23**(3) 419-428.

Gatignon, H., J. Eliashberg, and T.S. Robertson. 1989. Modeling Multinational Diffusion Patterns: An Efficient Methodology. *Marketing Science*. **8**(3) 231-247.

Gauri, D.K., A. Bhatnagar, and R. Rao. 2008. Role of word of mouth in online store loyalty. *Commun. ACM*. **51**(3) 89-91.

Ghose, A., P.G. Ipeirotis, and B. Li. 2012. Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*. **31**(3) 493-520.

Gilbride, T.J., and G.M. Allenby. 2004. A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. *Marketing Science*. **23**(3) 391-406.

Givon, M. 1984. Variety Seeking through Brand Switching. *Marketing Science*. **3**(1) 1-22.

Godes, D., and D. Mayzlin. 2004. Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*. **23**(4) 545-560.

Godes, D., and J.C. Silva. 2012. Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*. **31**(3) 448-473.

Goh, K.-Y., C.-S. Heng, and Z. Lin. 2013. Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. *Information Systems Research*. **24**(1) 88-107.

Goldenberg, J., S. Han, D.R. Lehmann, and J.W. Hong. 2009. The Role of Hubs in the Adoption Process. *Journal of Marketing*. **73**(2) 1-13.

Golub, B., and M.O. Jackson. 2012. How Homophily Affects the Speed of Learning and Best-Response Dynamics. *The Quarterly Journal of Economics*. **127**(3) 1287-1338.

Grover, R., and W.R. Dillon. 1985. A Probabilistic Model for Testing Hypothesized Hierarchical Market Structures. *Marketing Science*. **4**(4) 312-335.

Gu, B., J. Park, and P. Konana. 2012. The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*. **23**(1) 182-196.

Guadagni, P.M., and J.D.C. Little. 1983. A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science*. **2**(3) 203-238.

Harlam, B.A., and L.M. Lodish. 1995. Modeling Consumers' Choices of Multiple Items. *Journal of Marketing Research*. **32**(4) 404-418.

Harris, J. "How Often Is The Times Tweeted?," 2010.

Hartmann, W., P. Manchanda, H. Nair, M. Bothner, P. Dodds, D. Godes, K. Hosanagar, and C. Tucker. 2008. Modeling social interactions: Identification, empirical methods and policy implications. *Marketing Letters*. **19**(3-4) 287-304.

Herpen, E.v., and R. Pieters. 2002. The Variety of an Assortment: An Extension to the Attribute-Based Approach. *Marketing Science*. **21**(3) 331-341.

Hill, S., F. Provost, and C. Volinsky. 2006. Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. *Statistical Science*. **21**(2) 256-276.

Holme, P. 2005. Network reachability of real-world contact sequences. *Physical Review E*. **71**(4) 046119.

Holme, P., and J. Saramäki "Temporal networks," 2011.

Huang, P., N.H. Lurie, and S. Mitra. 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.

iResearch "Sina Commands 56% of China's Microblog Market," 2011.

Iribarren, J.L., and E. Moro. 2009. Impact of Human Activity Patterns on the Dynamics of Information Diffusion. *Physical Review Letters*. **103**(3) 038702.

Iyengar, R., A. Ansari, and S. Gupta. 2007. A Model of Consumer Learning for Service Quality and Usage. *Journal of Marketing Research (JMR)*. **44**(4) 529-544.

Iyengar, R., C. Van den Bulte, and T.W. Valente. 2011. Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science*. **30**(2) 195-212.

Jung, H.S., and H.H. Yoon. 2012. Why Do Satisfied Customers Switch? Focus on the Restaurant Patron Variety-Seeking Orientation and Purchase Decision Involvement. *International Journal of Hospitality Management*. **31**(3) 875-884.

Kahn, B.E. 1995. Consumer Variety-Seeking among Goods and Services : An Integrative Review. *Journal of Retailing and Consumer Services*. **2**(3) 139-148.

Kahn, B.E., M.U. Kalwani, and D.G. Morrison. 1986. Measuring Variety-Seeking and Reinforcement Behaviors Using Panel Data. *Journal of Marketing Research*. **23**(2) 89-100.

Kahn, B.E., and T.A. Louie. 1990. Effects of Retraction of Price Promotions on Brand Choice Behavior for Variety-Seeking and Last-Purchase-Loyal Consumers. *Journal of Marketing Research*. **27**(3) 279-289.

Kahn, B.E., and J.S. Raju. 1991. Effects of Price Promotions on Variety-Seeking and Reinforcement Behavior. *Marketing Science*. **10**(4) 316-337.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*. **47**(2) 263-291.

Kapoor, H., and L.A. Heslop. 2009. Brand Positivity and Competitive Effects on the Evaluation of Brand Extensions. *International Journal of Research in Marketing*. **26**(3) 228-237.

Katona, Z., P.P. Zubcsek, and M. Sarvary. 2011. Network Effects and Personal Influences: The Diffusion of an Online Social Network. *Journal of Marketing Research*. **48**(3) 425-443.

Katz, E., and P.F. Lazarsfeld (1955), "Personal influence; the part played by people in the flow of mass communications," Glencoe, Ill. , Free Press.

KÖSzegi, B., and M. Rabin. 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics*. **121**(4) 1133-1165.

Kumar, N., and I. Benbasat. 2006. The Influence of Recommendations and Consumer Reviews on Evaluations of Websites. *INFORMATION SYSTEMS RESEARCH*. **17**(4) 425-439.

Kumar, V., and T.V. Krishnan. 2002. Research Note: Multinational Diffusion Models: An Alternative Framework. *Marketing Science*. **21**(3) 318-330.

Laroche, M., and J.E. Brisoux. 1989. Incorporating Competition into Consumer Behavior Models: The Case of the Attitude-Intention Relationship. *Journal of Economic Psychology*. **10**(3) 343-362.

Laroche, M., M. Hui, and L. Zhou. 1994. A Test of the Effects of Competition on Consumer Brand Selection Processes. *Journal of Business Research*. **31**(2-3) 171-181.

Lee, C., H. Kwak, H. Park, and S. Moon "Finding influentials based on the temporal order of information adoption in twitter," in: *Proceedings of the 19th international conference on World wide web*, ACM, Raleigh, North Carolina, USA, 2010, pp. 1137-1138.

Lemp, J.D., and K.M. Kockelman. 2012. Strategic sampling for large choice sets in estimation and application. *Transportation Research Part A: Policy and Practice*. **46**(3) 602-613.

Li, X., and L.M. Hitt. 2008. Self-Selection and Information Role of Online Product Reviews. *Information Systems Research*. **19**(4) 456-474.

Li, X., and L.M. Hitt. 2010. Price Effects in Online Product Reviews: an Analytical Model and Emperical Analysis. *MIS Quarterly*. **34**(4) 809-A805.

Li, X., L.M. Hitt, and Z.J. Zhang. 2011. Product Reviews and Competition in Markets for Repeat Purchase Products. *Journal of Management Information Systems*. **27**(4) 9-41.

Liu, Q., and N. Arora. 2011. Efficient Choice Designs for a Consider-Then-Choose Model. *Marketing Science*. **30**(2) 321-338.

Liu, Y. 2006. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*. **70**(3) 74-89.

Loewenstein, G.F., and D. Prelec. 1993. Preferences for Sequences of Outcomes. *Psychological Review*. **100**(1) 91-108.

Lu, X., S. Ba, L. Huang, and Y. Feng. 2013. Promotional Marketing or Word-of-Mouth? Evidence from Online Restaurant Reviews. *Information Systems Research, published online ahead of print January 28, 2013*.

Luo, X. 2009. Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices. *Marketing Science*. **28**(1) 148-165.

Lynch, J.G., Jr., D. Chakravarti, and A. Mitra. 1991. Contrast Effects in Consumer Judgments: Changes in Mental Representations or in the Anchoring of Rating Scales? *Journal of Consumer Research*. **18**(3) 284-297.

Mahajan, V., E. Muller, and F.M. Bass. 1990. New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing*. **54**(1) 1-26.

Manski, C.F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*. **60**(3) 531-542.

Mayzlin, D. 2006. Promotional Chat on the Internet. *Marketing Science*. **25**(2) 155-163.

McAlister, L. 1982. A Dynamic Attribute Satiation Model of Variety-Seeking Behavior. *Journal of Consumer Research*. **9**(2) 141-150.

McAlister, L., and E. Pessemier. 1982. Variety Seeking Behavior: An Interdisciplinary Review. *Journal of Consumer Research*. **9**(3) 311-322.

McAlister, L., G. Sonnier, and T. Shively. 2012. The Relationship Between Online Chatter and Firm Value. *Marketing Letters*. **23**(1) 1-12.

McFadden, D. 1978, Modeling the Choice of Residential Location. In: *Spatial Interaction Theory and Planning Models*, A. Karlqvist (Ed.),Amsterdam, North-Holland., pp. 75-96.

McFadden, D., A. Talvitie, S. Cosslett, and B.I.o.T.S. University of California (1977), "Demand model estimation and validation," Institute of Transportation Studies.

McFadden, D.L., and K.E. Train. 1996. Consumers' Evaluation of New Products: Learning from Self and Others. *The Journal of Political Economy*. **104**(4) 683-703.

Meade, N., and T. Islam. 2006. Modelling and forecasting the diffusion of innovation – A 25-year review. *International Journal of Forecasting*. **22**(3) 519-545.

Mehta, N., S. Rajiv, and K. Srinivasan. 2003. Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation. *Marketing Science*. **22**(1) 58-84.

Menon, S., and B.E. Kahn. 1995. The Impact of Context on Variety Seeking in Product Choices. *Journal of Consumer Research*. **22**(3) 285-295.

Moe, W.W., and D.A. Schweidel. 2012. Online Product Opinions: Incidence, Evaluation, and Evolution. *Marketing Science*. **31**(3) 372-386.

Moe, W.W., and M. Trusov. 2011a. The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research (JMR)*. **48**(3) 444-456.

Moe, W.W., and M. Trusov. 2011b. The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*. **48**(3) 444-456.

Mohan, G., B. Sivakumaran, and P. Sharma. 2012. Store Environment's Impact on Variety Seeking Behavior. *Journal of Retailing and Consumer Services*. **19**(4) 419-428.

Moody, J. 2002. The Importance of Relationship Timing for Diffusion. *Social Forces*. **81**(1) 25-56.

Moynihan, R. 2008. Key opinion leaders **-** Independent experts or drug representatives in disguise? *British Medical Journal*. **336**(7658) 1402-1403.

Mudambi, S.M., and D. Schuff. 2010. What Makes a Helpful Online Revie? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*. **34**(1) 185-200.

Nair, H.S., P. Manchanda, and T. Bhatia. 2010. Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders. *Journal of Marketing Research (JMR)*. **47**(5) 883-895.

Nelson, P. 1970. Information and Consumer Behavior. *Journal of Political Economy*. **78**(2) 311.

Nelson, P. 1974. Advertising as Information. *Journal of Political Economy*. **82**(4) 729.

Nerella, S., and C. Bhat. 2004. Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models. *Transportation Research Record*. **1894** 11-19.

Netzer, O., R. Feldman, J. Goldenberg, and M. Fresko. 2012. Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*. **31**(3) 521-543.

Pan, R.K., and J. Saramäki. 2011. Path lengths, correlations, and centrality in temporal networks. *Physical Review E*. **84**(1) 016105.

Petrin, A., and K. Train. 2010. A Control Function Approach to Endogeneity in Consumer Choice Models. *Journal of Marketing Research*. **47**(1) 3-13.

Poston, R.S., and C. Speier. 2005. Effective Use of Knowledge Management Systems: A Process Model of Content Ratings and Credibility Indicators. *MIS Quarterly*. **29**(2) 221-244.

Punj, G.N., and R. Staelin. 1983. A Model of Consumer Information Search Behavior for New Automobiles. *Journal of Consumer Research*. **9**(4) 366-380.

Putsis, W.P., S. Balasubramanian, E.W. Kaplan, and S.K. Sen. 1997. Mixing Behavior in Cross-Country Diffusion. *Marketing Science*. **16**(4) 354-369.

Ratner, R.K., B.E. Kahn, and D. Kahneman. 1999. Choosing Less-Preferred Experiences for the Sake of Variety. *Journal of Consumer Research*. **26**(1) 1-15.

Roberts, J.H., and J.M. Lattin. 1991. Development and Testing of a Model of Consideration Set Composition. *Journal of Marketing Research*. **28**(4) 429-440.

Rogers, E.M. (1995), "Diffusion of innovations," New York, Free Press

Roy, R., P.K. Chintagunta, and S. Haldar. 1996. A Framework for Investigating Habits, “The Hand of the Past,” and Heterogeneity in Dynamic Brand Choice. *Marketing Science*. **15**(3) 280-299.

Sajeesh, S., and J.S. Raju. 2010. Positioning and Pricing in a Variety Seeking Market. *Management Science*. **56**(6) 949-961.

Scott, J.E., and S.K. Keiser. 1984. Forecasting Acceptance of New Industrial Products with Judgment Modeling. *Journal of Marketing*. **48**(2) 54-67.

Seetharaman, P.B., and H. Che. 2009. Price Competition in Markets with Consumer Variety Seeking. *Marketing Science*. **28**(3) 516-525.

Shocker, A., M. Ben-Akiva, B. Boccara, and P. Nedungadi. 1991. Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions. *Marketing Letters*. **2**(3) 181-197.

Simonson, I. 1990. The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior. *Journal of Marketing Research*. **27**(2) 150-162.

Sonnier, G.P., L. McAlister, and O.J. Rutz. 2011a. A Dynamic Model of the Effect of Online Communications on Firm Sales. *Marketing Science*. **30**(4) 702-716.

Sonnier, G.P., L. McAlister, and O.J. Rutz. 2011b. A Dynamic Model of the Effect of Online Communications on Firm Sales. *Marketing Science*.

Staiger, D., and J.H. Stock. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica*. **65**(3) 557-586.

Stigler, G.J. 1961. The Economics of Information. *Journal of Political Economy*. **69**(3) 213-225.

Sun, M. 2012. How Does the Variance of Product Ratings Matter? *Management Science*. **58**(4) 696-707.

Susarla, A., J.-H. Oh, and Y. Tan. 2012. Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube. *Information Systems Research*. **23**(1) 23-41.

Szabo, G., and B.A. Huberman. 2010. Predicting the popularity of online content. *Commun. ACM*. **53**(8) 80-88.

Tang, J., M. Musolesi, C. Mascolo, and V. Latora "Temporal distance metrics for social network analysis," in: *Proceedings of the 2nd ACM workshop on Online social networks*, ACM, Barcelona, Spain, 2009, pp. 31-36.

Tang, J., M. Musolesi, C. Mascolo, and V. Latora. 2010a. Characterising temporal distance and reachability in mobile and online social networks. *SIGCOMM Comput. Commun. Rev.* **40**(1) 118-124.

Tang, J., M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia "Analysing information flows and key mediators through temporal centrality metrics," in: *Proceedings of the 3rd Workshop on Social Network Systems*, ACM, Paris, France, 2010b, pp. 1-6.

Tang, J., S. Scellato, M. Musolesi, C. Mascolo, and V. Latora. 2010c. Small-world behavior in time-varying graphs. *Physical Review E*. **81**(5) 055101.

TechCrunch.com "Eric Schmidt: Every 2 days we create as much information as we did up to 2003," 2010.

Tirunillai, S., and G.J. Tellis. 2012. Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance. *Marketing Science*. **31**(2) 198-215.

Train, K. (2003), "Discrete Choice Methods With Simulation," Cambridge University Press.

Trusov, M., R.E. Bucklin, and K. Pauwels. 2009. Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site. *Journal of Marketing*. **73**(5) 90-102.

Urbany, J.E., P.R. Dickson, and W.L. Wilkie. 1989. Buyer Uncertainty and Information Search. *Journal of Consumer Research*. **16**(2) 208-215.

Van den Bulte, C., and Y.V. Joshi. 2007. New Product Diffusion with Influentials and Imitators. *Marketing Science*. **26**(3) 400-421.

Verna, P. "User-Generated Content: More Popular than Profitable," eMarketer, 2009.

Wang, C.A., X. Zhang, and I.-H. Hann "Social Bias in Online Product Ratings: A Quasi-Experimental Analysis," in: *Workshop on Information Systems and Economics (WISE)*, St. Louis, USA, 2010.

Wang, Q., K.Y. Goh, and J. Chu. 2013. Learning from Self and the Crowd: the Informational Role of User-Generated Content for Frequently Purchased Products. *working paper*.

Warren, C. "When Are Facebook Users Most Active? ," Mashable Business, 2010.

Wooldridge, J.M. (2002), "Econometric Analysis of Cross Section and Panel Data," MIT Press.

Xu, K., S.S. Liao, J. Li, and Y. Song. 2011. Mining comparative opinions from customer reviews for Competitive Intelligence. *Decision Support Systems*. **50**(4) 743-754.

Yoganarasimhan, H. 2012. Impact of social network structure on content propagation: A study using YouTube data. *Quantitative Marketing and Economics*. **10**(1) 111-150.

Zandt, T.V. 2004. Information Overload in a Network of Targeted Communication. *The RAND Journal of Economics*. **35**(3) 542-560.

Zhang, Y., and W.W. Moe "The Discontinuous Diffusion of User-Generated Content," in: *working paper*, 2012.

Zhao, Y., S. Yang, V. Narayan, and Y. Zhao. 2013. Modeling Consumer Learning from Online Product Reviews. *Marketing Science*. **32**(1) 153-169.

Zhu, F., and X. Zhang. 2010. Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*. **74**(2) 133-148.

APPENDICES

1. Research Context for Study ONE and TWO

1.1 When consumers log into the restaurant reviews and UGC website to search and browse online UGC related to restaurants, they first need to choose the city that they are located in. They can then search for a restaurant based on geographical areas, cuisine types, location landmarks, restaurant category tags, or directly search by the keywords they had in mind.

1.2 For example, if a consumer chooses Sichuan cuisine at the above step, the website provides her with a list of restaurants which offers Sichuan cuisine in the city which the consumer is located.

1.3 If this consumer is interested in a specific restaurant, she can click on the link of the restaurant. This will lead her to the reviews homepage of the restaurant, which shows both the aggregate reviews summary information, as well as the individual reviews posted by customers that patronized the restaurant earlier.

2. Robustness Checks for Model Estimations in Study ONE

Table A1: Including a Quadratic Term of *NewRestSearchPercent*

Variables	Model 1: Panel Probit	Model 2: Panel Logit	Model 3: Panel Probit with IV	Model 4: Panel Logit with IV
	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)
NewRestSearchPercent	0.80^{**}	1.40^{**}	4.86^{**}	8.11^{**}
	(0.31)	(0.52)	(1.70)	(2.86)
NewRestSearchPercent _sq	0.44	0.68	0.48^{+}	0.76
	(0.28)	(0.47)	(0.28)	(0.47)
ExperienceVolume	0.13	0.21	0.21^{+}	0.34^{+}
	(0.11)	(0.19)	(0.12)	(0.20)
ExperienceQuality	-0.53^{*}	-0.90^{*}	-0.50^{*}	-0.85^{*}
	(0.25)	(0.41)	(0.24)	(0.41)
ExperiencePrice	0.70^{***}	1.18^{***}	1.05^{***}	1.75^{***}
	(0.21)	(0.35)	(0.25)	(0.43)
OldRestNum	-0.033^{***}	-0.055^{***}	-0.024^{***}	-0.041^{***}
	(0.0064)	(0.011)	(0.0072)	(0.012)
InterPurchaseTime	0.0090^{*}	0.015^{+}	-0.026^{+}	-0.043^{+}
	(0.0045)	(0.0077)	(0.015)	(0.026)
NumOfPerson	0.0083^{+}	0.015^{+}	0.0084^{*}	0.016^{+}
	(0.0043)	(0.0082)	(0.0043)	(0.0082)
Control function residual			-4.12^{*}	-6.81^{*}
			(1.69)	(2.85)
Constant	-0.18	-0.30	-3.45^{*}	-5.71^{*}
	(0.39)	(0.65)	(1.40)	(2.35)
Consumer fixed effect	-included-	-included-	-included-	-included-
Number of consumers	798	798	798	798
Number of observations	3335	3335	3335	3335
Log-likelihood	-2036.9	-2036.7	-2033.9	-2033.8
AIC	4093.8	4093.3	4089.8	4089.6
BIC	4154.9	4154.4	4157.0	4156.8

Notes:
(1) We add the squared term of the variable measuring the extent of new product search.
(2) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2-1: Alternative Measurements of Consumer Search

	Model 1: Panel Probit	Model 2: Panel Logit	Model 3: Panel Probit with IV	Model 4: Panel Logit with IV
Variables	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)
NewUniquePercent	1.24^{***}	2.09^{***}	-0.28	-0.44
	(0.081)	(0.14)	(0.28)	(0.47)
ExperienceVolume	0.13	0.22	0.097	0.16
	(0.11)	(0.19)	(0.11)	(0.19)
ExperienceQuality	-0.52^{*}	-0.88^{*}	-0.55^{*}	-0.93^{*}
	(0.24)	(0.40)	(0.25)	(0.41)
ExperiencePrice	0.67^{**}	1.12^{**}	0.57^{**}	0.96^{**}
	(0.20)	(0.34)	(0.21)	(0.35)
OldRestNum	-0.031^{***}	-0.052^{***}	-0.036^{***}	-0.061^{***}
	(0.0063)	(0.011)	(0.0066)	(0.011)
InterPurchaseTime	0.012^{**}	0.021^{**}	0.023^{***}	0.038^{***}
	(0.0045)	(0.0077)	(0.0048)	(0.0082)
NumOfPerson	0.0083^{+}	0.015^{+}	0.0083^{+}	0.015^{+}
	(0.0042)	(0.0080)	(0.0043)	(0.0082)
Control function residual			1.52^{***}	2.52^{***}
			(0.27)	(0.45)
Constant	-0.26	-0.46	1.02^{*}	1.69^{*}
	(0.38)	(0.63)	(0.45)	(0.76)
Consumer fixed effect	-included-	-included-	-included-	-included-
Number of consumers	798	798	798	798
Number of observations	3335	3335	3335	3335
Log-likelihood	-2056.3	-2055.8	-2039.6	-2039.3
AIC	4130.5	4129.5	4099.3	4098.6
BIC	4185.6	4184.5	4160.4	4159.7

Notes:
(1) *NewUniquePercent* indicates the percentage of unique new restaurants searched.
(2) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2-2: Alternative Measurements of Consumer Search

	Model 1: Panel Probit	Model 2: Panel Logit	Model 3: Panel Probit with IV	Model 4: Panel Logit with IV
Variables	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)
UserTripSearchNum	0.017^{***}	0.029^{***}	0.015^{***}	0.026^{***}
	(0.0037)	(0.0063)	(0.0038)	(0.0065)
ExperienceVolume	0.11	0.19	0.11	0.19
	(0.11)	(0.18)	(0.11)	(0.19)
ExperienceQuality	-0.51^{*}	-0.85^{*}	-0.53^{*}	-0.89^{*}
	(0.23)	(0.38)	(0.24)	(0.41)
ExperiencePrice	0.52^{**}	0.86^{**}	0.57^{**}	0.97^{**}
	(0.19)	(0.32)	(0.20)	(0.35)
OldRestNum	-0.033^{***}	-0.055^{***}	-0.035^{***}	-0.059^{***}
	(0.0060)	(0.010)	(0.0063)	(0.011)
InterPurchaseTime	0.024^{***}	0.042^{***}	0.021^{***}	0.036^{***}
	(0.0046)	(0.0081)	(0.0045)	(0.0076)
NumOfPerson	0.0079^{+}	0.014^{+}	0.0078^{+}	0.014^{+}
	(0.0042)	(0.0074)	(0.0043)	(0.0080)
Control function residual			1.24^{***}	2.09^{***}
			(0.078)	(0.14)
Constant	0.64^{+}	1.03^{+}	0.65^{+}	1.10^{+}
	(0.36)	(0.59)	(0.38)	(0.63)
Consumer fixed effect	-included-	-included-	-included-	-included-
Number of consumers	798	798	798	798
Number of observations	3335	3335	3335	3335
Log-likelihood	-2056.3	-2055.8	-2039.6	-2039.3
AIC	4130.5	4129.5	4099.3	4098.6
BIC	4185.6	4184.5	4160.4	4159.7

Notes:
(1) *UserTripSearchNum* indicates the total number of searches for a specific trip of a consumer.
(2) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2-3: Alternative Measurements of Consumer Search

	Model 1: Panel Probit	Model 2: Panel Logit	Model 3: Panel Probit with IV	Model 4: Panel Logit with IV
Variables	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)
NewSearchNum	0.055^{***}	0.098^{***}	0.027^{***}	0.046^{***}
	(0.0051)	(0.0093)	(0.0054)	(0.0095)
OldSearchNum	-0.11^{***}	-0.18^{***}	-0.024^{+}	-0.040^{+}
	(0.011)	(0.019)	(0.012)	(0.021)
ExperienceVolume	0.12	0.19	0.12	0.19
	(0.11)	(0.18)	(0.11)	(0.19)
ExperienceQuality	-0.44^{+}	-0.74^{+}	-0.50^{*}	-0.86^{*}
	(0.24)	(0.39)	(0.24)	(0.41)
ExperiencePrice	0.52^{**}	0.87^{**}	0.57^{**}	0.96^{**}
	(0.20)	(0.33)	(0.20)	(0.35)
OldRestNum	-0.032^{***}	-0.053^{***}	-0.035^{***}	-0.058^{***}
	(0.0062)	(0.010)	(0.0064)	(0.011)
InterPurchaseTime	0.014^{**}	0.023^{**}	0.019^{***}	0.031^{***}
	(0.0046)	(0.0078)	(0.0045)	(0.0077)
NumOfPerson	0.0080^{+}	0.014^{+}	0.0079^{+}	0.014^{+}
	(0.0042)	(0.0077)	(0.0043)	(0.0080)
Control function residual			1.05^{***}	1.76^{***}
			(0.097)	(0.17)
Constant	0.55	0.91	0.63^{+}	1.06^{+}
	(0.37)	(0.61)	(0.38)	(0.63)
Consumer fixed effect	-included-	-included-	-included-	-included-
Number of consumers	798	798	798	798
Number of observations	3335	3335	3335	3335
Log-likelihood	-2088.0	-2086.0	-2026.1	-2025.8
AIC	4196.0	4192.0	4074.1	4073.6
BIC	4257.1	4253.1	4141.3	4140.8

Notes:
(1) *NewSearchNum* indicates the number of searches for new restaurants.
(2) *OldSearchNum* indicates the number of searches for old restaurants.
(3) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3-1: Alternative Measurements of Consumer Prior Experience

	Model 1: Panel Probit	Model 2: Panel Logit	Model 3: Panel Probit with IV	Model 4: Panel Logit with IV
Variables	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)
NewRestSearchPercent	1.27***	2.14***	1.83	3.07
	(0.078)	(0.14)	(1.27)	(2.13)
LastChoiceVolume	0.032	0.055	0.034	0.058
	(0.045)	(0.075)	(0.045)	(0.075)
LastChoiceQuality	0.067	0.11	0.068	0.11
	(0.098)	(0.16)	(0.098)	(0.16)
LastChoicePrice	0.11	0.18	0.12	0.20
	(0.080)	(0.13)	(0.083)	(0.14)
OldRestNum	-0.030***	-0.050***	-0.028***	-0.047***
	(0.0063)	(0.011)	(0.0070)	(0.012)
InterPurchaseTime	0.011*	0.017*	0.0059	0.0097
	(0.0044)	(0.0076)	(0.012)	(0.019)
NumOfPerson	0.0088*	0.017*	0.0089*	0.017*
	(0.0042)	(0.0082)	(0.0042)	(0.0082)
Control function residual			-0.56	-0.93
			(1.27)	(2.13)
Constant	-0.82***	-1.38***	-1.22	-2.06
	(0.18)	(0.30)	(0.94)	(1.57)
Consumer fixed effect	-included-	-included-	-included-	-included-
Number of consumers	798	798	798	798
Number of observations	3335	3335	3335	3335
Log-likelihood	-2042.3	-2041.9	-2042.2	-2041.8
AIC	4102.6	4101.7	4104.4	4103.5
BIC	4157.6	4156.7	4165.5	4164.6

Notes:
(1) *LastChoiceVolume*, *LastChoiceQuality*, *LastChoicePrice* indicate the consumer's most recent or last consumption experience in terms of number of UGC, average quality rating, and price respectively.
(2) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3-2: Alternative Measurements of Consumer Prior Experience

	Model 1: Panel Probit	Model 2: Panel Logit	Model 3: Panel Probit with IV	Model 4: Panel Logit with IV
Variables	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)	Estimates (Std. Err.)
NewRestSearchPercent	1.26^{***}	2.12^{***}	2.23	3.67
	(0.078)	(0.14)	(1.43)	(2.39)
MaxPriorVolume	-0.0064	-0.011	-0.00032	-0.0014
	(0.035)	(0.058)	(0.036)	(0.060)
MaxPriorQuality	-0.030	-0.057	-0.011	-0.027
	(0.15)	(0.26)	(0.16)	(0.26)
MaxPriorPrice	0.10^{*}	0.17^{+}	0.12^{*}	0.20^{*}
	(0.052)	(0.088)	(0.058)	(0.097)
OldRestNum	-0.035^{***}	-0.059^{***}	-0.035^{***}	-0.058^{***}
	(0.0078)	(0.013)	(0.0078)	(0.013)
InterPurchaseTime	0.011^{*}	0.018^{*}	0.0028	0.0049
	(0.0045)	(0.0076)	(0.013)	(0.021)
NumOfPerson	0.0087^{*}	0.016^{*}	0.0088^{*}	0.016^{*}
	(0.0043)	(0.0082)	(0.0043)	(0.0082)
Control function residual			-0.97	-1.56
			(1.43)	(2.40)
Constant	-0.60^{+}	-1.01^{+}	-1.34	-2.20
	(0.32)	(0.54)	(1.14)	(1.92)
Consumer fixed effect	-included-	-included-	-included-	-included-
Number of consumers	798	798	798	798
Number of observations	3335	3335	3335	3335
Log-likelihood	-2042.7	-2042.3	-2042.5	-2042.1
AIC	4103.4	4102.7	4105.0	4104.2
BIC	4158.5	4157.7	4166.1	4165.4

Notes:
(1) *MaxPriorVolume*, *MaxPriorQuality*,*MaxPriorPrice* indicate the consumer's best prior consumption experience in terms of number of UGC, average quality rating, and price respectively.
(2) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Two-Stage Model Results with Different Disaggregate Quality Ratings

Variables	*Model 1: Taste*		*Model 2: Ambience*		*Model 3: Service*	
	Estimates	Std. Err.	Estimates	Std. Err.	Estimates	Std. Err.
	First Stage: Variety Seeking Decision					
NewRestSearchPercent	-7.99^{***}	1.207	-7.95^{***}	1.217	-8.01^{***}	1.202
NewRestSearchPercent_sq	7.10^{***}	0.721	7.07^{***}	0.728	7.08^{***}	0.720
Control function residual (linear term)	-1.55^{+}	0.929	-1.52	0.942	-1.55^{+}	0.917
Control function residual (squared term)	0.71^{+}	0.400	0.69^{+}	0.412	0.73^{+}	0.392
IV^c	0.71^{***}	0.048	0.71^{***}	0.048	0.73^{***}	0.048
α_i *(Mean)*	2.88^{***}	0.431	3.05^{***}	0.314	2.90^{***}	0.308
α_i *(SD)*	-0.14^{+}	0.087	-0.14^{+}	0.087	-0.14	0.088
InterPurchaseTime	-0.01	0.029	-0.01	0.029	-0.01	0.029
NumberOfPerson	0.02^{*}	0.008	0.02^{*}	0.008	0.02^{*}	0.008
OldRestNum	-0.01	0.049	-0.01	0.050	-0.01	0.048
	Second Stage: Product Alternative Decision					
Search Num	0.24^{***}	0.015	0.24^{***}	0.015	0.24^{***}	0.015
TripNum	0.15^{***}	0.010	0.15^{***}	0.010	0.15^{***}	0.010
Promotion	-0.18^{**}	0.057	-0.16^{***}	0.057	-0.26^{***}	0.058
TagNum	0.35^{***}	0.064	0.30^{***}	0.060	0.37^{***}	0.060
UserRestDistance	-0.16^{***}	0.013	-0.17^{***}	0.012	-0.17^{***}	0.012
Volume (Mean)	-0.80^{***}	0.115	-0.73^{***}	0.110	-0.74^{***}	0.108
Volume (SD)	0.46^{***}	0.122	0.36^{***}	0.138	0.40^{***}	0.118
UGCRating (Mean)	0.67^{**}	0.210	0.96^{***}	0.150	0.03	0.111
UGCRating (SD)	-0.27	0.300	-0.19	0.156	-0.13	0.206
VarianceOfUGCRating (Mean)	-0.63	0.432	-1.21^{*}	0.478	1.14^{***}	0.252
VarianceOfUGCRating (SD)	-0.49	0.833	-1.56^{*}	0.778	-0.04	0.423
Price (Mean)	-0.81^{***}	0.168	-1.19^{***}	0.183	-1.13^{***}	0.179
Price (SD)	0.72^{***}	0.154	0.81^{***}	0.173	0.81^{***}	0.161
VarianceOfPrice (Mean)	0.02	0.170	0.09	0.174	0.14	0.171
VarianceOfPrice (SD)	0.15	0.121	0.13	0.132	0.16	0.124
*Volume * NewRest*	0.45^{***}	0.094	0.47^{***}	0.091	0.45^{***}	0.090
*UGCRating * NewRest*	-0.16	0.266	-0.36^{+}	0.195	0.02	0.146
*VarianceOfUGCRating * NewRest*	0.58	0.525	0.62	0.603	-0.01	0.331
*Price * NewRest*	-0.59^{**}	0.207	-0.39^{+}	0.224	-0.49^{*}	0.219
*VarianceOfPrice * NewRest*	0.11	0.192	0.07	0.200	0.07	0.195
CuisineDummy	-included-		-included-		-included-	
Number of coefficients	46		46		46	
Number of consumers	798		798		798	
Number of observations	3335		3335		3335	
Log-likelihood	-5318.7		-5290.1		-5300.4	
AIC	10729.4		10672.3		10692.9	
BIC	11010.6		10953.4		10974.1	

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$