Understanding, Analyzing and Predicting

Online User Behavior

by

CHUNXIAO LI

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Business Administration

Approved April 2019 by the
Graduate Supervisory Committee:

Bin Gu, Chair
Pei-yu Chen
Hui Xiong

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Due to the growing popularity of the Internet and smart mobile devices, massive data has been produced every day, particularly, more and more users' online behavior and activities have been digitalized. Making a better usage of the massive data and a better understanding of the user behavior become at the very heart of industrial firms as well as the academia. However, due to the large size and unstructured format of user behavioral data, as well as the heterogeneous nature of individuals, it leveled up the difficulty to identify the SPECIFIC behavior that researchers are looking at, HOW to distinguish, and WHAT is resulting from the behavior. The difference in user behavior comes from different causes; in my dissertation, I am studying three circumstances of behavior that potentially bring in turbulent or detrimental effects, from precursory culture to preparatory strategy and delusory fraudulence. Meanwhile, I have access to the versatile toolkit of analysis: econometrics, quasi-experiment, together with machine learning techniques such as text mining, sentiment analysis, and predictive analytics etc. This study creatively leverages the power of the combined methodologies, and apply it beyond individual level data and network data. This dissertation makes a first step to discover user behavior in the newly boosting contexts. My study conceptualize theoretically and test empirically the effect of cultural values on rating and I find that an individualist cultural background are more likely to lead to deviation and more expression in review behaviors. I also find evidence of strategic behavior that users tend to leverage the reporting to increase the likelihood to maximize the benefits. Moreover, it proposes the features that moderate the preparation behavior. Finally, it introduces a unified and scalable framework for delusory behavior detection that meets the current needs to fully utilize multiple data sources.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude and appreciation to my advisor and chair of my dissertation committee, Dr. Bin Gu. Thank you so much for your guidance and support during all stages of my doctoral journey. I am so grateful for all of the hours that he has dedicated to me over the six year to teach me, encourage me and lead me to become who I am today. Thanks for motivating me with your passion for the pursuit of scientific truth, fervor for academic research and love for new explorations. These have not only inspired me in my research work but also contributed a lot to my personality.

I am very much indebted to Dr. Peiyu Chen for her instrumental guidance and enormous help through all the difficulties both in academic area and my life. I would like to deeply thank Dr. Hui Xiong for every idea we discussed at our meetings. They helped me a lot to dive into new dimensions. I thank all my committee members, for their constructive feedback and insightful comments, which have helped improve this dissertation. To my coauthers – Dr. Yanjie Fu, Dr. Chenhui Guo and Dr. Hongchang Wang – great thanks for your contribution and expertise, I appreciate for allowing me to lean on them for support and friendship.

I thank the Department of Computer and Information Systems at the W.P.Carey School of Business for accepting and supporting me. I am sincerely grateful to Dr. Michael Goul and Dr. Raghu Santanam for teaching the doctoral seminar and providing me with the fundamental knowledge for research. To my cohort – Dr. Irfan Kanat, Dr. Seyedreza Mousavi, Dr. Ying Liu, Xueyan Yin, Chen Liang, Ziru Li and Qinglai He, I thank you for the company during teatime and dinner dates. I would also thank Mr. Haihua Ma from *Boohee.Com* as well as Dr. MingJie Zhu and Wei Min from *CreditX Technology* for supporting my work.

I thank my BA advisors Dr. Xiansheng Shen and Lili Ding from University of Science and Technology of China as well as MA advisor Dr. Krupa Viswanathan from Temple University for guiding me into the path of research and encourage to pursue a Ph.D at Arizona State University.

From the bottom of my heart, I would like to express my gratitude to my friends Xin Gan, Han Jiang, Jisun Kim, Xiaoxiao Ma, Madhu Lakshmanan, Jin Qin, X.Y. Sun, Youchen Wu, Jerry Xue, Long Yu,

Tianlong You, Qi Zhu and many others, for their friendship which have made this six-year journey so much more enjoyable.

Lastly, but by no means least, I thank my family for being with me all the ups and downs during the school time of my life. I am eternally grateful for my mother Yiqiu Zhang and father Chuanhuai Li, and my grandparents for their unceasing support and unconditional love. I would extremely like to thank my husband Liang who have always been there for me with love, care and wisdom. I am so blessed to have Victoria Li in my life as my daughter, I am thankful for her sweet smiles, warm hugs and genuine love. I also thank Mosla Li for keep me company, without whom this work would not be possible.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

After the boom of big data beginning in 2010, the massive data from various platforms irons the fire for studying user behaviors. Understanding what, how and why users behave when they are on various online platforms is now one of the most challenging task for the enterprises and institutions. Those who master this art are Titans, and they have the power and privilege to create better interface design (Agichtein, Brill, & Dumais, 2006), better information filtering (Morita, & Shinoda ,1994), better recommendation systems, richer social interactions (Benevenuto, Rodrigues, Cha, & Almeida, 2009), precise targeted advertising (Cui, Shivakumar, Carobus, Jinda, & Lawrence, 2005), etc. And those who fail to understand the behavior of their users, would suffer and die out.

The current information systems studies on this topic is still in a early stage in terms of understanding the users. In the meantime, practitioners have been also exploring user behavior data as a way to manage threads, sustain profit /performance and defend against negative effects, for instance, accounting strategic and detect anomaly intrusion (Oh & Lee, 2003) etc. Despite the substantial existing literature, our knowledge on user behavior is still limited.

In particular, I am interested in examining three circumstances of behavior that potentially bring in turbulent or detrimental effects to the online platforms, from precursory culture to preparatory strategy and delusory fraudulence.

1.1 Research Overview and Questions

1.1.1 Precursory Culture

With the pervasiveness of social media websites, online reviews that carry “Word of mouth” keep booming at an exponential pace. A considerable number of studies for Information systems and marketing have been focus on online reviews. Among them, some studies have pointed out that the effectiveness of online reviews depends largely on the user’s characteristics. However, they seldom explore the characteristics of review authors as possible antecedents of review behavior.

This study aims to extend on this and answer the calls for research on the cross-cultural differences in the production of electronic word of mouth (eWOM) (King, Racherla, & Bush, 2014)

By considering the effects of precursory cultural background on review behavior such as dissenting opinions and emotion expression, I focus on the distinction between individualism and collectivism values as well as on behaviors that are relevant to i) online review authorship and ii) individualist–collectivist cultural values. From the definition, Collectivist values are generally characterized by a preference for preserving harmony, avoiding confrontation, and promoting conformity; individual initiatives and deviations from the dominant opinion of the group are thereby discouraged (Hofstede 2001, House et al. 2004).

In the online reviews context, when a consumer writes a review about a merchant, the pressured to "conform" may effect on his review behavior (Muchnik et al. 2013, Lee et al. 2014, Wang et al. 2015), thus, he may stick to the group opinion as expressed in previous reviews about the same merchant. This behavior tends to emerge because the current average rating of a merchant is prominently shown on the webpage of a review website and may serve as an anchor for subsequent consumers (Adomavicius et al. 2013). It has been widely espoused that collectivist values encourage social harmony and bonding within groups (Triandis 1995, Lam et al. 2009) Therefore, I anticipate that consumers from collectivist cultures are more likely to demonstrate review conformity. The cultural psychology literature includes several studies that suggest a strong relationship between culture and emotion. Previous studies have shown that individuals from individualist cultures tend to be more vocal and expressive, whereas those from collectivist cultures speak in ways intended to maintain harmony and avoid controversy (e.g., using indirect language) (Holtgraves 1997). Similarly findings also occur in Butler et al. (2007) and Niedenthal et al. (2006), whereas those from individualist cultures are more likely to express their emotions, particularly negative emotions (Takahashi et al. 2002) Accordingly, these studies have suggested that the reviews written by individuals from individualist cultures are more likely to contain emotions. Therefore, I propose the following questions:

RQ1: How does individualism (collectivism) influence deviation from (conformity to) prior opinion in online reviews?

RQ2: How does individualism (collectivism) influence emotional expression (suppression) in online reviews?

1.1.2 Preparatory strategy

Nowadays firms often use different incentives to stimulate user activities and enhance their performance in the mobile context. Although financial incentives are widely used, their impact is still unclear, as non-mobile device survey results may not be applicable to mobile settings (Kwon et al. 2016). On the one hand, these incentives may impulse users to take advantages of mobile features, such as mobility, flexibility (Ghose and Han 2011), and social connectivity (Yan and Tan 2014). Specifically, mobile devices enable users to upload and download information anytime, anywhere and expand their social connections to people with similar interests or goals. Given that users have more access to mobile based information channel than traditional channels such as TV and prints (Ghose and Han 2014), mobile enabled financial incentives could have a greater impact on users.

On the other hand, financial incentives on mobile apps may induce preparatory strategic behavior—it may encourage users to take "hidden" actions to increase their chances of receiving financial rewards, which may, in turn, lead to failure of incentives (Mayer et al. 2014) . While previous literature laid the groundwork for understanding the effects of various economic incentives, they rarely conducted any critical tests on strategic behavior triggered by incentives. Without any remedies, such strategic behavior may not only increase the cost of deploying incentive programs, but also decrease the outcome of health intervention and even jeopardize the long-term health statuses of users. Therefore, a deeper understanding of incentive-induced strategic behavior is critical to the success of financial incentives on mobile health apps. In addition, it is worth exploring the possible tools to mitigate strategic behaviors in this contexts. In this aspect, my study explore online social networking features, to understand whether social connections and social activities moderate such behavior. In summary, this paper aims to examine the following research questions:

RQ1: Do financial incentives induce strategic behavior of users in mobile health apps?

RQ2: If there is strategic behavior, what is the net effect of financial incentives on user health outcome?

RQ3: What online social network factor can mitigate strategic behavior?

1.1.3 Delusory fraudulence

The prevalence of online services and mobile devices also has exposed business like e-commerce sites and financial services agencies with online platforms to fraudulent attacks. Statistics (Roberson, 2016) show that fraudulent activities cost about 11.2 billion dollars worldwide in 2012, and the number has increased almost by 100% up to 21.84 billion dollars as of 2016. More seriously, the fraud rates is on the rise with an increase of over 30% on ecommerce fraud attacks in 2017 compared to 2016, while many much more were unidentified in online loads and financials. This is actually a global issue facing by all online Financial related services. In this case, a smarter fraud detective strategy with user behavior analytics is badly demanded to reduce the amount of pretty frauds in Internet finances.

However, detecting fraud base on behavioral data is not a simple outlier detection problem. Cunning mobile frauds are usually not under the radar because they may be more familiar with the traditional detection system than the agencies which have been spending awesomely on resources and manpower. That means, they may not be lying out enough. Moreover, it is because 1) the fraudulence behavior are much less likely to happen than the normal behavior. A large number of user behavior are being collected by the mobile application or website, only a tiny fraction of which are fraudulence behaviors; 2) normal behavior may not always be so "normal". Sometimes normal users wander around pages for certain reason and this could be classified as fraudulent behavior, which induce bias and result in exaggerated fraud rate and blocking lots of normal user from deserved financial supports.

Thanks to the breakthroughs in information technology and fast increasing of data processing capacities, it becomes possible to gather more user behavior pattern data with their daily smartphone usage and the network dynamics through analyzing massive contacting records. To this end, I am drawing a fraud portrait to sort out online fraudulent behavior and identify frauds. I propose to design a framework to predict loan default risk using non-financial data. Moreover, this research turns its focus on the networking of the potential fraud, which do not generate much attention in the past. By exploring the known fraud and the information nodes in one's network, we have a better picture to identify the fraudulent user.

1.2 Major challenges

When studying user behavior data, our researchers normally suffer from following challenges, my goal in this study is to using econometrics and machine learning techniques to overcome them:

Complexity of Data. Online user activities usually are often collected from different websites, smart mobile devices, wearable devices, etc. Data that can be used to model behavior are incredibly diverse. Data cleaning and preprocessing session usually is to turn unstructured data into structure ones and to format data from multiple sources in a proper way.

Volume of Data. Nowadays the mobile apps and websites could collect TB level of data in seconds. Traditional models such as logistic regressions could suffer from a lack of degree of freedom or overfitting because of the high dimensionality.

Selection bias. There may be endogeneity issue behind treatment-control setting due to self-selection since users with specific personality are have more possibility to enter a certain group. Thus may result in potential estimation bias.

False positives. Using regular outlier detection methods tends to result in a large number of false positives because of the unbalance between the two kinds of behavior. In particular, not every piece of behavior of a normal user look perfectly "normal", giving the facts that even regular load applicants may make irregular decisions and behavior for a variety of reasons from time to time.

1.3 Academic Contributions

This research aims to the mobile Internet and related technologies which is developing and changing rapidly. The perspective is to understands the precursory culture, the preparatory strategy and delusory fraudulence, and provides a reference example and ideas for the study of user behavior in mobile applications.

First, this study comprehensively uses the theoretical knowledge of information science, psychology, behavioral economics and other disciplines to study the process of stimulating users' behavior in the empowerment of mobile applications, and to break through the traditional service research in the tasks, objects and methods of research. The limitations of the single-disciplinary perspective is broaden, so is the scope of research theme. Second, using Quasi-Experiment Design methods to better combine theory and practice, our research further clarifies the use of financial incentives, and the mechanism of the action. Moreover, the establishment and verification of strategic behaviors triggered by financial incentives on mobile applications fills gaps in financial research on the mobile Internet. Third, this study uses a technical framework combining econometrics and a variety of advanced artificial intelligence techniques, using difference-in-difference models, deep learning, natural language processing and other research methods to extract features of users' behaviors in mobile applications. Analysis provides the contribution of Design Science and is of great significance to future research on user behavior.

1.4 Managerial Contribution

First, our preliminary study found that strategic behavior occurred after the financial incentive announcement but before its implementation. This finding underscores the importance of considering strategic behavior when evaluating incentive programs. Researchers often quantify the effects of financial incentives by comparing the results before and after the intervention; however, fail to capture potential strategic behavior may lead to overestimation of short-term effects.

Second, the use of social networking features can affect user activity and policy behavior. The existence of social networks is an additional source of monitoring and may trigger a concentrating effect, so participants are less likely to engage in strategic behavior. In the application, we should make full use of this feature of social network, integrate into the health function of the ecosystem, increase user activity and improve health management.

Third, this study analysis different structures and different types of user behavior data (behavior event data, social data, etc.) by using ID-mapping and other technologies, and constructing a unified data model. This is instructive to the follow-up scholars and practitioners; at the same time, the application platform can use both classified models and integrated model to analysis and provides services .

Fourth, the Behavior Language Processing Framework has been established, and this technology has been combined with mobile finance scenario to apply to user activity analysis, behavior monitoring, fraudulence prediction, etc.

CHAPTER 2

CULTURE, CONFORMITY, AND EMOTIONAL SUPPRESSION IN ONLINE REVIEWS

2.1 Background

Much research in various business disciplines, particularly information systems and marketing, has focused on online reviews. Several studies have noted that the effect of online reviews greatly depends on their characteristics. Specifically, negative reviews tend to be more influential than positive reviews (Chevalier & Mayzlin, 2006), whereas an author who expresses emotion in a review can affect the perceived helpfulness of the review (Yin, Bond, & Zhang, 2014) and consumer conversion (Ludwig et al., 2013). Moreover, the disagreement among prior reviews (e.g., higher variance in star ratings) can have varying effects on product sales and the characteristics of subsequent reviews (Nagle & Riedl, 2014; Sun, 2012). Interestingly, few studies have explored the characteristics of review authors as possible antecedents of review content.

To extend prior literature on the antecedents of online reviews (Goes, Lin, & Yeung, 2014; Huang, Burtch, Hong, & Polman, 2016), we focus on the potential role of the cultural background of reviewers (particularly individualism vs. collectivism values)1. In the process, we answer the recent calls for research on the cross-cultural differences in the production of electronic word of mouth (eWOM) (King, Racherla, & Bush, 2014). Anecdotal and scientific evidence jointly suggest that cultural differences have significant potential to explain the variations in review characteristics. By evaluating the Amazon marketplaces in the United Kingdom (UK), Japan, Germany, and the United States (US), Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee (2009) observed “noticeable differences between reviews” in terms of their average helpfulness and rating variance. Other studies that have examined the cross-cultural differences in the production and consumption of online reviews have also reported similar results (Chung & Darke 2006; Fang, Zhang, Bao, and Zhu, 2013; Koh, Hu, & Clemons, 2010). For instance, consumers from collectivist cultures are less likely to write reviews with low valence (i.e., one-star ratings) (Fang et al., 2013). Underreporting biases, which refer to an author’s tendency to write reviews following extreme experiences, are more prevalent among consumers from individualist cultures (Koh et al., 2010). Consumers from individualist cultures are more likely to write reviews for products or services that enable self-expression (Chung & Darke, 2006). However, many questions remain despite these contributions to our understanding of the role of culture in the review process. According to King et al. (2014, p.175), “Understanding these differences and being able to adapt the review process to meet these needs are critical to retailers, so that they can design systems that provide this information in the best manner possible.”.

The majority of the studies on individualism versus collectivism values have focused on their implications on individuals’ tendency to conform or stand out. Accordingly, we focus on the following characteristics of online reviews that are directly linked to conformity and are likely to be influenced by an author’s individualist or collectivist cultural values: 1) conformity to (or deviation from) prior opinion and 2) emotional suppression (or expression). We address the following questions:

RQ1: How does individualism (collectivism) influence deviation from (conformity to) prior opinion in online reviews?

RQ2: How does individualism (collectivism) influence emotional expression (suppression) in online reviews?

RQ3: In turn, how do these cultural influences affect the perceived helpfulness of online reviews?

While Americans say, “the squeaky wheel gets the grease”, the Japanese say, “the nail that stands out gets pounded down” (Goleman, 1990). Such variation in cultural values is not merely anecdotal: much research has established that it exists. For example, researchers in cultural psychology (Hofstede, 2001; House, Hanges, Javidan, Dorfman, & Gupta, 2004) have systematically documented that individuals from collectivist cultures are more likely to exhibit conformity to group opinion (Bond and Smith 1996, Ng et al. 2000) and are less likely to express emotion (Butler, Lee, & Gross, 2007, Niedenthal, Krauth-Gruber, & Ric, 2006). These observations suggest that online reviews that consumers from collectivist cultures write are less likely to deviate from prior opinion and less likely to include emotional expressions than those from individualistic cultures.

In this paper, we empirically evaluate these expectations to extend the literature and answer the calls for research into cross-cultural differences in eWOM (King et al., 2014). First, our work builds on the small body of literature that addresses the cultural differences in the production of online reviews by using data on users from various countries and cultural backgrounds. This technique contrasts those of the majority of previous studies, which have mostly relied on two-country designs (e.g., comparing American and Chinese consumers), which limits the generalizability of their findings (Fang et al., 2013; Koh et al., 2010). Second, previous studies have considered the role of cultural differences in determining the volume and valence of reviews in an absolute sense (Fang et al., 2013). We extend such work by considering the self-group differences in online reviews (i.e., relative valence in terms of deviation from prior opinion) and emotional expression.

Drawing on the cultural psychology literature, we formulated and evaluated several hypotheses using a unique dataset that integrates online restaurant reviews from TripAdvisor.com with country-level measures of individualism/collectivism values (House et al., 2004). We then estimated the effects of these values on the measures of review conformity and emotional suppression. We also examined the subsequent effect of the characteristics of reviews on their perceived helpfulness. We obtained three key findings. First, we found that consumers from countries with a higher level of individualism were more likely to deviate from the prior average rating when writing a review. Second, these reviewers were more likely to express their emotions in the review text. Third, conformity and emotional expression generally had a negative relationship with review helpfulness2.

Our work offers important practical implications for online review platforms. First, recent studies suggest that the approaches that many leading review websites use to aggregate reviews (e.g., averaging) tend to ignore reviewer-specific differences in producing reviews (Dai, Jin, Lee, & Luca, 2012). However, our findings reveal previously undocumented systematic differences in reviewer culture that review websites should consider when aggregating reviews. Second, several features that improve or damage the perceived helpfulness of online reviews (in terms of "helpful" votes) are more likely to systematically manifest when consumers come from a particular culture. Therefore, online practitioners, who are cognizant of these issues, must consider approaches that encourage or deter certain review characteristics. For example, Yelp offers mobile users with "example" reviews to encourage them to produce longer and informative content. Based on an individual's location or review history, one may propose a similar strategy to encourage individuals to include or exclude textual features that do or do not contribute to a “helpful” review.

This paper proceeds as follows. In Section 2, we review the previous studies on online reviews and particularly those that focus on conformity and emotional suppression. We specifically focus on the cultural psychology literature that deals with conformity, language use, and emotional suppression. In Section 3, we propose several hypotheses for empirical examination. In Section 4, we present the research methodology and report our results. In Section 5, we discuss the implications and limitations of our work.

2.2 Literature Review

2.2.1 Online Reviews

Relative to traditional mass communication, online reviews uniquely feature bi-directionality, which emphasizes the need to study both the consumers who created the reviews and effects of these reviews on other consumers (Dellarocas, 2003; Goes et al., 2014). Online reviews enable consumers to share their evaluations and opinions of products or services to an extremely large audience (Dellarocas, 2003; Lee & Bradlow, 2011; Lu, Ba, Huang, & Feng, 2013). Following the pioneering works of Ba and Pavlou (2002) and Dellarocas (2003), many studies from the information systems field have begun to investigate the downstream effects of reviews in terms of sales (Li & Hitt, 2008), helpfulness (Mudambi & Schuff, 2010), and market competition (Kwark, Chen, & Raghunathan, 2014). More recently, researchers have begun to look at how to better design review systems (Liu, Chen, & Hong, 2014) and what factors stimulate online reviews (Burtch, Hong, Bapna, & Griskevicius, forthcoming). We consider the antecedents of review characteristics, which have received relatively less attention in the literature (Goes et al., 2014), by focusing on reviews’ textual characteristics and reviewers’ conformity to (or deviation from) the prior average ratings.

Recent studies have reported evidence on reviewers’ broad conformity (Muchnik, Aral, & Taylor, 2013; Lee, Hosanagar, & Tan, 2015; Wang, Zhang, & Hann, forthcoming). Muchnik et al. (2013) experimentally demonstrate that individuals exposed to a positive prior rating have an increased probability of submitting a positive rating. Similarly, Wang et al. (forthcoming) and Lee et al. (2015) report that individuals' opinions correlate positively with those of their friends. However, contrary to reactance theory (Brehm & Brehm, 1981), a related stream of research (Wu & Huberman 2008; Moe & Schweidel, 2012; Godes & Silva, 2012) has revealed that some individuals are motivated to "stand out" from the crowd by deviating from others' opinions. Conditional on a purchase, a consumer will decide whether to post a review. Wu and Huberman (2008) argue that consumers are motivated, at least in part, by the expected influence of their reviews on the average rating and, implicitly, on the actions or preferences of others. These researchers have empirically revealed that buyers are most likely to post reviews when the expected effect is high (i.e., when only few reviews are present or when their experience extensively deviates from the prevailing average).

Only a handful of studies have investigated reviews' textual characteristics, and the majority of these works have focused on the consequences of textual features. Several textual features affect review helpfulness and product sales. For example, Goes et al. (2014) show that consumers who are more popular in a review community tend to write highly objective reviews. Yin et al. (2014) demonstrate that individuals are likely to perceive certain types of negative emotions (i.e., anxiety) as more helpful than other emotions (i.e., anger). In their study, Ahmad and Laroche (2015) considered the relationship between different types of expressed emotions (i.e., hope, happiness, anxiety, and disgust) and the perceived helpfulness of reviews and observed differential effects across each emotion. Ghose et al. (2011) report that spelling mistakes and review subjectivity are negatively associated with helpfulness and product sales. Huang et al. (2015) examine the effect of anonymity and social presence on review characteristics. We build on the review text literature by considering the antecedent of review emotion (namely, the individualism value of the review author). In Section 2.2, we review the literature on cultural values, conformity, and language use.

2.2.2 Cultural Values, Conformity, and Language Use

Researchers have used national cultural dimensions such as those that Robert House (House et al., 2004) and Geert Hofstede (Hofstede, 2001) introduced to study various phenomena in information systems (Leidner & Kayworth, 2006). However, few studies have explored the role of cultural values in online reviews (Chung & Darke, 2006; Koh et al., 2010; Fang et al., 2013). These researchers contrast the review authorship or consumption between individuals residing in a collectivist country and those residing in an individualist country. Chung and Darke (2006) found that self-relevance has a greater effect on the user-generated content in individualist cultures than that in collectivist cultures. Koh et al. (2010) found that underreporting is more prevalent among U.S. customers than among Chinese or Singaporean customers. Fang et al. (2013) report on a number of several descriptive differences between American and Chinese reviewers. For example, Chinese reviewers provide more positive reviews and place a higher weight on negative reviews. Although previous studies have explored the differences in the behavior of individuals from various cultures, scholars have yet to consider two notable aspects: opinion conformity and emotional suppression. Both aspects tend to differ across cultures, particularly with respect to collectivism versus individualism. First, with respect to conformity, many studies have reported that individuals from collectivist cultures are more likely to conform in judgment and evaluation (Bond & Smith, 1996), behavior (Cialdini, Wosinska, Barrett, Butnet, & Gornik-Durose, 1999), and opinion (Huang, 2005). Second, with respect to emotional expression, several studies have determined that people from individualist cultures are more likely to express emotions (Takahashi, Ohara, Antonucci, & Akiyama, 2002), while those from collectivist cultures are more likely to suppress emotion (Niedenthal, 2006) (particularly negative ones) (Butler et al., 2007).

2.3 Hypothesis Development

We propose several research hypotheses that we empirically test. We divide the research framework into several components. We examine the antecedents in the first stage in which we propose the formal hypotheses about the effects of consumers' cultural backgrounds (countries that exhibit higher levels of individualism versus collectivism) on review characteristics (rating deviation and review textual characteristics). We empirically examine in the second stage the potential relationships between review characteristics and perceived review helpfulness. Figure 2.1 presents the research framework.

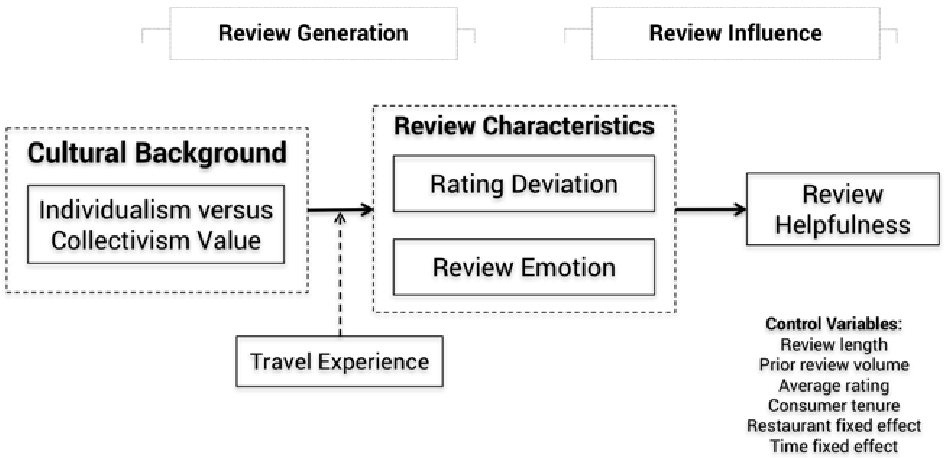


Figure 2.1 Research Framework

2.3.1 Cultural Background and Review Characteristics

By considering the effects of cultural background on dissenting opinions and emotion expression, we focus on the distinction between individualism and collectivism values and on behaviors that are relevant to 1) online review authorship and 2) individualist/collectivist cultural values.

Collectivist values generally feature a preference for preserving harmony, avoiding confrontation, and promoting conformity; as such, such values discourage individual initiatives and deviations from the dominant opinion of the group (Hofstede, 2001; House et al., 2004). Researchers have documented conformity to group pressure in experiments since the 1950s (Asch, 1955). Through a meta-analysis of 133 conformity studies similar to Asch (1955), scholars have also systematically verified that conformity effects are much stronger among individuals from collectivist cultures (Bond & Smith, 1996). Similarly, other studies have observed greater conformity among individuals from collectivist cultures in terms of actual behavior (Cialdini et al., 1999) and opinion formation (Huang, 2005). These findings have a direct bearing on our study context through their suggestions that those reviews written by individuals from collectivist (individualist) cultures are more likely to conform to (deviate from) prior opinions.

In the online reviews context, when a consumer writes a review about a merchant, the consumer may feel pressured to “conform” (Muchnik et al., 2013; Lee et al., 2015; Wang et al., forthcoming) to the group opinion as expressed in previous reviews about the same merchant. This behavior tends to emerge because the webpage of a review website prominently shows a merchant's current average rating, which may serve as an anchor for subsequent consumers (Adomavicius, Bockstedt, Curley, & Zhang, 2013). Yaveroglu and Donthu (2002) argue that individuals from collectivist cultures (e.g., China and Japan) are more likely to conform to the views of others to fit in, gain social understanding, and be accepted by others in a group. Those societies that espouse collectivist values encourage social harmony and bonding in groups (Triandis, 1995; Lam, Lee, & Mizerski, 2009). Therefore, we anticipate that consumers from collectivist cultures will more likely demonstrate review conformity.

In some cases, a consumer may also observe and deviate from prior opinion (Moe & Schweidel, 2012; Wu & Huberman, 2008). As we discuss when reviewing the cultural psychology literature, countries with high individualism values encourage individual autonomy and individualist behavior and discourage conformity. Therefore, individuals from individualist cultures will likely be more "opinionated" because they want to stand out from the others or to have their voices heard. Accordingly, we expect individuals from countries with high individualist values to deviate from prior opinion. Thus, we propose the following:

H1A: On average, those ratings that consumers from individualist (versus collectivist) cultural backgrounds submit are more likely to deviate from (less likely to conform to) the prior average rating.

Several studies in the information systems literature have also examined the role of online review text. Early studies in this line of research have reported the influence of textual content over and above numerical ratings (Pavlou & Dimoka, 2006; Chevalier & Mayzlin, 2006). Recent studies have considered the effects of basic textual features, such as readability and spelling mistakes (Ghose & Ipeirotis 2011; Goes et al., 2014), on review helpfulness. Other studies have explored highly nuanced features, such as semantic style (Cao, Duan, & Gan, 2011) and objectivity versus subjectivity (Ghose & Ipeirotis, 2011).

Scholars have recently examined review texts to identify their emotional and affective content (Ludwig et al., 2013; Yin et al., 2014). They have revealed that such content can strongly affect the perceived helpfulness of a review and its influence on customer conversion. One natural extension is to explore a review author's individualist/collectivist cultural background as a potential antecedent of emotional content in a review.

The cultural psychology literature includes several studies that suggest a strong relationship between culture and emotion. The literature has reported that the tendency toward emotional expression differs according to an individual's cultural background. Previous studies have shown that individuals from individualist cultures tend to be more vocal and expressive, whereas those from collectivist cultures speak in ways intended to maintain harmony and avoid controversy (e.g., using indirect language) (Holtgraves, 1997). Some studies have demonstrated that people from collectivist cultures tend to suppress or withhold their emotions when communicating with others (Butler et al., 2007; Niedenthal et al., 2006), whereas those from individualist cultures are more likely to express their emotions (particularly negative emotions) (Takahashi et al., 2002). These tendencies manifest early in life because children are socialized to meet the standards of their culture (Friedlmeier, Corapci, & Cole, 2011).

Individuals in individualist cultures generally consider expressing emotions publically to be acceptable, but individuals in collectivist cultures generally frown on it. Tsai, Miao, Seppala, Fung, and Yeung (2007) found that American culture typically supports high arousal states, such as excitement and enthusiasm, because these emotions are more effective in influencing others. By contrast, collectivist cultures espouse low arousal states, such as calmness, which better suit adapting to and accommodating others. When writing an online review, a consumer expresses an opinion and performs an evaluation, which often involves a public display of emotion (Yin et al., 2014). Accordingly, these studies have suggested that the reviews written by individuals from individualist cultures are more likely to contain emotions. Therefore, we propose the following:

H1B: Consumers from individualist cultural backgrounds tend to express more emotions in their reviews.

2.3.2 Review Characteristics and Helpfulness

Review helpfulness (generally measured by “helpful” votes) has important implications for both review curators and consumers. To draw practical implications from our study, one must understand how the systematic differences in reviewer behavior may be associated with the perceived helpfulness of reviews. Consumers generally seek different opinions toward the same restaurant prior to consumption to assess whether such establishment can match their tastes (Sun, 2012; Hong, Chen, Hitt, 2013; Liu et al., 2014). Ratings that deviate (either positively or negatively) from prior opinion are likely to stand out and offer unique information by presenting the “other side of the argument” (Cao et al., 2011). Indeed, previous studies have presented consistent evidence that negative reviews, in particular, are likely to be perceived as more helpful because of a “negativity bias’; that is, negative reviews tend to be seen as more informative (Mudambi & Schuff, 2010; Chen & Lurie, 2013). Similarly, we anticipate that rating deviation (i.e., extreme valence relative to past reviews) will result in the higher perceived helpfulness of the review. Therefore, we propose the following:

H2A: Rating deviation is positively associated with review helpfulness.

Several pioneering studies have also employed text mining techniques vis-à-vis the effect of review content. Lee, Hosanagar, and Nair (2013) determine that the presence of informational content in a message may be more or less useful depending on the type of product one considers. Ghose and Ipeirotis (2006) found that objective content is more helpful than subjective content. Considering these past studies and the idea that emotions tend to be perceived as less rational or objective, consumers may perceive reviews that contain greater expressions of emotion as less helpful. Therefore, we propose the following:

H2B: Review emotion is negatively associated with review helpfulness.

2.4 Research Methodology

2.4.1 Data

We collected data from several archival data sources (Table 2.1).First, we collected online reviews (posted between 2003 and 2014) from a leading review platform, TripAdvisor (www.tripadvisor.com), using a Web crawler. Our data included online reviews for approximately 3,750 restaurants located in six major U.S. cities (namely, Chicago, Houston, Los Angeles, New York, Phoenix, Philadelphia, and Seattle). We ensured data accuracy by manually verifying a randomly selected set of 150 reviews. Figure 2.2 presents a screenshot of a review in TripAdvisor.

Table 2.1 Archival Data Sources

Data	Source
Review, reviewer data	TripAdvisor
Review emotion	TripAdvisor reviews processed by Linguistic Inquiry and Word Count (LIWC)
Cultural values	House et al. (2004), World Value Survey

Figure 2.2 Screenshot of a Sample TripAdvisor Review

We constructed our panel by collecting the entire review history of each restaurant and ordering the reviews based on their time stamps. From each review, we obtained the star rating, sequence (order) position, time stamp, and actual review text. We also obtained data on the characteristics of the review authors, including their historical reviewing activity, website registration date, and country of residence. We then measured emotional expression by examining the review text in an automated fashion using the text-mining tool Linguistic Inquiry and Word Count (LIWC), which we describe further in Section 4.2.

Second, we collected data on cultural values from several sources based on prior literature. Several scholars and institutions have attempted to measure national cultural values over the years. Researchers from various fields have extensively used the cultural value data that Hofstede (2001) collected; however, these data are subject to severe limitations because Hofstede collected them from a selected group of IBM employees, which biases them. House et al. (2004) provide a more detailed set of cultural value measures that includes collectivism versus individualism. Researchers have widely used these latest measures in recent years (e.g., Schoorman, Mayer, & Davis, 2007). Researchers in other disciplines have also operationalized cultural values based on the World Values Survey (WVS) (e.g., Giannetti & Yafeh, 2012; Burtch, Ghose, & Wattal, 2014; Hong & Pavlou, 2014)3. By analyzing the results of WVS, Inglehart and Welzel (2010) observed that two factors can explain more than 70 percent of the variance in responses , one of which is survival versus self-expression (the extent to which a society emphasizes values related to survival as opposed to self-expression); these factors capture much of the same information that House et al.'s (2004) the collectivism/individualism measure captures (Inglehart & Oyserman, 2004).

Many researchers consider House et al.'s (2004) culture measure as the most up-to-date and comprehensive because this measure builds on Hofstede (2001), Inglehart and Oyserman (2004), and several other cultural studies. Therefore, we focused on this measure in our primary analysis and subsequently performed a series of robustness checks using the measures from WVS that Inglehart and Welzel (2010) propose. We assigned a consumer (review author) with an individualism value score based on the consumer's self-reported country of residence.

2.4.2 Key Measures

2.4.2.1 Dependent Variables:

Rating deviation: we measured rating deviation as the absolute difference between the rating of a focal review and the average prior rating. TripAdvisor uses a half-star average rating system; therefore, the published average ratings fall in the set (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5). To compute for deviation, we reconstructed the average restaurant rating at the time immediately before the focal review (nth position in the sequence) as follows: $\bar{\mathrm{r}}_n = \frac{i}{n-1} * \sum_{i=1}^{n-1} r_i$. Afterward, we obtained the observed average rating $\hat{r}_n$ by rounding $\bar{\mathrm{r}}_n$ to the nearest half star. For example, for $\bar{\mathrm{r}}_n$=3.24, $\hat{r}_n$=3 (3 stars); for $\bar{\mathrm{r}}_n$=3.26, $\hat{r}_n$=3.5 (3 and a half stars); and for $\bar{\mathrm{r}}_n$=3.76, $\hat{r}_n$=4 (4 stars). In case where $\bar{\mathrm{r}}_n$=3.25, we rounded the value to 3.5. Once can write rating deviation (distance between the nth rating r_n and the observed prior average rating $\hat{r}_n$) as follows:

$$Rating\ Deviation_n = abs(r_n - \hat{r}_n)$$

As a robustness check, we considered the unrounded prior average and formulated an alternative measure of deviation; we obtained almost identical results using the unrounded measure (in terms of the magnitudes and statistical significance of the parameter estimates). However, we expected this result because the actual and observed (rounded) deviation have a 99 percent correlation.

Review emotion: we used LIWC, text-analysis software that identifies sentiment and emotion in textual content, to obtain the measures of emotion (e.g., happy, cried, and abandon), positive emotion (e.g., love, nice, and sweet), and negative emotion (e.g., hurt, ugly, and nasty) (Pennebaker, Francis, & Booth, 2001). LIWC has recently attracted frequent use in the information systems and marketing literatures (Sridhar & Srinivasan, 2012; Yin et al., 2014; Goes et al., 2014). Before calculating the textual measures, we cleaned the textual data to remove special characters. Using LIWC, we operationalized the review emotion measures as the percentage of emotional (overall, positive, and negative) words out of the total number of words.

Review helpfulness: in line with prior literature (Mudambi & Schuff, 2010; Chen & Lurie, 2013), we measured review helpfulness in terms of the total number of “helpful” votes received by a review. Given the highly skewed distribution of votes, we used the log transformation of the raw value in our analyses.

2.4.2.2 Independent Variables

Individualism/collectivism values: we used the collectivism/individualism data from House et al. (2004). These data (from survey responses from 17,300 individuals) are highly consistent with the “survival versus self-expression” measure of WVS and the individualism measure from Hofstede (2001). The collectivism data from House et al. measure the degree to which individuals express pride, loyalty, and cohesiveness in their organizations or families. We employed the negative value of collectivism to measure its polar opposite, individualism. Therefore, higher values of collectivism indicate a greater individualism or lesser collectivism. We plotted the data from House et al. and the self-expression measure of Inglehart and Welzel (2010) based on the most recent wave of

WVS. Figure 2.3 shows that Sweden, the Netherlands, New Zealand, the UK, Denmark, Germany, and the US rank highly on individualism, whereas Russia, China, Georgia, Morocco, Zimbabwe, Hungary, and Albania rank low on individualism. Note that researchers often assume that cultural values are country-level constructs that a country's members inherit. For example, the Chinese or Japanese are generally less assertive and more prone to conformity than Americans. However, the cultural values of individuals in a country may vary because of individual heterogeneity and immigration. Nonetheless, similar to prior literature, we assumed that the inheritance of cultural values holds for the majority of a country's residents because that society embraces those values. Therefore, to avoid confusion with respect to the level of measurement and the ecological fallacy, we referred to subjects' cultural backgrounds instead of their cultural values.

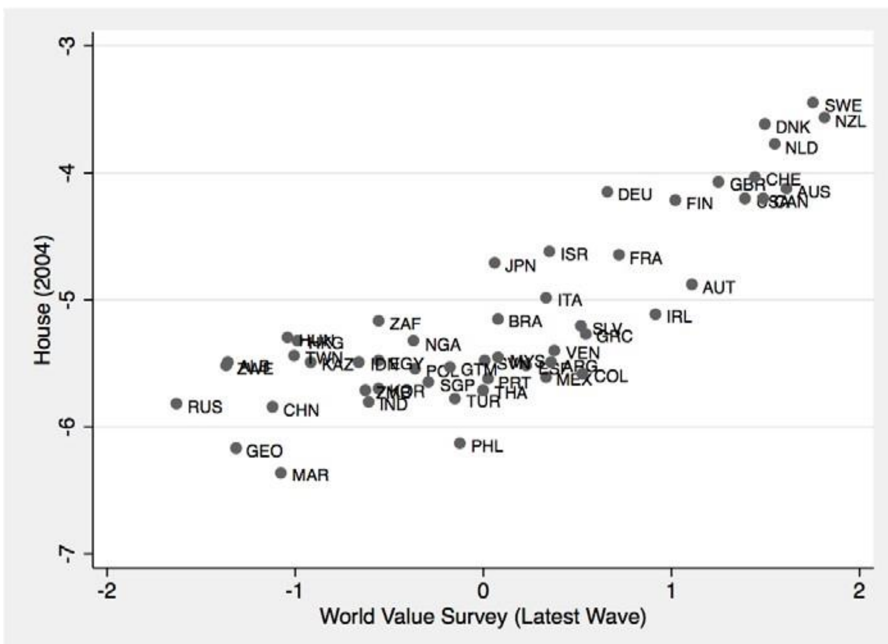


Figure 2.3 Individualism (versus Collectivism) Value by Countries

Travel experience: we measured a consumer's travel experience as the number of countries that the consumer had traveled to as reflected in the TripAdvisor data. Travel experience indicates an individual's exposure to different cultures. Traveling to different countries allows an individual to encounter people of different cultural backgrounds, which makes the individual more receptive to other cultures and exhibit only few of the systematic differences that we have hypothesized. Given its skewness, we log-transformed the travel experience variable.

Prior review volume: prior review volume may affect rating deviation because late arrivals (in terms of the sequence of reviews written for a restaurant) may have different motivations and preferences than the early adopters. For example, a late arriver may have a higher motivation to deviate from prior opinion to make his/her review “stand out”.

Average rating: we controlled for the average rating of a consumer because prior research has noted that some consumers are systematically more positive or negative in their reviewing behavior (Dai et al. 2012).

Consumer tenure: we controlled for consumer tenure (number of months since website registration) for several reasons. First, consumers may grow more positive or negative as they accumulate review experience. We log-transformed this variable in our analyses because of its skewed distribution.

Review age: we controlled for review age (number of days since the review became live and available for consumer voting) because older reviews are exposed to more viewers and have a greater opportunity to accrue helpful votes.

Time effects: we controlled for time effects by employing monthly dummy variables. The reviews written at different periods may be systematically different because of unobserved shocks or trends (e.g., degradation in restaurant quality).

Tables 2.2 and 2.3 present the descriptive statistics and correlation matrix of our key variables, respectively.

Table 2.2 Descriptive Statistics

Variable	**Mean**	**St.d.**	**Min**	**Max**	**Median**
1. Rating deviation	0.79	0.67	0	4.50	0.5
2. Review emotion	8.72	6.52	0	100	7.41
3. Positive emotion	7.96	6.53	0.00	100	6.67
4. Negative emotion	0.74	1.73	0.00	100	0
5. Prior volume	181.46	269.40	1.00	2561	83
6. Individualism	–4.32	0.35	–6.37	–3.46	–4.22
7. Experience	9.52	11.23	1	207	5
8. Average rating	3.77	1.27	1	5	4.1
9. Consumer tenure	28.57	27.79	0	139	21

Table 2.3 Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9
1.Rating deviation	1.00								
2.Review emotion	-0.11	1.00							
3.Positive emotion	-0.16	0.96	1.00						
4.Negative emotion	0.21	0.11	-0.15	1.00					
5.Prior volume	-0.05	-0.03	-0.03	-0.01	1.00				
6.Individualism	0.01	0.04	0.04	0.02	-0.01	1.00			
7.Experience	-0.05	0.01	0.01	0.00	-0.01	-0.14	1.00		
8.Average rating	-0.15	-0.07	-0.05	-0.07	0.06	0.00	0.11	1.00	
9.Consumer tenure	-0.06	-0.05	-0.05	-0.01	0.05	0.03	0.22	0.22	1.00

2.4.3 Empirical Model

Although we operationalized cultural values at the country level, assigning these values to consumers is reasonable in this scenario for several reasons. First, those consumers who are born and raised in a particular country are likely to inherit that country's cultural values. Second, the interaction between country-level cultural values and consumer-level travel experience can help one further identify the effects of cultural values.

We identified the effects of cultural values, travel experience, and the interaction between these two by examining within-restaurant variance in reviews via within transformation (i.e., a standard fixed effect estimation (FE: δj) while controlling for time effects via dummy variables ($\sum_T \tau_t * M_t$). Additionally, we controlled for consumer-level heterogeneity using the abovementioned controls. We formulated the estimation equations for rating deviation and votes as follows: in these two equations, i indexes consumers, j indexes restaurants, and t indexes time; δj is the restaurant fixed effect that controls for restaurant-level, time-invariant unobserved factors; and $\sum_T \tau_t * M_t$ is the vector of monthly time dummies. They key parameters of interest are $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$.

$$\begin{aligned}\text{Rating Deviation}_{ijt} &= \alpha_1 * individualism_i + \alpha_2 * \ln(experience_i) + \alpha_3 \\ &* (individualism * \ln(experience))_i + \alpha_4 * review\ emotion_{itj} + \delta_j + \sum_T \tau_t * M_t \\ &+ control_{ijt} + \varepsilon_{ijt}\end{aligned}$$

$$\text{Review Emotion}_{ijt} = \beta_1 * individualism_i + \beta_2 * \ln(experience_i) + \beta_3 * (individualism * \ln(experience))_i + \beta_4 * rating\ deviation_{itj} + \delta_j + \sum_T \tau_t * M_t + control_{ijt} + \varepsilon_{ijt}$$

$$\ln(helpfulness)_{ij} = \gamma_1 * Rating\ Deviation_{ij} + \gamma_2 * \text{lnwords}_{ij} + \gamma_3 * Rating\ Deviation_{ij} * \text{lnwords}_{ij} + \gamma_4 * review\ emotion_{itj} + \gamma_5 * \text{ReviewAge}_{ij} + \delta_j + control_{ij} + \varepsilon_{ij}$$

2.4.4 Estimation Results and Hypotheses Testing

In this section, we report the estimation results of our main analyses. Following the structure of the hypothesis development, we began by examining the effects of individualism on rating deviation. As Table 2.4 shows, individualism values significantly increased rating deviation, which offers clear support for Hypothesis 1a.

Table 2.4 Effect of Individualism Value on Rating Deviation

DV:	**(1) Rating deviation**	**(2) Rating deviation**	**(3) Rating deviation**
Individualism	0.026***(0.004)	0.015***(0.005)	0.036***(0.014)
ln(experience)		–0.032***(0.002)	–0.069***(0.022)
Individualism*ln(experience)			–0.009*(0.005)
ln(prior volume)		0.000(0.007)	0.000(0.007)
Average rating		–0.066***(0.002)	–0.066***(0.002)
ln(consumer tenure)		–0.004***(0.001)	–0.004***(0.001)
Review emotion		–0.014***(0.000)	–0.014***(0.000)
Constant	1.498*(0.832)	0.000(0.007)	0.036***(0.014)
Restaurant FE	Yes	Yes	Yes
Observations	256,810	256,810	256,810
R-squared (within)	0.015	0.057	0.061
# of restaurants	3,735	3,735	3,735
Notes: robust standard errors are enclosed in parentheses. Std. err. is adjusted for clusters in restaurants. The coefficients are significant at levels *** p < 0.01, ** p < 0.05, and * p < 0.1.			

Although consumers in countries that promote individualist values tend to deviate from the prior average rating, a variation may still exist among consumers in the same country. For example, some consumers from the US may be more conformist, whereas some consumers from China may be more individualistic. One may attribute this variation in cultural values to travel experience, which potentially exposes people to different cultural values. Such exposure makes people more tolerant of other worldviews. Therefore, we further examined the potential moderating role of consumer travel experience on the relationship between individualism values and rating deviation. In particular, we calculated the marginal effects and conducted a spotlight analysis (Spiller, Fitzsimons, Lynch, & McClelland, 2013) to assess both the main and interaction effects. As Figure 2.4 shows, first, the main effect of individualism on rating deviation remained positive across the spectrum of values for different travel experiences. Second, travel experience significantly moderated the effect of individualism on rating deviation. In sum, as individuals gain travel experience, they are potentially exposed to different cultures and become less affected by their own cultural backgrounds.

Table 2.5 Effect of Individualism Value on Review Emotion

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
DVs :	**Overall emotion**	**Overall emotion**	**Overall emotion**	**Positive emotion**	**Positive emotion**	**Positive emotion**	**Negative emotion**	**Negative emotion**	**Negative emotion**
Individualism	0.518*** (0.052)	0.553*** (0.052)	1.452*** (0.128)	0.440*** (0.051)	0.480*** (0.050)	1.295*** (0.122)	0.086*** (0.011)	0.081*** (0.011)	0.177*** (0.026)
ln(experience)		0.014 (0.013)	–1.585*** (0.194)		–0.002 (0.013)	–1.452*** (0.188)		0.017*** (0.004)	–0.153*** (0.044)
Individualism* ln(experience)			–0.375*** (0.045)			–0.340*** (0.044)			–0.040*** (0.010)
ln(prior volume)		0.240*** (0.046)	0.241*** (0.046)		0.225*** (0.045)	0.226*** (0.045)		0.018 (0.011)	0.018 (0.011)
Average rating		–0.094*** (0.014)	–0.095*** (0.014)		–0.041*** (0.014)	–0.042*** (0.014)		–0.053*** (0.004)	–0.053*** (0.004)
ln(consumer tenure)		–0.151*** (0.009)	–0.150*** (0.009)		–0.160*** (0.009)	–0.160*** (0.009)		0.010*** (0.002)	0.011*** (0.002)
Rating deviation		–0.971*** (0.020)	–0.972*** (0.020)		–1.410*** (0.021)	–1.410*** (0.020)		0.439*** (0.007)	0.439*** (0.007)
Constant	13.370*** (0.220)	10.456*** (1.682)	14.286*** (1.741)	12.466*** (0.215)	10.527*** (1.436)	13.997*** (1.503)	0.954*** (0.047)	–0.048 (0.316)	0.360 (0.331)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	256,810	256,810	256,810	256,810	256,810	256,810	256,810	256,810	256,810
R-squared (within)	0.042	0.049	0.049	0.039	0.060	0.060	0.013	0.043	0.043
# of restaurants	3,735	3,735	3,735	3,735	3,735	3,735	3,735	3,735	3,735

Notes: robust standard errors are enclosed in parentheses. Std. err. is adjusted for clusters in restaurants. The coefficients are significant at levels *** p < 0.01, ** p < 0.05, and * p < 0.1.

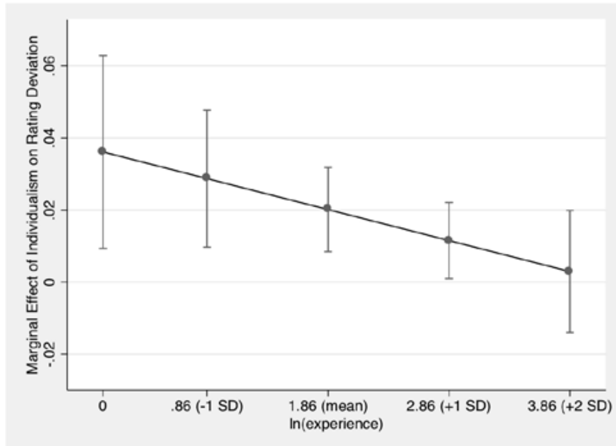


Figure 2.4 Spotlight Analysis of the Interaction Effect on Rating Deviation

We then examined the effects of cultural background on review emotion and, specifically, overall emotion (Column 1 of Table 2.5), positive emotion (Column 3), and negative emotion (Column 5). First, consumers from individualistic cultures were more likely to express both positive and negative emotions, which supports Hypothesis 1b. Travel experience attenuated all these estimated direct effects of cultural values on review text (Figure 2.5 visualizes the main and interaction effects). Table 2.5 and Figure 2.5 show that consumers from an individualist cultural background always expressed a higher level of overall, positive, and negative emotions and that travel experience attenuated the positive effects. Interestingly, we observed that rating deviation was positively correlated with the presence of negative emotion yet negatively correlated with the presence of positive emotion.

(a) Overall Emotion (b) Positive Emotion (c) Negative Emotion

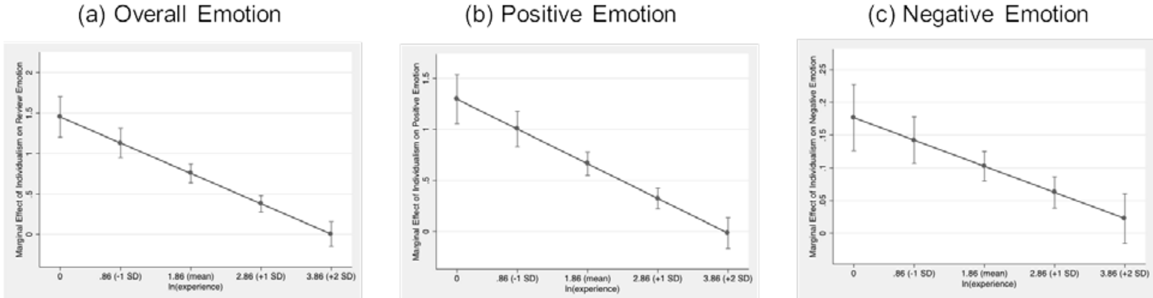


Figure 2.5 Spotlight Analysis of the Interaction Effect on Review Emotion

Table 2.6 Effect of Review Characteristics on Review Helpfulness

	(1)	(2)	(3)	(4)	(5)
Rating deviation	–0.044*** (0.005)	–0.049*** (0.005)	–0.029*** (0.007)	–0.025*** (0.007)	
ln(words)	0.048*** (0.002)	0.047*** (0.002)	0.050*** (0.002)	0.051*** (0.002)	
Rating deviation * ln(words)	0.022*** (0.001)	0.023*** (0.001)	0.020*** (0.002)	0.019*** (0.002)	
Emotion	–0.000** (0.000)		0.000** (0.000)		
Rating deviation * emotion			–0.001*** (0.000)		
Positive emotion		–0.000*** (0.000)		0.001*** (0.000)	
Negative emotion		0.003*** (0.000)		–0.001** (0.001)	
Rating deviation * positive emotion				–0.002*** (0.000)	
Rating deviation * negative emotion				0.003*** (0.000)	
Individualism					0.005* (0.003)
Review age	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Constant	–0.139*** (0.011)	–0.133*** (0.011)	–0.152*** (0.011)	–0.155*** (0.011)	0.120*** (0.013)
Restaurant FE	Yes	Yes	Yes	Yes	Yes
Observations	298,458	298,458	298,458	298,458	298,458
R-squared	0.070	0.070	0.070	0.071	0.040
# of restaurants	3,747	3,747	3,747	3,747	3,747

Notes: robust standard errors are enclosed in parentheses, Std. err. is adjusted for clusters in restaurants. The coefficients are significant at levels *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

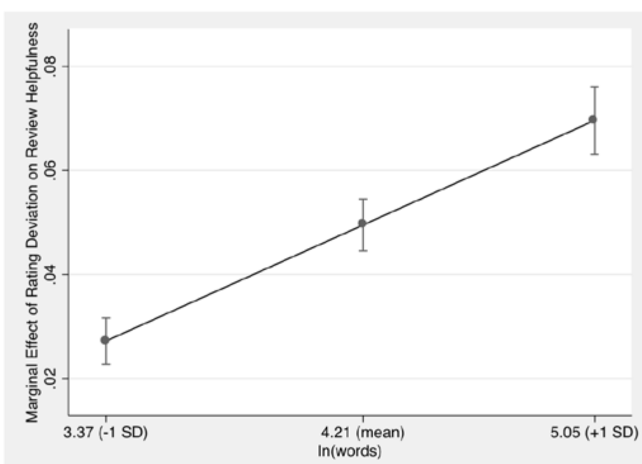


Figure 2.6 Interaction Effect of Rating Deviation and Review Length on Helpfulness

We then examined the effects of review characteristics (rating deviation and review emotion) on review helpfulness in terms of “helpful” votes. As the regression results in Table 2.6 and the plot in Figure 2.6 show, rating deviation increased the perceived helpfulness of a review, which supports Hypothesis 2a. We also found a positive interaction between rating deviation and review length (see Figure 2.6), which suggests that deviation exerts a greater influence when the textual content conveys more information.

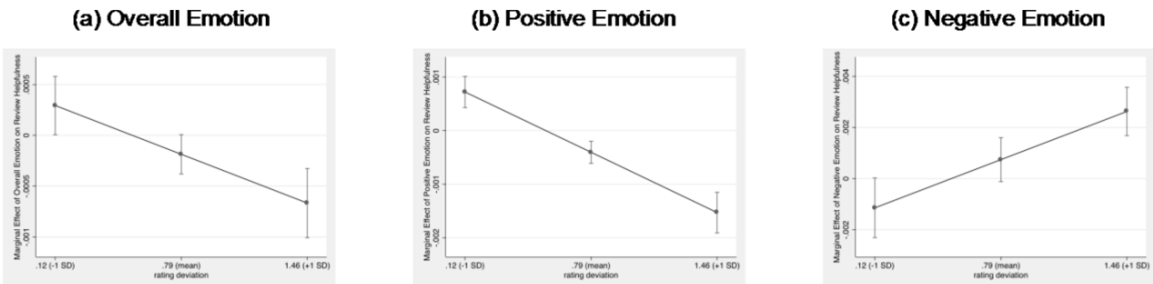


Figure 2.7: Interaction Effect of Rating Deviation and Review Emotion on Helpfulness

Given the influence of review emotion, we observed that overall emotion had a negative effect on review helpfulness, which supports Hypothesis 2b. When we further broke down the positive and negative emotions, we found that positive emotions led to lower review helpfulness, whereas negative emotions increased review helpfulness. This finding is consistent with the “negativity bias” in online reviews that prior research has demonstrated (Chen & Lurie, 2013). Beyond the main effects, we observed significant interaction effects between rating deviation and review emotion (Figure 2.7), which indicates that positive emotion and rating deviation have a significant negative interaction effect (substitutive effect) on review helpfulness and that negative emotion and rating deviation have a positive interaction effect (complementary effect) on review helpfulness. Column 5 of Table 2.6 shows that individuals generally perceive those reviews written by consumers from individualist cultural backgrounds to be more helpful.

2.4.5 Robustness Checks

We validated the robustness of our results in several ways. We first considered alternative measures of cultural values by re-running our analyses using data from WVS (see Section 4.5.1).

The first set of robustness checks aimed to demonstrate that the measurement of cultural values did not drive the observed results. We also considered an alternative estimation approach (seemingly unrelated regression (SUR)) by allowing the error terms of Equations (2) and (3) to correlate with each other (see Section 4.5.2).

2.4.5.1 Robustness Check 1: Alternative Measures of Cultural Background

We obtained an additional dataset on cultural values from WVS and re-estimated our models to confirm the stability of our results. We observed a high correlation between the measures of House et al. (2004) and WVS (ρ = 0.90) in our sample4. Given the lack of temporal variation in the WVS data, we used the most recent set of survey responses. Our main results remained stable regardless of our chosen measure (see Table 2.7).

Table 2.7 Robustness Check: Estimation Using Alternative Measure

DVs:	**(1) Rating deviation**	**(2) Overall emotion**	**(3) Positive emotion**	**(4) Negative emotion**
Individualism	0.027*** (0.010)	0.917*** (0.110)	0.803*** (0.107)	0.117*** (0.019)
ln(experience)	–0.017*** (0.006)	0.446*** (0.067)	0.376*** (0.066)	0.070*** (0.014)
Individualism* ln(experience)	–0.009** (0.004)	–0.255*** (0.039)	–0.224*** (0.038)	–0.032*** (0.008)
ln(prior volume)	0.001(0.007)	0.240*** (0.046)	0.223*** (0.045)	0.019* (0.011)
Average rating	–0.066*** (0.002)	–0.096*** (0.014)	–0.042*** (0.014)	–0.054*** (0.004)
ln(consumer tenure)	–0.004*** (0.001)	–0.154*** (0.009)	–0.163*** (0.009)	0.010*** (0.002)
Review emotion	–0.014*** (0.000)			
Rating deviation		–0.973*** (0.020)	–1.412*** (0.020)	0.440*** (0.007)
Constant	1.663*** (0.640)	6.547*** (1.664)	7.124*** (1.411)	–0.593* (0.318)
Restaurant FE	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes
Observations	259,460	259,460	259,460	259,460
R-squared (within)	0.046	0.049	0.060	0.042
# of restaurants	3,736	3,736	3,736	3,736
Notes: Robust standard errors are enclosed in parentheses. Std. Err. is adjusted for clusters in restaurants. The coefficients are significant at levels *** p < 0.01, ** p < 0.05, and * p < 0.1.				

2.4.5.2 Robustness Check 2: Alternative Estimation Approach

We also obtained results using the SUR model, which controls for the possibility that review deviation and emotion are co-determined. SUR allows one to correlate the error terms of Equations (2) and (3) and jointly estimates these equations. The estimation results, as Table 2.8 shows, are consistent with our main results, which indicates their robustness.

Table 2.8 Robustness Check: SUR Estimation

DVs:	**(1) Rating deviation**	**(2) Overall emotion**	**(3) Rating deviation**	**(4) Negative emotion**	**(5) Positive emotion**
Individualism	0.102*** (0.012)	2.238*** (0.099)	0.068*** (0.012)	0.207*** (0.027)	2.045*** (0.099)
ln(experience)	–0.153*** (0.020)	–2.565*** (0.167)	–0.117*** (0.020)	–0.187*** (0.045)	–2.386*** (0.166)
Individualism * ln(experience)	–0.028*** (0.005)	–0.601*** (0.039)	–0.020*** (0.005)	–0.053*** (0.010)	–0.551*** (0.039)
Emotion	–0.029*** (0.000)				
Rating deviation		–2.003*** (0.016)		0.843*** (0.004)	–2.758*** (0.016)
Negative emotion			0.145*** (0.001)		
Positive emotion			–0.033*** (0.000)		
ln(prior volume)	–0.018*** (0.001)	–0.064*** (0.009)	–0.019*** (0.001)	0.016*** (0.002)	–0.078*** (0.009)
Average rating	–0.068*** (0.001)	–0.160*** (0.009)	–0.054*** (0.001)	–0.025*** (0.003)	–0.128*** (0.009)
ln(consumer tenure)	–0.007*** (0.001)	–0.148*** (0.009)	–0.009*** (0.001)	0.011*** (0.002)	–0.158*** (0.009)
Constant	2.303*** (0.376)	19.197*** (3.127)	2.130*** (0.368)	–0.235 (0.839)	19.336*** (3.125)
Time effect	Yes	Yes	Yes	Yes	Yes
Observations	256,810	256,810	256,810	256,810	256,810
R-squared	0.038	0.041	0.050	0.016	0.043
Notes: standard errors are enclosed in parentheses. The coefficients are significant at levels *** p < 0.01, ** p < 0.05, and * p < 0.1.					

2.5 Discussion

2.5.1 Key Findings

This study is the first to theoretically conceptualize and empirically test the effect of cultural values on rating deviation and review emotion in online restaurant reviews. First, we demonstrate that consumers from an individualist cultural background are more likely to deviate from prior opinion. Second, consumers from an individualist cultural background are more likely to express emotion in their reviews. Third, these two characteristics of online reviews can have important implications for review helpfulness.

2.5.2 Implications

This study offers several theoretical implications. Our work is the first to consider that cultural differences may affect consumers' tendency to deviate from (or conform to) past reviews. Recent work has suggested that the present review aggregation approach employed by many leading platforms (e.g., Yelp) tends to ignore the systematic differences in reviewer behavior and conformity in the review-generation process (Dai et al., 2012). Our findings point to a previously undocumented driver of the systematic differences in review characteristics. This driver relates to reviewers' cultural background, which not only has the potential to exacerbate or mitigate herding in review generation but also has similar negative implications for the optimality of existing review aggregation techniques.

Our work is also the first to consider how cross-cultural differences manifest in reviews' textual characteristics beyond their length. Consumers from individualist cultural backgrounds express more emotion in their reviews. In turn, both conformity (lacking rating deviation) and review emotion lead to lower review helpfulness. Our work is among the first to draw a connection between the cultural background (values) of authors and the perception of audiences toward the quality of the review content. These results imply that the operators of online review sites must recognize the systematic, cross-cultural differences in the produced content and that they must consider some approaches to mitigate biases whenever they damage a review's perceived helpfulness. For instance, review platforms may offer examples of "helpful" reviews to consumers that consider the reviewer's reviewing history or country of residence. Alternatively, review platforms may seek and solicit reviews from individuals with a particular cultural background to elicit helpful reviews for others.

Online review aggregators, such as those presented in this study, are likely to be of greatest use for products and services that cater to various customer segments; namely, consumers from different cultural backgrounds. If one aggregates reviews based on the reviewing tendencies of consumers (e.g., weighting reviews based on the cultural background of authors and their anticipated likelihood of under- or over-stating divergent opinions), one may improve the consumer search process, reduce search costs, and expect better purchase decisions.

Previous studies that have considered cross-cultural differences in online reviews have almost exclusively employed a two-country design to explore cross-cultural differences in the production and consumption of reviews (Chung & Darke, 2006; Koh et al., 2010; Fang et al., 2013). By contrast, this study leverages a large observational dataset of reviews that consumers from 52 countries wrote. Therefore, our findings have external validity.

2.5.3 Limitations and Opportunities

Similar to other studies, our work is subject to some limitations. First, we measured cultural values at the national level and then ascribed them to individuals based on their country of residence. A more accurate measure should employ a survey of each consumer based on the original measures of Hofstede (2001) or House et al. (2004). However, this approach entails surveying a large number of TripAdvisor users, which is impractical because of our limited access. We acknowledge this limitation and interpret the observed effects as derived from consumers' "cultural backgrounds" rather than their "cultural values"5. Future research may employ a different research design to address this limitation. For example, researchers may recruit reviewers, survey their cultural values at the individual level, and then ask them to complete a review task.

Second, a person may be born and raised in one country but immigrate to another country. Unfortunately, we cannot observe this behavior in our archival data. Nevertheless, this limitation is unlikely a prevalent issue in our sample and considering it would introduce noise into our estimations, which would prevent us from identifying the hypothesized effects. Given that we have observed significant estimates in our regressions, this limitation does not pose a significant problem for this study. We infer that our estimates are conservative.

Third, we could not measure the dynamics of helpful votes for each review (i.e., we lacked time stamps on helpful votes and could only observe the total number of votes that had accrued as of the data-collection period). Implicitly, our analyses assume that all helpful votes arrive immediately after one publishes a review. Ideally, we prefer to analyze the arrival of helpful votes dynamically because a review's conformity or deviation will vary over time as others write other, subsequent reviews. In other words, after its writing, a review may be in high or low agreement with all prior reviews but begin to agree with the overall body of opinion as subsequent reviews appear and, thus, affect the rate at which helpful votes arrive. However, this limitation does not pose a serious concern for our analyses. First, we observed a strong positive correlation (rho = 0.89) between the conformity of the author at the time of authorship (i.e., agreement with prior reviews) and the author's conformity to the overall body of reviews that were authored during the data-collection period (i.e., agreement with prior and subsequent reviews). This finding indicates that review deviation and conformity are relatively static values. Second, after repeating our analyses, we observed similar results in terms of signs and significance even after limiting our sample of data to those reviews that were published in the previous two weeks. Therefore, our inability to identify the dynamics of helpful vote arrival unlikely affected our results.

Fourth, we could not determine the degree to which self-selection in review authorship versus the compositional differences that emerged drove the variation in online review characteristics associated with authors' cultural background. In other words, consumers from collectivist cultures are more likely to opt out of reviewing when they are emotionally charged or hold a "different" opinion from prior reviewers. However, this limitation is a serious concern for our study. First, social psychology presents evidence that individuals from collectivist cultures are more susceptible to the opinions of peers and are more likely to conform to such opinions (see Bond and Smith (1996) for a review of this topic). Therefore, differences in the reviewing behavior may exist over and above the decision of whether or not to write a review. Second, despite its presence, self-selection does not have substantive implications for our results or estimates. Our hypotheses and empirical estimations draw relationships between cultural backgrounds and the characteristics of published reviews written by individuals from such backgrounds. We observed systematic differences in the review content regardless of whether one attributes such differences to deviations in opinion conditional on authorship or self-selection into authorship.

This study offers several opportunities for future research. First, future research may investigate different U.S. states as a source of heterogeneity to examine the effect of individualism on online reviews. Second, future studies may examine whether other dimensions of cultural values (e.g., uncertainty avoidance) can affect consumer behavior when producing or consuming online reviews. For example, future research could delve into cross-cultural differences in review consumption under different levels of product uncertainty (Dimoka, Hong & Pavlou, 2012; Hong & Pavlou, 2014). Third, our analyses of perceived helpfulness abstract away the possibility that the effects are moderated by the cultural values of the primary audience for a service provider. For instance, a recent work has provided early evidence of cross-cultural differences in online review consumption by reporting that individuals from collectivist cultures place greater value on negative reviews (Fang et al., 2013). Future studies may explore other differences in perceived helpfulness across cultures, such as whether individuals from collectivist or individualist cultures exhibit a similar preference for review deviation or conformity.

CHAPTER 3

STRATEGIC BEHAVIOR IN MOBILE HEALTH PLATFORMS

3.1 Background

The rapid development of mobile technology has facilitated innovations in health management, bringing in a variety of mobile health applications into the healthcare industry as a potential alternative to traditional health management (Fox and Duggan 2010). The Health Care app is expected to help individuals manage their health, promote healthy living, and provide relevant on-demand medical information. These appliactions have grown rapidly in the last 10 years and the market size is expected to reach 31 billion U.S. dollars by 2020 (Statista 2017). These mobile health applications have served as essential tools to enable cost-efficient health management and medical treatment, and have attracted considerable attention from both researchers and practitioners. As of 2018, mainstream mobile app stores contain more than 165,000 mobile health apps, and the total number of worldwide downloads had reached 3 billion (Dogtiev, 2018). The developing countries are paying much attention to the health eco-system too. The "Healthy China 2030" Planning Outline issued by the State Council stated that the scale of the health service industry should reach 16 trillion yuan in 2030. According to the survey of relevant market institutions, the scale of China's health service industry in 2017 was 4.9 trillion RMB, so a big chasm lays between the two. It can be foreseen that with the upgrading of consumption structure, aging and urbanization, and the further improvement of the medical insurance system, the health service industry will enter a stage of accelerated growth. These data show that the global mobile healthcare market is growing rapidly, and health care applications will play an increasingly important role in the overall health care landscape of countries. This new service model is in line with global's supply-side structural reform requirements, meets the potential needs of consumers, and injects new momentum into economic development.

The huge demand for medical health applications is mainly due to the improvement of people's living standards, the emphasis on health, and the rising incidence of obesity and chronic diseases with abundance of life. Wu Xiaolan, deputy director of the China Center for Aging Research, proposed in the "Actively Advancing the Strategy of Healthy Ageing" that the health service model should shift from the past "disease test" model to the "health maintenance" model (Wu, 2018). Therefore, it is not surprising that mobile applications that perform weight loss, fitness, and health management account for the largest share of all mobile health applications. Despite the promising benefits such mobile health apps have introduced, and despite the analysts' confidence in the market potential of healthcare applications, doctors are still skeptical about the value of such applications to users-- the effectiveness of the apps for individual health interventions is unclear, as previous studies have mixed results on the effect of mobile health apps on users' health outcomes. More importantly, the success of mobile health interventions not only depends on short-term progress but also requires changes to user behavior in the long run (Charness & Gneezy 2009). Moreover, the health applications are different from most mobile applications that they require long-term usage to help users manage and improve their health. Further, these health applications require users to enter a lot of information on a regular basis, docking smart hardware, and long-term use in order to keep abreast of the individual's specific situation, develop an immediate health plan, provide meaningful content output, and more.

Based on the above characteristics of the health care application, this study attempts to explore a health management model based on big data combined with the use of incentive mechanisms. We have established an analytical framework for user behavior to explore the incentive mechanism design to maintain the user's regular use, improve user activity, enhance user interaction, and increase the final impact of the application on user health. Previous academic research has explored many incentives is still known as "one of the oldest and most reliable ways to motivate people (Park 2015)." The usage of financial incentives is widespread in the offline context. For example, in order to attract new customers, gym will offer a free/ low-cost fitness class to attract peoples to experience it (Patel et al., 2011). Similarly, in a smoking cessation program, participants will receive financial rewards if they do not smoke within a specified period of time (Donatelle et al., 2004). However, the reports that financial incentives produce positively impacts are in an offline environment. If you switch directly to an online environment, the results are full of uncertainty (Kwon et al. 2016). Compared to offline environments, financial incentives are used in online environments may move in two distinct directions: On the one hand, financial incentives may have stronger effects when users take advantages of mobile features, such as mobility, flexibility (Ghose and Han 2011), and social connectivity (Yan and Tan 2014). Specifically, mobile devices enable users to upload and download information anytime and anywhere and expand their social connections to people with similar interests or goals. Given that users are more exposed to mobile based information channel than traditional channels such as TV and prints (Ghose and Han 2014), mobile enabled financial incentives could have stronger influences on users. On the other hand, financial incentives on mobile apps may induce unintended strategic behavior—it may encourage users to take "hidden" actions to increase their chances of receiving financial rewards, which may, in turn, lead to failure of incentives (May et al. 2014) .

While previous literature has provided the foundation for understanding the impact of a variety of economic incentives, they have seldom conducted any critical examinations on incentive-induced strategic behavior. Without any remedies, such strategic behavior may not only increase the cost of deploying incentive programs but also decrease the outcome of health intervention and even jeopardize the long-term health statuses of users. Therefore, a deeper understanding of incentive-induced strategic behavior is critical to the success of financial incentives on mobile health apps. In addition, it is worth exploring possible tools to mitigate strategic behaviors in this contexts. With this regard, online social networking features could be such a tool serving as a check on user strategic behavior. In summary, this paper aims to examine the following research questions:

1. Do financial incentives induce strategic behavior of users in mobile health apps?

2. If there is strategic behavior, what is the net effect of financial incentives on user health outcome?

3. Can online social network features mitigate strategic behavior?

We study the above questions by leveraging a quasi field experimental design on one of the leading mobile-based weight management app in China. Since 2013, the mobile app has created several weight loss campaigns with "deposit contracts" to incentivize users to reduce their body weights. To join the campaigns, participants are asked to provide their initial body weight with photographic evidence and deposit 50 RMB into a public money pool. If participants could achieve the goal of losing four percent of their body weight within 28 days after the start of the campaign, they are privileged to share the money in the pool. Otherwise, the deposit will not be refunded. According to previous literature, such financial incentive should incentivize weight loss performance of participants (Gneezy 2011). However, knowing the exact schedule and the threshold of campaigns in advance, the participants can game the incentive program, by over-reporting (it may not be over reporting. Maybe we can say strategic manipulating intentionally or unintentionally) their initial body weight so that it becomes easier to reach the four percent threshold . We leverage quasi field experiments conducted by the mobile app to assess the effect of financial incentives on user behavior. We define the users who registered and participated in the campaign as the treatment group and users who initiated the registration but did not eventually join the campaigns due to payment processing issues as the primary control group. To construct an alternative control group, we include users who had never participated in any campaigns in the past but would participate in at least one of the future campaigns. We show that, in the quasi-experiment, users in the treatment and control groups have identical weight loss progress before the official announcement of the campaigns, teasing out possible self-selection bias. Our identification strategy is to gauge the "additional increase/decrease" in the performance of users in the treatment group, compared to the performance of users in baseline control group. Using a fixed effects difference-in-differences (DID) model, we quantify the short and long-term (post-intervention) impact of financial incentive and identify evidence for strategic behavior.

Our key findings are as follows. First, we find evidence that financial incentives have a positive short-term effect in mobile weight interventions. The results suggest that users in the treatment group lose more weight than users in the control group during the treatment period, with estimated marginal effects ranging from 0.92% to 1.46% of body weight in the four independent campaigns across four seasons of one year. There is an increase in average body weight during the post-intervention period, but overall, users in the treatment group achieved better weight-loss progress than those in the control group. More importantly, we find evidence that users who participate in the campaigns have a slowdown in weight-loss progress relative to users who do not participate in the campaigns between the campaign announcement date and the start of the campaign while the two groups of users have similar weight-loss progress before the announcement date. This finding indicates the existence of strategic behavior. Third, we further uncover that, interestingly, strategic behavior is less prevalent among users with more intensive social networking activities. This suggests that social activities such as tweeting and being mentioned by others may exert monitoring pressure on the users for reducing strategic behavior. Fourth, we find there exist heterogeneous effects for the users with different levels of body mass index (BMI), gender, and age. Finally, we perform several robustness checks: utilize propensity score matching to form matched sample, replicate the analysis with weekly panel data, and use alternative control group settings. All the results are consistent with to the previous analyses.

The remainder of the paper is structured as follows: In the next section, we present a literature review and discuss the current progress in financial incentives and highlight the mobile context. We describe research context, data, and methodology in Section 3. Section 4 presents the empirical study. In Section 5, we discuss contributions and limitations. We conclude the paper in Section 6.

3.2 Literature Review

As this study examines the impact of financial incentive, two streams of research are relevant. We begin with the literature on digital intervention where we highlight the role of mobile IT artifacts in facilitating behavioral interventions in the health management context. We then explore research on financial incentives and their impact on behavioral interventions and discuss recent literature on strategic behavior associated with financial incentives.

3.2.1. Mobile Health Interventions

The advancement of information technologies has enabled novel tools that facilitate a booming digital transformation of the healthcare systems (Agarwal et al. 2010). Meanwhile, IT has also enabled the formation of online platforms where people interconnect to each other. As a result, IT-enabled behavioral intervention platforms have achieved success at an unprecedented scale unconstrained by spatial and temporal restrictions.

Compared with traditional settings (no matter online or offline), mobile-based apps have unique characteristics that affect behavioral interventions. First, recent information systems studies have highlighted several features of mobile devices (e.g., portability and flexibility) that allow users to get access to information content anytime and anywhere (Ghose and Han 2011). In contrast, traditional interventions often suffer from the lack of timeliness or limited accessibility. For instance, Patrick et al. (2009) found that compared to printed materials and meetings (including face-to face and telephone meetings) from a health counselor, text messaging is a more efficient communication tool to encourage weight loss behavior of over-weighed people. Moreover, recent advances in IT and data science provide the unprecedented behavioral informatics and computational modeling for health management (Pavel et al. 2015), which in turn could increase the likelihood of health behavior.

Second, many mobile apps have the advantage in visualization and user behavior analysis. These mobile apps enables users to extract their health data efficiently through self-reporting or digital health monitoring systems (e.g., heart rate monitoring), which empower them to keep track of personal daily activities and progress to the intervention targets. In the weight management context, users track the amount of food and water they consume, the number of steps they walk and run, the amount of time spent on exercises, the number of tasks they accomplish, and so on. In addition, health apps provide granular data, real-time dashboards, and summary statistics on the users' health condition, and facilitate advanced healthcare informatics (Pantelopoulos and Bourbakis 2010). In this regard, the visualize elements from mobile devices is more vibrant than that from paper-based notes or calendars.

Third, incorporate social networking features that enable users with common interests or goals to build online social connections, interact with each other virtually, and receive social support for the community (Yan and Tan 2014). The existent literature on offline social networks has revealed that influential members (e.g., family, friends, and colleges) from one's social network may exert normative influence and support on intentions to health behavior (Cohen-Cole and Fletcher 2008). Barrera et al. (2002) found that the groups with social support have a significant improvement from diabetes, and the effect is stronger when the interventions are given through the Internet. On mobile platforms, social interaction is more accessible since users on the same app usually share the same goals. For instance, Yan and Tan (2014) have shown that social support from online social networks helps generate positive outcomes to health-related activities. Users can befriend with others to establish social connections and observe activities of others, and even exchange information via instant messaging tools (Yan et al. 2015). Babar et al. (2018) illustrate the short-term effects of feedback from users' social network on running performance (Babar et al., 2018). In addition, the achievements of other members of the social network have brought social pressure and demonstration to their peers, making them the same effort. Therefore, social networks for health management have a positive impact on health behaviors. In addition, achievements of other members in social networks create social pressure on their peers to exert the same level of effort. Therefore, social networks for health management have a positive impact on health behavior.

While the above-mentioned positive effects have emerged from the researches of various health platforms, The current evidence of mobile health's impact is still equivocal. Some studies also found insignificant or even negative effects of digital health interventions and other studies cast doubts about the generalizability of the positive findings (Chaudhry et al. 2006). For instance, Cavallo et al. (2012) showed that the use of online self-monitoring and social network groups does not lead to better social support or physical activity outcomes. One plausible explanation is that the impact of a digital intervention is highly contextual, depending on the type of technology, as well as user motivation.

The above review indicates that digital intervention platforms have the potential to play an essential role in health intervention in the future. But, to our knowledge, few extant studies have explored this direction. Although previous studies have considered text-message-based interventions (for a review, see Siopis et al. 2015), the literature on smartphone-enabled intervention is very limited due to the novelty of technology. More importantly, the difference between the two technologies—text messaging vs. smartphone-enabled intervention is significant with smartphone-enabled intervention having the advantage of information richness but requiring user self-monitoring and engagement. For instance, evidence from an experiment suggests that text-message-assisted intervention is effective in weight control (Lin et al. 2014). However, Svetakey et al. (2015)

concluded that interventions delivered by the interactive smartphone application do not lead to superior weight loss performance.

3.2.2. User Behavior in Mobile Apps

Understanding user behavior is important to the design, operation, and effectiveness of mobile applications. Research on mobile application users collects a large amount of activity logs and records from mobile devices and programs. Researchers such as Do collected data on application access, location and Bluetooth from 77 Nokia smartphone users for a period of nine months, and they found that the use of the application depended on the user's location (Do et al., 2011). Their research highlights the importance of identifying the physical and social use environments of built applications. Falaki et al. collected application usage data from 255 Android and Windows Mobile users (Falaki et al., 2010). They found huge differences between users. For example, the average number of smartphone interactions per user per day ranged from 10 to 200, and suggested that the frequency and program of interaction between the application and the user should vary from user to user. Bohmer et al. collected data related to application status information from 4,125 Android users, such as installation, opening and closing (Bohmer et al., 2011). Other studies have revealed interesting application usage patterns, for example, new information is most popular in the morning, and gaming applications are most popular at night. (Rahmati et al., 2012).

Second, high-accuracy, multi-dimensional demographic data helps researchers better understand profiles and explore user interests and preferences. For example, in addition to the activity log of 117 users of the Swiss Nokia N95 smartphone, Chittaranjan et al. also collect demographic information (eg gender, age, nationality) and self-reported personality traits (Chittaranjan et al., 2012). Their research found that male users are more likely to use game function applications, while introverted female participants are more likely to use social function applications (Chittaranjan et al., 2012).

Recent literature has also focused on the study of user behavior in specific applications, such as user activity and churn. For example, Henze et al. released five game apps on the Android market and monitored how apps are used. They collected data from 6,907 users, and research shows that many users have abandoned these applications in a short period of time. The results suggest improving application quality and providing incentives for users to promote long-term use of applications (Henze et al., 2011). Other studies have analyzed network traffic to understand the browsing habits of Internet users (Adar et al., 2008; Obendorf et al., 2007). Researchers have also built more specific user behavior models to study user search intent (Park et al., 2015) and Wikipedia editing model (Geiger and Halfaker, 2013) to predict the performance of crowdsourced workers (Rzeszotarski and Kittur) , 2011), and detect malicious accounts in online social networks (Wang et al., 2013).

User behavior research generally uses click sequence data for mining (Srivasatava et al., 2000). Researchers use simple methods such as Markov chains to capture the user's navigational path on the site (Benevenuto et al., 2009; Lu et al., 2005; Sadagopan and Li, 2008). However, these models focus on simple aspects of user behavior (eg, web pages that users like) and cannot model more complex user behavior. Other methods use clustering techniques to identify groups of users who perform similar activities in their applications (Gunduz and Ozsu, 2003; Su and Chen, 2015; Ting et al., 2005; Wang et al., 2013). The generated clusters can be used to infer user interest (Su and Chen, 2015) or to predict future user behavior (Gunduz and Ozsu, 2003). However, existing cluster-based models are largely supervised (or semi-supervised) and require a large amount of ground truth data samples to train or fine tune model parameters [Sadagopan and Li, 2008; Ting et al., 2005 ; Wang et al, 2013). In addition, many behavioral models are constructed as “black boxes” for classification tasks, with little explanation of how users behave and why (Gunduz and Ozsu, 2003; Wang et al, 2013). Therefore, based on mobile application research, it is urgent to establish an unsupervised click sequence behavior model and explain the model intuitively.

3.2.3 Financial Incentives and Strategic Behavior

Practitioners have often implemented financial incentives in behavioral interventions. There is a pervasive literature attempting to understand how economic incentives alter human behavior or elicit effort. Prior studies have shown that financial incentives have positive influences on a variety of individually or socially desirable behavior, including saving money (Ashraf et al. 2006; Beshears et al. 2015; Thaler and Benartzi 2004), education (Barrera-Osorio et al. 2008), exercise (Acland and Levy 2015; Charness and Gneezy 2009; Milkman et al. 2013), weight-loss (Jeffery et al. 1990; John et al. 2011; Volpp et al. 2008), and smoking cessation (Donatelle et al. 2004; Gine et al. 2008; Volpp et al. 2009). Most of the previous studies, however, were conducted in offline contexts. To our knowledge, extant literature has seldom explored the role of financial incentive in an online mobile setting.

It is widely believed that in the short term, financial incentives (e.g., monetary rewards such as cash and gift) can significantly change people's behavior and encourage more inputs in most cases. For example, in educational studies, financial incentives have been found to increase enrollment, attendance, grades, and graduation rates (Schultz 2004; Behrman et al. 2005; Barrera-Osorio et al. 2008; Barrow et al. 2014). In healthcare contexts, Donatelle et al. (2004) demonstrated that financial incentives enhance short-term smoking cessation and reduction, especially when rewards are considerably large.

In contrast to the short-term effect, researchers have conflicting findings on the impact of financial incentives after their removal. On the one hand, intervention driven changes in behavior may make people form alternative habits (Charness and Gneezy 2009) and overcome initial resistance to engaging in beneficial activities (Angrist and Lavy 2009), indicating that financial incentives may lead to targeted outcomes after the removal of financial incentives. For instance, students who are paid for studying are more likely to attend college, have higher college GPAs, and are more likely to remain in college after their first year (Jackson 2010). Based on this evidence, it is plausible that financial incentives can be a powerful impetus to motivate desirable behavior and enhance performance. On the other hand, economic incentives may lead to unintended consequences and may even become counterproductive in the long run. In the context of education, it is found that the outcome of financial incentives depends on characteristics of the subjects (Bettinger 2012; Angrist and Lavy 2009). For donation behavior, donors' willingness to contribute could be undermined by financial incentive after the incentive is removed (Meier 2007). Similarly, for health behavior, Volpp et al. (2009) found that incentive is only effective when the reward is above a certain threshold, and the high performance vanishes in six months. The long-term effect of financial incentive is still under debate, since extrinsic incentives may conflict with other motivations (Gneezy et al. 2011).

Furthermore, one of the reasons for the unintended consequences induced by financial and other types of external incentives is that economic agents may behave strategically in response to incentive programs. Strategic behavior refers to the use of information asymmetry and the manipulation of data or materials to obtain goals or rewards that would otherwise not be possible (Courty and Marschke, 2004; Oyer, 1998). Economics literature suggests that nonlinear relationship between rewards and performance drives such strategic behavior (Oyer 1998). One example of a strategic behavior is timing behavior —for example, under fiscal-year report systems with a discrete bonus for a given sales quota, salespeople relocate sales performance among different time periods to maximize their payoffs (Oyer 1998). Similarly, Similarly, Kapeller found that authors and editors of academic journals may predict the rules and prejudices contained in JIF calculations, thereby changing their publishing behavior to improve the performance of their journals in rankings (Kapeller, 2010). Bastani et al. (2017) found that healthcare providers often upcode, i.e., mis-report hospital-acquired infections (HAIs) to increase reimbursement or avoid financial penalties. Moreover, Courty and Marschke (2004) documented that training agencies decide to manipulate graduation timing of trainees at the end of the school year to maximize their rewards, which is dependent on the relative position of performance levels towards certain thresholds.

Strategic behavior is found to result in a reduction of individual effort, as well as cause organization inefficiency and welfare loss (Courty and Marschke 2004). Therefore, it is of great importance to avoid the negative effect of strategic behavior and to retain the effectiveness of financial incentives. The most direct remedy to strategic behavior is to monitor the behavior of agents more strictly so that they pay a higher cost to game the incentive contracts. However, monitoring is often costly or infeasible, especially in online or mobile settings and prior literature has rarely considered strategic behavior in such settings. At the same time, studies find abundant evidence of strategic behavior in offline settings. In this paper, we show that one of the possible solutions to strategic behavior lies in leveraging social networking feature on the mobile platforms. Agents who exhibit strategic behavior may be regarded as “unethical” to some extent and risk their reputations in online social networks. Therefore, social norms may pressure them not to behave strategically. In other words, social networking provides a distributed monitoring mechanism that restrain agents from strategic behavior.

3.3 Methodology

3.3.1 Institutional Background and Data

Our study focuses on a common health behavioral intervention practice in the world—weight losing. According to literature, more than one-third (34.9%) of US adults suffer from overweight or obesity (Ogden et al. 2014). Even a modest weight loss—such as a 5% decrease in body weight—can lead to a significant reduction in chronic disease risk (Blackburn 1995; Pasanisi et al. 2001). Therefore, a significant amount of research has been devoted to designing effective interventions for weight management. Digital health management applications provide a useful solution to the problem. As of 2016, major app stores contained more than 165,000 mobile health apps, and the total number of worldwide downloads reached 3 billion (Research 2 Guidance 2016).

We investigate one of the largest digital fitness & weight management apps in China that has a total of 40 million registered users as of June 2016. The free-to-use mobile app has been widely known as the leading digital community for users who are looking for weight intervention or fitness on mobile devices. The platform offers various weight control instructions and diet plans, in addition to user profiles (see Figure 3.1). More specifically, users have visualized dashboards demonstrating daily records of body weight, calories taken from food, calories burned by exercising, and individual weight loss targets. Moreover, the mobile platform adopts social networking features that allow users to follow and be followed by other users, as well as interact with each other by tweeting and mentioning.

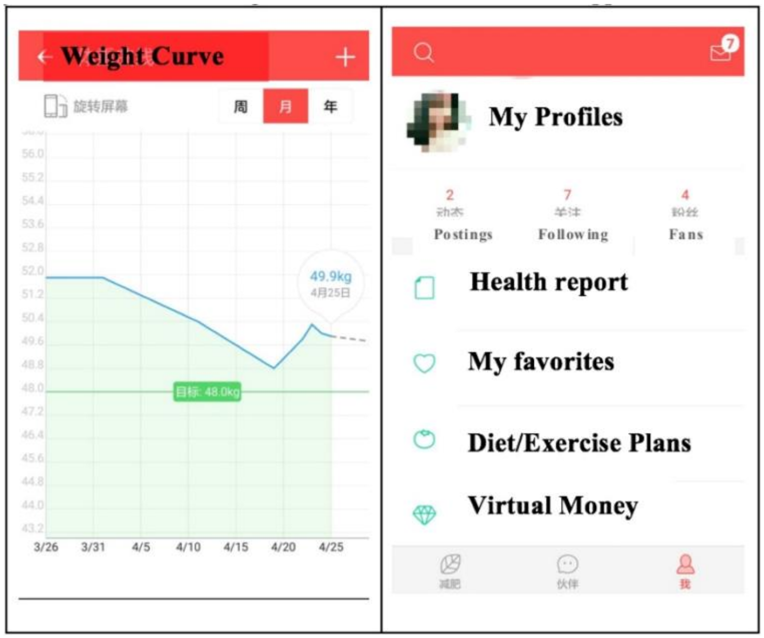


Figure 3.1 Screenshots of the Mobile App (Translated into English Version)

The mobile app holds several weight loss campaigns with financial incentives (i.e., incentive programs) called “I bet I will be slimmer” since 2013. All the users are informed of the campaigns through notifications. To participate in the campaigns, users register by providing required information and depositing a fixed amount of money (i.e., 50 RMB or roughly 8 USD, equals to the price of two working lunches in China) into a money pool. This special incentive setting is also known as “deposit contract” in the economics literature. The campaigns require participants to reach the goal of losing four percent of body weight within 28 days. Participants who reached the threshold by the deadline will share the money in the pool and receive additional gifts (such as a T-shirt) from the company that runs the platform, while participants who have not reached the threshold cannot get a refund (and therefore forfeit the deposit). The platform creates a strict validation process to filter out unqualified users or potential frauds. It requires participants to take high-resolution photos of themselves standing on a weight scale from different angles with clear numbers on the scale, both before and after the campaigns. The incentive program provides us an opportunity to test both the short-term and posterior (post-intervention) effects of monetary reward on health-related behavior.

The design of the campaign creates incentives for strategic behavior among the participants, because of the nonlinear nature of the compensation scheme. Imagine a participant who knows the schedule of the campaigns and the four percent threshold rule for getting the reward, in advance to the campaigns. The participants may want to maximize the probability of winning by adopting the following strategy: hold their weight loss progress or even gain some weight before the campaign. This would make it easier for them to achieve the 4 percent goal. Since this strategic behavior may hinder the users' social image, we suspect that social activities may moderate such strategic behavior: when the participant has intensive exposure in the social network, his/her motivation for such strategic behavior weakens.

To conduct the study, we obtain the complete user activity data from Oct 2013 to July 2016. Our dataset includes daily records on a population of users' demographics, personal weight, diet, and calorie records, social network structures, and social activities. Furthermore, we acquire user's participations (or attempts to participate) in the weightloss campaigns. We also build a social network of users based on their following relationship and social interactions. Eventually, we establish an unbalanced panel dataset from these records.

3.3.2 Identification Strategy

We aim to identify the impact of financial incentives and the potential strategic behaviors associated with them in a mobile environment. One of the major identification challenges of the study is the self-selection bias in participation decisions. We resolve it using a quasi field experimental design, as causal inference requires high similarity between observations in the treatment and control groups. In our context, users sign up for the incentive programs on their own initiative, thus introducing a non-random selection bias. Specifically, it is possible that unobserved user characteristics, such as opportunity costs, commitment to weightloss, etc, would simultaneously lead to participation in the campaigns and weight-loss performance. We take advantage of the registration process to address the self-selection issue. We have complete records of users who had initiated the registration process, regardless of whether they were approved for participants. Therefore, we define the users who participated in the campaign as the incentive group (treatment group) and users who initiated the registration but did not eventually join the campaigns as the control group. This is better than using the users how have never initiated the registration process to form the control group since the two groups would have similar interests in losing weight, and population sizes of the two groups are more balanced. Later, we will show that the two groups demonstrate very similar weightloss patterns before the start of the campaigns but their trajectory differs significantly afterwards.

Nevertheless, there is still a possibility that it suffers from the selection bias since the treatment group eventually attend the program. For instance, users who trust the system may be more likely to deposit money to the platform, and they may also have better weight loss performance. To mitigate this self-selection bias, we further apply propensity score matching (PSM) to match treatment group users with control group users and conduct the same analyses.

We exploit timing assumption for our quasi field experiment design. Since the campaign lasts four weeks, we use four weeks as one period and define a total of four four-week periods. In the Pre-Announcement period, the campaign has not yet been announced; users in both treatment and control groups use the mobile app as usual. Therefore, there should be little differences between the two groups. In the Pre-Intervention period, the platform announces about the campaign users are allowed to register for the campaign. Practically, there are still no external incentives given to either of the group; there should be still no differences between the two groups except strategic behavior. However, since the users in the treatment group know about the threshold rule of the incentive program, they may undertake specific behavior to "game" the incentive contract. During the Intervention Period, the treatment group has the chance to receive the reward by achieving the goal—losing four percent of body weight; the control group does not receive any monetary rewards, even if they reach four percent, and their weight loss effort purely depends on intrinsic or social motivations. In the Post-Intervention Period, for the treatment group, no matter whether they get the monetary reward or not, the financial incentive is removed, and there are no other rewards assigned to them. We keep track of percentage change in body weight of the two groups along the four periods so that the performance difference in the Intervention period reflects the short-term effect of financial incentive, and the performance difference in the Post-Intervention period reflects the long-term (post-intervention) effect. Moreover, the performance difference between the two groups in the Pre-Intervention period indicates the strategic behavior effect. We illustrate the timeline of the quasi-experiment in Figure 3.2, for the third campaign.

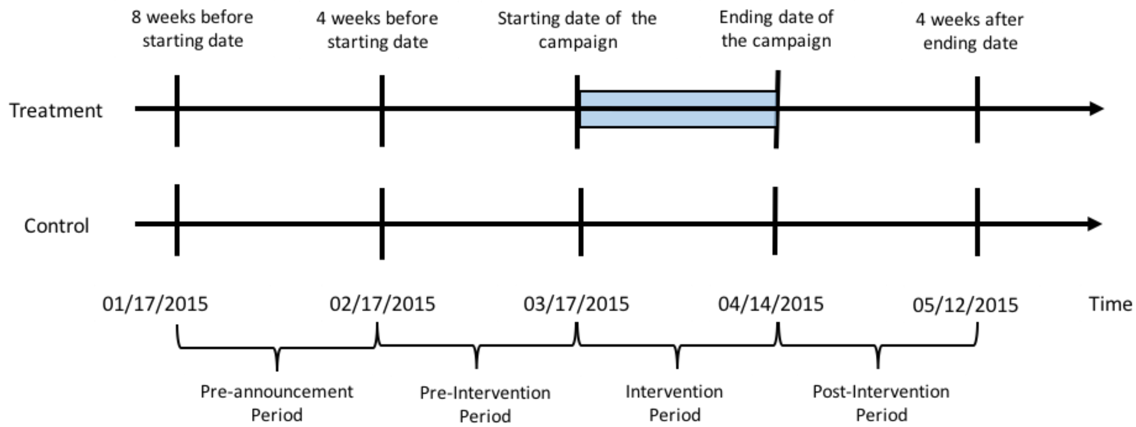


Figure 3.2 Timeline of a Field Quasi-Experimental Design (the 3rd Campaign)

3.4 Empirical Analysis

In this section, we report the steps and results of econometric analyses. Among all the "I bet I will be slimmer" campaigns, we utilize the first four campaigns in our main analysis because all four campaigns have the same deposit requirement, intervention time window, and threshold. Campaigns are held in the four seasons respectively (i.e., fall 2014, winter 2014, spring 2015, and summer 2015), providing season-specific results throughout one calendar year. Table 3.1 summarizes these four campaigns, and for more detailed timelines, please refer to Appendix 1.

Table 3.1 Overview of Four Platform-Sponsored Campaigns

Campaign	# of Participants	Start Date	End Date	Season	Success Rate
1	4205	2014-09-04	2014-10-01	Fall	20.9%
2	4661	2014-12-25	2015-01-22	Winter	24.8%
3	5259	2015-03-16	2015-04-12	Spring	29.6%
4	6955	2015-05-24	2015-06-22	Summer	17.4%
Note: For all campaigns, the deposit is 50RMB; participants have 28 days to reduce 4% of body weight.					

We further split the entire unbalanced panel dataset into treatment and control groups, and present summary statistics of the main variables (except period and group dummies) by the group, as shown in Table 3.2. In total, we have 14,979 users in the treatment groups and 37,076 users in the control groups, while users who have never attempted to join these campaigns are excluded from the main analysis. Importantly, our quasi-experimental design limits dissimilarity in characteristics between the individuals in the treatment and control groups, and yields similar group sizes.

Table 3.2 Summary Statistics (Four Campaigns)

Variable	**Mean**	**SD**	**Min**	**Max**
	Treatment (54,364 Obs.)			
Weight_{it}	62.464	10.460	40.100	130.271
$\mathrm{DiffWeight}_{it}$	-0.268	1.557	-9.900	9.884
$\mathrm{PercentWeightChange}_{it}$	-0.332	2.421	-17.662	17.673
$\mathrm{LogFollowee}_{it}$	3.140	1.238	0.000	7.639
$\mathrm{LogFollower}_{it}$	2.278	1.937	0.000	14.906
$\mathrm{LogPost}_{it}$	1.467	1.282	0.000	6.457
$\mathrm{LogMention}_{it}$	0.652	1.138	0.000	7.305
	Control (125,503 Obs.)			
Weight_{it}	62.414	11.038	40.000	140.000
$\mathrm{DiffWeight}_{it}$	-0.198	1.414	-9.921	10.000
$\mathrm{PercentWeightChange}_{it}$	-0.203	2.167	-20.056	18.888
$\mathrm{LogFollowee}_{it}$	2.846	1.236	0.000	9.814
$\mathrm{LogFollower}_{it}$	1.622	1.896	0.000	16.267
$\mathrm{LogPost}_{it}$	0.693	1.140	0.000	6.863
$\mathrm{LogMention}_{it}$	0.381	0.920	0.000	8.762
Note: Number of users in the treatment group is 14,979 and number of users in control group is 37,076.				

Before conducting regression analyses, we provide some model-free evidence for the impact of the incentives. We compare the average percentage weight change in the two groups in the four seasons in Figure 3.3, where the blue and red-colored bars represent the control and treatment groups respectively. For all four campaigns, the model-free evidence shows that the two groups have similar pattern before the announcement, which confirms the similarity between the two groups. Next, treatment groups reduce significantly less weight in the pre-intervention, indicating the existence of strategic behavior. During the intervention, we observe a much greater reduction in weight among treatment groups than the control group. Remarkably, we observe that treatment groups would increase body weight in the post-intervention periods in some cases.

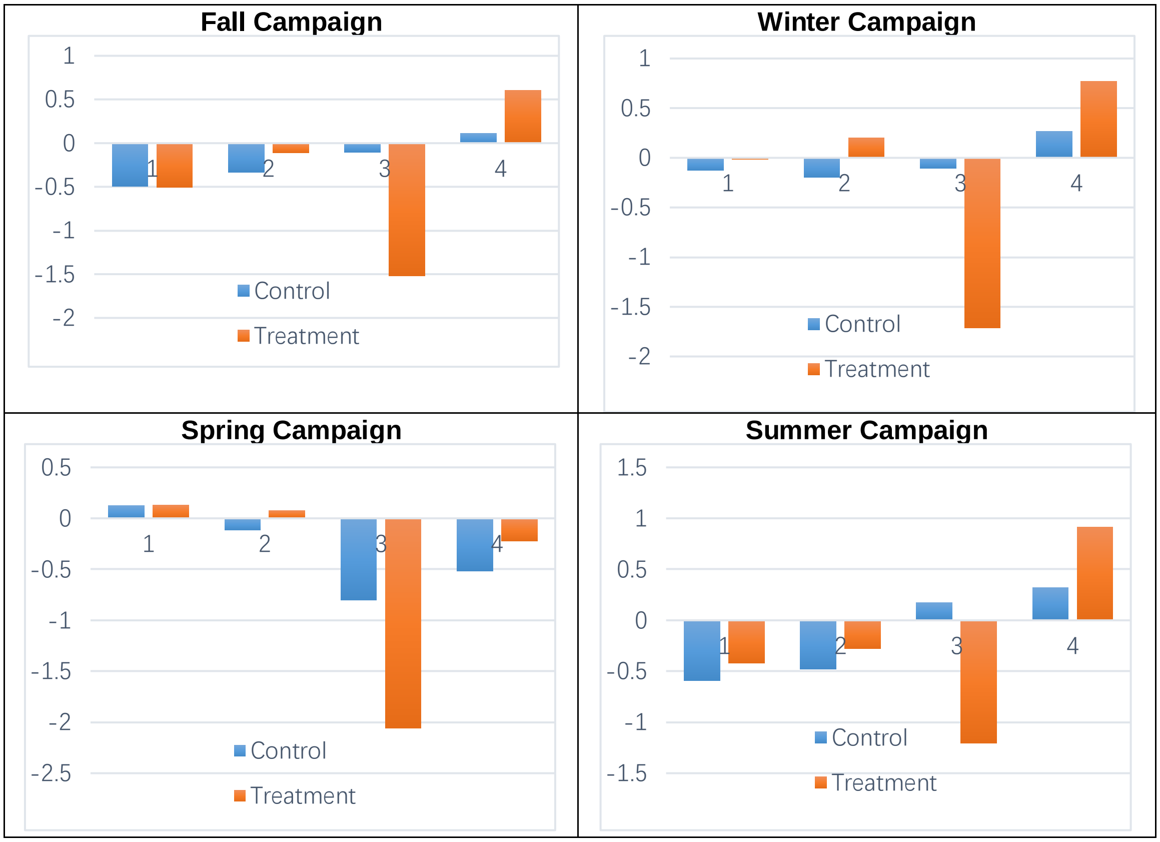


Figure 3.3 Comparing Percentage Weight Change

3.4.1. The Effect of Incentive Program

We then turn to regression analysis to verify the effect of incentive programs in weight management. Our identification strategy relies on comparing the difference in percentage weight change between treated users and untreated users, before, during, and after each campaign. In addition, since social network measures and social activities may have impacts on their weight loss performance as well, they are included in the regression as control variables. In particular, we apply panel data difference-in-differences fixed effects regression model to estimate the impact of a financial incentive on weight loss performance.

$$
\begin{aligned}
PercentWeightChange_{it} &= \beta_1 PeriodDummies_t + \beta_2 Treatment_i \times PeriodDummies_t \\
&+ \beta_3 LogNumFollowee_{it} + \beta_4 LogNumFollower_{it} + \beta_5 LogNumPost_{it} \\
&+ \beta_6 LogNumMention_{it} + \alpha_i + \epsilon_{it}
\end{aligned}
$$

The dependent variable $PercentWeightChange_{it}$ measures the weight loss performance of individual i at time t. In the main analysis, we adopt percentage change in weight as the measure ($PercentWeightChange_{it} = \frac{Weight_{it} - Weight_{it-1}}{InitialWeight_i} \times 100$) where initial weight is the users' body weight at the beginning of the campaign's pre-intervention period. One potential issue with the weight data is sparsity. In our context, body weight values are self-reported by the users, and in online platforms, it is common that very few users are consistently active and report their weight frequently . For users with at least two weight records , we apply a MatLab-based interpolation algorithm to fill in missing daily weight records between two existing records. As a robustness check, we apply "left" interpolation method, in which we use previous weight record values until a new value enters, and the results are highly consistent.

We use period dummies to identify the period (out of four periods) in which the observation was made. Treatment is a binary variable that equals to 1 if the individual is in the treatment group and 0 otherwise. Hence, the coefficients of period dummies capture the percentage weight change of users in the control group, while the coefficients of interactions between period dummies and treatment capture the impact of the incentive program on percentage weight change in each period. Regarding social network measures, we include log-transformed number of followees and number of followers user i has until time t, which are equivalent to out-degree and in-degree of the user in the social network. We also include log-transformed number of tweets and mentions to reflect the intensity of social activities user i has made at time t . Notice that they reflect different social activities, since posting is action taken by the focal users and mention is action taken by their peers. We also include user fixed effects α_i in the model to control for unobserved individual heterogeneity. As a result, the estimated effect of the incentive program is free of time-invariant confounding factors, and the inclusion of other regressors further mitigates the impact of time-varying confounders .

3.4.1.1 Short-Term Effect of Incentives

Result 1 (Short-Term Intervention Effect): Individuals who have participated in financial incentive based campaigns have significantly better weight loss performance in the intervention period.

We present the estimates of the difference-in-differences model upon four campaigns in Table 3.3. In all specifications, the coefficients of period dummies reflect the weight loss patterns of users in the control groups, which vary across different campaigns. Relying on the DID setting, the coefficients of interaction terms represent the additional effect of an incentive program on weight loss performance of users in the treatment groups. We verify the short-term effect since the coefficients of Intervention×Treat in all campaigns are statistically significant (coefficient varies from -0.92 to -1.46, p-values are smaller than 0.001). The marginal effect is an additional reduction of body weight from 0.9 to 1.4 percent in users incentivized by a reward, which is a relatively large magnitude since the weight loss target is four percent.

Table 3.3 Estimation Result of DID

Campaign	**Fall**	**Winter**	**Spring**	**Summer**
Pre-Intervention	0.259***	-0.023	-0.188***	0.160***
	(0.031)	(0.043)	(0.030)	(0.042)
Intervention	0.349***	-0.074	-0.806***	0.569***
	(0.035)	(0.046)	(0.035)	(0.046)
Post-Intervention	0.499***	0.187***	-0.565***	0.628***
	(0.037)	(0.048)	(0.036)	(0.047)
Pre-Intervention ×Treat	0.541***	0.519***	0.570***	0.322***
	(0.066)	(0.073)	(0.070)	(0.068)
Intervention ×Treat	-1.057***	-1.455***	-0.920***	-1.305***
	(0.073)	(0.076)	(0.073)	(0.073)
Post-Intervention ×Treat	0.532***	0.473***	0.388***	0.445***
	(0.071)	(0.075)	(0.071)	(0.071)
LogFollowee_{it}	0.125+	0.026	-0.000	0.183+
	(0.065)	(0.087)	(0.070)	(0.097)
LogFollower_{it}	0.354***	0.398***	0.288***	0.426***
	(0.048)	(0.057)	(0.048)	(0.072)
LogPost_{it}	-0.414***	-0.337***	-0.433***	-0.409***
	(0.021)	(0.026)	(0.024)	(0.023)
LogMention_{it}	-0.127***	-0.170***	-0.186***	-0.297***
	(0.024)	(0.033)	(0.031)	(0.031)
Observations	43,019	30,654	40,065	33,025
Number of users	12,244	8,703	11,107	9,161
R-squared	0.087	0.116	0.111	0.119

Note. The dependent variable is *PercentageWeightChange*. Robust standard errors are under the coefficients.
***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.

3.4.1.2 Evidence of Strategic Behavior

Result 2 (Strategic Behavior Effect): Individuals who have participated in incentive-based campaigns undertake strategic behavior to "game" the incentive contract by retaining their body weight in the pre-intervention period.

We find that the coefficients of Pre-Intervention×Treat are positive and statistically significant for all four campaigns (coefficient varies from 0.32 to 0.57, significant at 0.001 confidence level). It indicates that in the pre-intervention period, percentage weight change increases more (or reduces by a smaller amount) among the users who are incentivized by the reward program compared with the users without incentive. It suggests that the users who participated in the campaigns were influenced by the financial incentive before the start of the campaigns. We attribute the estimated effect to strategic behavior instead of heterogeneity across the two groups as the regression analysis shows that the two groups have similar weight loss trajectory before the campaign announcements. The difference in weightloss performance after the campaign announcement but before the start of the campaign is most likely caused by campaign participants' desire to "postpone" their weight loss performance to the intervention period when performance is counted so that they have higher chance to reach the four percent threshold for the financial incentive.

3.4.1.3 Post-Intervention Effect of Incentives

Result 3 (Post-Intervention Effect): Individuals who have participated in financial incentive based campaigns have even worse weight loss performance in the post-intervention period (a.k.a. there is no post-intervention effect), although the overall long-term effect may be still positive.

We further analyze the post-intervention effect of the financial incentive after the conclusion of the campaigns. We limit the time window to 4 weeks after the intervention because longer time window would reduce the number of qualified users in the analysis. In Table 3.3, the coefficients of Post-Intervention×Treat are significantly positive which represent a smaller weight reduction or a larger weight gain (coefficient varies from 0.39 to 0.53, significant at 0.001 confidence level) i of the treatment group. These results indicate that the financial incentive affect the performance in the opposite direction in the post-intervention period and delimit the beneficial impact of the incentives. More importantly, when we combine the effects in three periods, we observe that two out of four campaigns (the summer and winter campaigns) have an overall positive effect of a financial incentive on weight loss performance. Hence, incentive campaigns may still lead to positive cumulative gains in weight loss performance. From the platform's perspective, the net benefit (Return of Investment) of incentive programs is positive

3.4.2 Moderators for Strategic Behavior

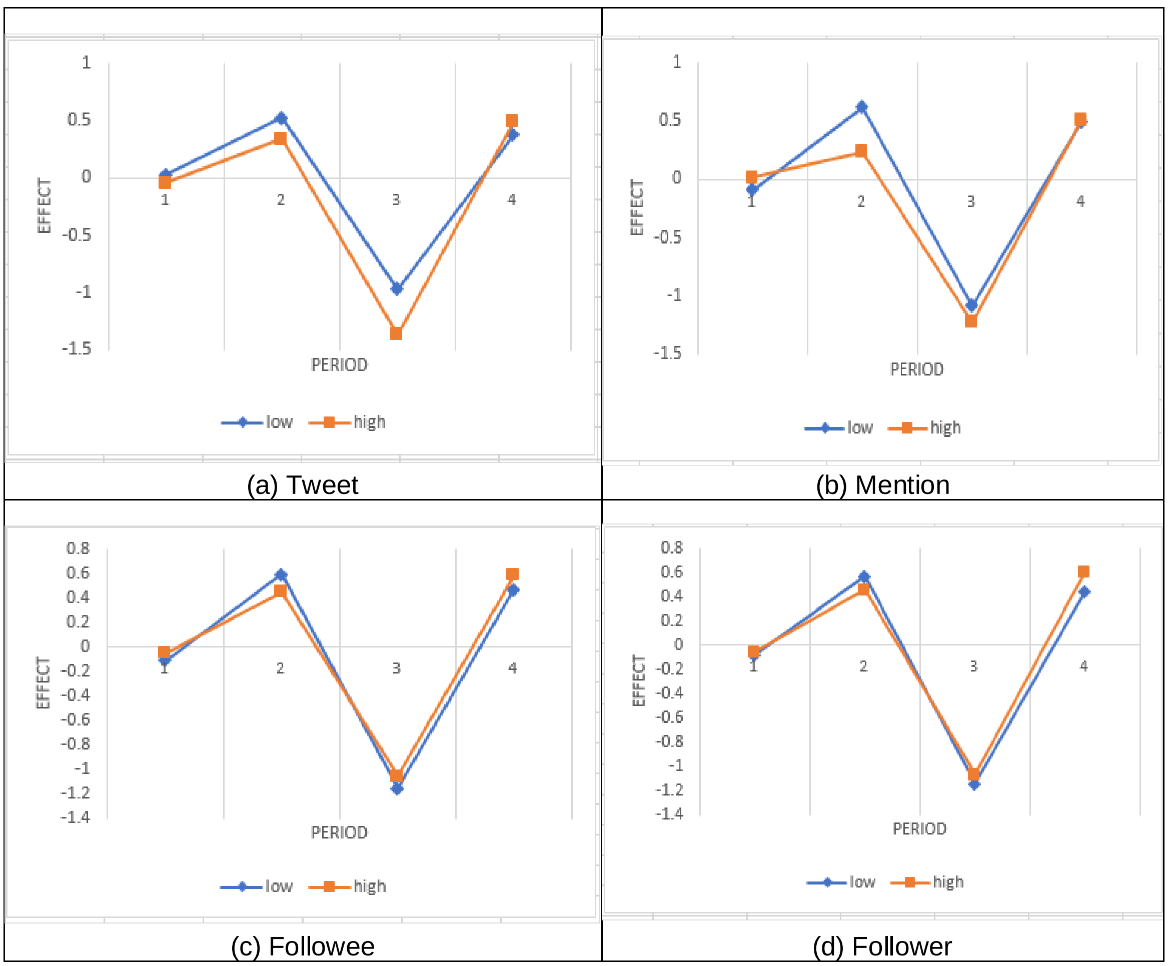
Result 4A: Users who have more social activities undertake less strategic behavior to incentives

Result 4B: Users who have more social connections undertake less strategic behavior to incentives

We next estimate the moderating effect of social networking measures of users on their strategic behavior toward incentives. As users social network structures change little during the campaigns, we fix social network measures for each user-campaign pair at the beginning of the pre-intervention period. We consider two sources of social networking measures in our setting: social activity measures such as number of tweets and number of mentions, and social network structure measures such as in-degree and out-degree of the users (i.e., number of followers and number of followees). We specify the moderators by whether users have high or low social activities as well as high or low number of social connections, to generate a binary variable $Social_i$ (High = 1/Low = 0) for each social networking measure. In the regression model, each of the social networking measures is interacted with $Treatment_i \times PeriodDummies_t$ to construct a three-way interaction term. We therefore identify the differential effect of incentive using the following modified regression model:

$$
\begin{aligned}
&PercentWeightChange_{it} \\
&\quad = \beta_1 PeriodDummies_t + \beta_2 Treatment_i \times PeriodDummies_t \\
&\quad + \beta_3 Social_i \times Treatment_i \times PeriodDummies_t + \beta_4 LogNumFollowee_{it} \\
&\quad + \beta_5 LogNumFollower_{it} + \beta_6 LogNumPost_{it} + \beta_7 LogNumMention_{it} + \alpha_i + \epsilon_{it}
\end{aligned}
$$

Figure 3.4(a) and 3.4(b) demonstrate the moderating effect of social activities regarding number of tweets and mentions (by the others) on the focal user's response to incentives . In the pre-intervention period, there is less strategic behavior when users are active in either of the social activities. In the intervention period, intensive tweets or mentions results in a stronger positive effect of incentive on weight loss performance. Figure 3.4(c) and 3.4(d) show the moderating effect of social connections regarding out-degree and in-degree on the users' response to incentives. We observe a similar pattern in the pre-intervention period as strategic behavior is smaller for users with high social connections, although the differences are much smaller. Furthermore, our analysis shows that weight loss performance for users with high social network connections do not differ significantly from those with low social network connections during the intervention period.



(a) Tweet (b) Mention

(c) Followee (d) Follower

Figure 3.4 Comparison between High and Low Social Network Characteristics

Table 3.4 reports the detailed estimates of the difference-in-differences model with the three-way-interaction terms. From Spec. 1 and 2, we observe that participants who post more tweets and who are more frequently mentioned by others undertake less strategic behavior in the pre-intervention period, better weight loss performance in the intervention period, and regain more body weight in the post-intervention period. The result suggests that, in the pre-intervention period, users with more tweets and mentions are less likely to “game” the incentive system as they are more monitored by the users in the online community. During the intervention period, users who post more tweets want to become popular in the community, and thus they have a stronger motivation to achieve the goal of the incentive programs. On the contrary, users with more mentions are already popular and have lower motivation to do so. Interestingly, users with more social activities are more likely to gain weight after the campaigns. Relatively, socially motivated users significantly reduce motivation after the campaigns terminate. Likewise, Spec. 3 and 4 compare the estimation results for users with a high and low number of followee and followers, the estimates also indicate a lower level of strategic behavior among the users with more social connections.

Table 3.4 Estimation Result of DID: Differential Effects by Social Networking Features

Specification	**1**	**2**		**3**	**4**
Moderator	Tweet	Mention	Moderator	Followee	Follower
Pre-Intervention×Treat	-0.186+		Pre-Intervention×Treat	-0.144*	
×High Post	0.070		×High Followee	0.067	
Intervention×Treat	-0.394***		Intervention×Treat	0.102	
×High Post	0.076		×High Followee	0.072	
Post-Intervention×Treat	0.116		Post-Intervention×Treat	0.118	
×High Post	0.075		×High Followee	0.073	
Pre-Intervention×Treat		-0.386***	Pre-Intervention×Treat		-0.114+
×High Mention		0.075	×High Follower		0.066
Intervention×Treat		-0.140+	Intervention×Treat		0.075
×High Mention		0.080	×High Follower		0.072
Post-Intervention×Treat		0.019	Post-Intervention×Treat		0.163*
×High Mention		0.080	×High Follower		0.072
Campaign Fixed Effects	Yes	Yes	Campaign Fixed Effects	Yes	Yes
Observations	146,763	146,763	Observations	146,763	146,763
# of users	36,394	36,394	# of users	36,394	36,394
Note. The dependent variable is *PercentageWeightChange*. Robust standard errors are under the coefficients. ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.					

3.4.3 Heterogeneous Effects

We have verified that financial incentive programs exert strong short-term effects but no post-intervention effects on weight loss behavior. One might argue that the weight gain of post-intervention effects is because it is harder for the slimmer users to reduce body weight further. In the same argument, one might argue that this campaign might not be so fair since the heavier obese users may lose weight easily compared to the slimmer users. If the claim about the differential effect of incentives due to distinctive body shape is correct, then we should observe the coefficients of interactions decrease smoothly after the end of the campaign. The effect of financial incentives may vary across users and we conduct additional analysis to study the heterogeneity. Our DID model is as follows:

$$
\begin{aligned}
PercentWeightChange_{it} & \\
&= \beta_1 PeriodDummies_t + \beta_2 Treatment_i \times PeriodDummies_t \\
&+ \beta_3 X_i \times Treatment_i \times PeriodDummies_t \ + \beta_4 LogNumFollowee_{it} \\
&+ \beta_5 LogNumFollower_{it} + \beta_6 LogNumPost_{it} + \beta_7 LogNumMention_{it} + \alpha_i + \epsilon_{it}
\end{aligned}
$$

We include interaction terms of BMI, age, gender with $Treatment_i \times PeriodDummies_t$, comparing users in the incentive and control groups. We create a set of dummy variables for each characteristic, as we label the subsample of users with their average body mass indices (BMI)—Low for below 25th quantile, Medium for between 25th and 75th quantiles (medium), and High for above 75th quantile. Likewise, we classify the users into different age groups—Low for less than 22, Medium for 22-28, and High for above 28, as well as into different genders—Unknown, Female, and Male. We introduce the sets of dummies to the regression model one-at-a-time, so that the heterogeneous effect is evaluated with all other factors equal.

Result 5: There is BMI, gender, and age difference in users' short-term response to financial incentives. However, gender is the only factor that leads to differential strategic behavior effect and post-intervention effect of the incentive program.

From Spec. 1 of Table 3.5, users with high BMI perform have worse performance in the intervention period (coefficient=0.576, p-value<0.001), meaning that high BMI users achieve worse performance during the intervention period than low BMI users. The analysis also shows That, in the pre- and post-intervention periods, there is no significant difference in weight loss performance for users with different BMI levels. To sum up, for users with different BMI, the short-term effect of financial incentive is heterogeneous, while the post-intervention effect and strategic behavior are homogenous.

Table 3.5 Estimation Result of DID: Heterogeneous Effects by Demographics

Specification	**1**		**2**		**3**
Moderator	BMI	Moderator	Gender	Moderator	Age
Pre-Intervention ×Treat ×Medium BMI	0.123	Pre-Intervention ×Treat ×Female	-0.073	Pre-Intervention ×Treat ×Medium Age	0.177
	0.160		0.092		0.139
Pre-Intervention ×Treat ×High BMI	0.028	Pre-Intervention ×Treat ×Male	-0.909***	Pre-Intervention ×Treat ×High Age	0.104
	0.181		0.256		0.151
Intervention ×Treat ×Medium BMI	0.094	Intervention ×Treat ×Female	-0.210*	Intervention ×Treat ×Medium Age	0.328*
	0.169		0.097		0.152
Intervention ×Treat ×High BMI	0.576**	Intervention ×Treat ×Male	-0.663*	Intervention ×Treat ×High Age	0.422*
	0.198		0.292		0.164
Post-Intervention ×Treat ×Medium BMI	-0.002	Post-Intervention ×Treat ×Female	-0.230*	Post-Intervention ×Treat ×Medium Age	0.033
	0.168		0.095		0.159
Post-Intervention ×Treat ×High BMI	0.236	Post-Intervention ×Treat ×Male	-1.153***	Post-Intervention ×Treat ×High Age	0.002
	0.204		0.316		0.171
Campaign Fixed Effects	Yes	Campaign Fixed Effects	Yes	Campaign Fixed Effects	Yes
Observations	41,366	Observations	100,332	Observations	71,287
Number of users	10,520	Number of users	25,847	Number of users	19,076

Note. The dependent variable is *PercentageWeightChange*. The cutoffs for Medium and High age are 22 and 28, respectively. Robust Standard errors are under the coefficients. ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.

We next compare users with different demographic background such as gender and age. Specifically, from Spec. 2, we find that male users have less strategic behavior in the pre-intervention period a stronger response to financial incentive during the intervention period and are more likely to keep their weight loss progress compared to female users. From Spec. 3, users above the median age (age>22) tend to lose less weight during the intervention period, potentially because they are less motivated by financial incentive as they have much higher income than low aged users.

3.5 Robustness Checks

3.5.1 Propensity Score Matching

Our identification strategy relies on the assumption that control group users are not significantly different from treatment group users so that the comparison is valid. As mentioned, users in the control groups have initiated the registration process but eventually have not participated in the campaigns. Our discussion with the platform reveals that most of the incomplete registrations were due to security concerns or unfamiliarity with mobile payment. This leads to potential endogeneity issue as certain user characteristics could be more sensitive to security concerns or less familiar with mobile payment and such characteristics could be correlated with their weight loss effort. To mitigate endogeneity, we further leverage the observed characteristics of users to construct propensity score matched panel datasets and conduct the same regression model, like previous studies such as Xu et al. (2016). More detailed information about the matching process is presented in Appendix 2. The comparison of key variables between the treatment group and matched control group is shown in Table A1, which displays the similarity of the users in the two groups.

We present the results in Table 3.6. Notably, the short-term, post-intervention and strategic behavior effects of incentive programs are very similar to the estimates without matching. the result indicates that users who have participated in the campaigns do not have significantly different motivation for weight loss behavior than users who initiated the registration process but dropped out from the campaign before the campaigns started.

Table 3.6 Estimation Result of DID: Matched Data

Campaign	Fall	Winter	Spring	Summer
Pre-Intervention	0.359***	0.199*	-0.086	0.279***
	(0.067)	(0.079)	(0.063)	(0.070)
Intervention	0.333***	-0.156*	-0.776***	0.637***
	(0.078)	(0.076)	(0.072)	(0.074)
Post-Intervention	0.412***	0.167*	-0.569***	0.607***
	(0.075)	(0.078)	(0.071)	(0.075)
Pre-Intervention ×Treat	0.411***	0.302**	0.497***	0.196*
	(0.090)	(0.101)	(0.090)	(0.088)
Intervention ×Treat	-1.058***	-1.363***	-0.918***	-1.363***
	(0.104)	(0.099)	(0.096)	(0.094)
Post-Intervention×Treat	0.608***	0.500***	0.408***	0.485***
	(0.098)	(0.099)	(0.094)	(0.091)
$\mathrm{LogFollowee}_{it}$	0.171+	-0.023	0.008	0.251+
	(0.093)	(0.106)	(0.104)	(0.138)
$\mathrm{LogFollower}_{it}$	0.327***	0.411***	0.295***	0.338***
	(0.067)	(0.072)	(0.063)	(0.087)
$\mathrm{LogPost}_{it}$	-0.379***	-0.344***	-0.461***	-0.395***
	(0.030)	(0.034)	(0.032)	(0.030)
$\mathrm{LogMention}_{it}$	-0.153***	-0.162***	-0.200***	-0.289***
	(0.035)	(0.042)	(0.042)	(0.039)
Observations	21,102	24,237	20,510	26,355
Number of users	5,715	6,582	5,504	7,097
R-squared	0.119	0.131	0.159	0.128

Note. The dependent variable is *PercentageWeightChange*. Robust standard errors are under the coefficients.
***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.

3.5.2 Main Effect with Weekly Panel Data

In our earlier analysis, we aggregate performance data into four periods and analyze pre-announcement, post-announcement-but-pre-intervention, and post-intervention effects of financial incentives. We find that the users exhibit strategic behavior, given the nonlinear financial incentive in which users get rewards for achieving the predetermined performance threshold. Our analysis suggests that users "hold" or "over-report" their body weight in the pre-intervention periods (for example, by reporting their body weight after meals). To strengthen our empirical results, we break down each time period into more granular time intervals to obtain the weekly effect of incentive programs. Ultimately, we have at most 16 periods for each user, and therefore include 15 week dummies in the regression model. In Table 3.7, we present the estimation result of weekly panel data model. Also, Figure 3.5 illustrates the trends of the coefficients of the interaction terms.

Table 3.7 Estimation Result of DID using Weekly Panel Data

Campaign		Fall	Winter	Spring	Summer
Pre-Announcement	Week2×Treat	0.004	-0.008	0.021	-0.012
		(0.029)	(0.026)	(0.025)	(0.033)
	Week3×Treat	0.025	0.048+	0.012	0.023
		(0.028)	(0.027)	(0.026)	(0.034)
	Week4×Treat	0.054+	0.043	0.042+	0.020
		(0.029)	(0.028)	(0.025)	(0.034)
Pre-intervention	Week5×Treat	0.040	0.019	0.085**	0.060+
		(0.028)	(0.028)	(0.026)	(0.034)
	Week6×Treat	0.088**	0.088**	0.073**	0.025
		(0.029)	(0.028)	(0.028)	(0.035)
	Week7×Treat	0.220***	0.251***	0.248***	0.278***
		(0.029)	(0.031)	(0.030)	(0.036)
	Week8×Treat	0.086**	0.093**	0.028	0.088*
		(0.032)	(0.035)	(0.034)	(0.037)
Intervention	Week9×Treat	-0.287***	-0.376***	-0.363***	-0.228***
		(0.031)	(0.031)	(0.031)	(0.035)
	Week10×Treat	-0.262***	-0.349***	-0.145***	-0.135***
		(0.030)	(0.030)	(0.030)	(0.034)
	Week11×Treat	-0.283***	-0.390***	-0.150***	-0.212***
		(0.029)	(0.030)	(0.029)	(0.034)
	Week12×Treat	-0.279***	-0.355***	-0.257***	-0.315***
		(0.030)	(0.031)	(0.030)	(0.035)
Post-Intervention	Week13×Treat	0.138***	0.139***	0.134***	0.079*
		(0.029)	(0.030)	(0.029)	(0.034)
	Week14×Treat	0.146***	0.083**	0.096**	0.125***
		(0.030)	(0.029)	(0.030)	(0.033)
	Week15×Treat	0.104***	0.047+	0.081**	0.051
		(0.028)	(0.028)	(0.030)	(0.032)
	Week16×Treat	0.072**	0.073**	0.002	0.039
		(0.028)	(0.028)	(0.030)	(0.033)
Observations		170,720	120,425	156,204	119,029
Number of users		11,951	8,384	10,579	8,289
R-squared		0.036	0.046	0.053	0.044

Note. The dependent variable is *PercentageWeightChange*. Week1×Treat is omitted from the regression. Robust standard errors are under the coefficients. For brevity, we only report the coefficients of the interaction terms.
***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.

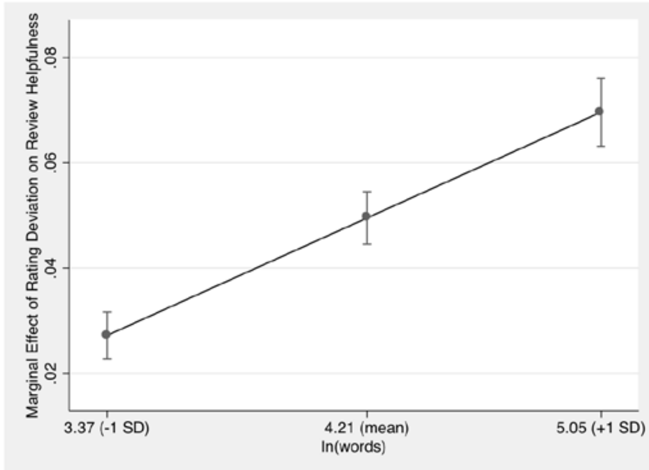


Figure 3.5 Graphical Presentation for Coefficients of Interaction Terms

We consistently obtain significant negative coefficients for the interactions between treatment and dummy variables for Intervention Period (Week 9 to 12), showing that incentive programs have a positive effect of motivating users' weight loss behavior. More importantly, the interactions for the Pre-Announcement Period (Week 2 to 4) consistently have insignificant coefficients, meaning the treatment group and the control group are statistically similar in their weight loss behavior before campaign announcements. The result provides evidence that our comparison between the treatment and control groups is valid since there is no effect in the pre-announcement period. Rather, the coefficients for Pre-Intervention Period (Week 6 to 8) are mostly significantly positive (in Week 7 the coefficients are from 0.22 to 0.28, p-values are smaller than 0.001), providing evidence of strategic behavior since most of the participants register for campaigns in these weeks. In addition, weekly panel data analysis demonstrates that after the weight loss campaigns terminate, participants exhibit undesirable side effects of incentive programs—participants in the treatment group reduce less weight (or even regain body weight) in the post-intervention period.

3.5.3 Main Effect with Alternative Control Groups

We show above that our treatment and control groups demonstrate parallel trend in weight loss behavior before campaign announcements. To further strengthen our result, we identify an alternative control group that consists of users who have never attended any campaigns so far but attended at least one of the future campaigns. Table 3.8 shows the primary effect results for alternative control group setting that there is a high and positive impact of the financial incentive on weight loss performance and a strong strategic behavior. The magnitude of the effects is very close to the results in Table 3.3.

Table 3.8 Estimation Result of DID: Alternative Control Groups

Campaign	**Fall**	**Winter**	**Spring**	**Summer**
Pre-Intervention	0.307***	0.065**	-0.119***	0.289***
	(0.027)	(0.024)	(0.027)	(0.031)
Intervention	0.386***	-0.069**	-0.499***	0.408***
	(0.029)	(0.026)	(0.030)	(0.033)
Post-Intervention	0.418***	0.077**	-0.456***	0.451***
	(0.031)	(0.028)	(0.031)	(0.037)
Pre-Intervention ×Treat	0.482***	0.500***	0.532***	0.252***
	(0.066)	(0.064)	(0.069)	(0.061)
Intervention ×Treat	-1.090***	-1.366***	-1.165***	-1.110***
	(0.073)	(0.067)	(0.071)	(0.066)
Post-Intervention×Treat	0.640***	0.667***	0.344***	0.621***
	(0.070)	(0.068)	(0.070)	(0.067)
$\mathrm{LogFollowee}_{it}$	0.072	0.015	-0.024	0.142+
	(0.062)	(0.076)	(0.068)	(0.084)
$\mathrm{LogFollower}_{it}$	0.317***	0.264***	0.219***	0.438***
	(0.049)	(0.057)	(0.053)	(0.063)
$\mathrm{LogPost}_{it}$	-0.366***	-0.373***	-0.443***	-0.484***
	(0.022)	(0.025)	(0.026)	(0.024)
$\mathrm{LogMention}_{it}$	-0.169***	-0.155***	-0.207***	-0.278***
	(0.028)	(0.032)	(0.031)	(0.033)
Observations	39,421	42,725	38,970	37,435
Number of users	10,276	11,150	10,081	9,800
R-squared	0.098	0.112	0.126	0.109
Note. The dependent variable is PercentageWeightChange. The control groups include those users who have not yet attended any campaigns but attended at least one of the future campaigns. Robust standard errors are under the coefficients. ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.				

3.6 Contributions and Implications

We examine the impacts of financial incentive and the associated strategic behavior in a mobile-based intervention practice. Our study provides evidence that financial incentive positively enhances weight loss progress during the treatment period. However, it leads to strategic behavior on user performance in the pre-intervention period, right before the incentive is implemented.

Moreover, we find that intensive usage on the social features has a moderating effect on the strategic behavior. In other words, participants have more social connections, and social activities are less likely to perform strategic behavior. In this section, we discuss the theoretical and managerial implications of our finding for health behavior interventions, with a focus on the mobile environments.

3.6.1 Theoretical Contributions

First, our research further sheds light on the use of financial incentives. Drawing on behavioral economics literature on incentives, we empirically find that agents who face performance-contingent financial incentives have significantly improved performance during the intervention period. However, after the removal of the financial incentive, the lack of follow-up stimuli causes agents to regress to the original performance levels in the long run.

Second, we provide a close look at the strategic behavior induced by financial incentives in the context of mobile based health management apps.. Importantly, we show that strategic behavior happens after the announcement of the financial incentive but before its implementation. Our finding highlights the importance of taking strategic behavior into consideration in assessment of incentive programs. Previously, researchers often quantify the effect of the financial incentives by comparing outcomes right before and after interventions. Our findings show that a part of the reduction in body weight during the intervention can be attributed to the mere timing shift in weight loss progress. Therefore, failing to capture the potential strategic behavior could lead to over-estimation of the short-term effect. Moreover, the mixed results of the post-intervention effect in previous studies may also due to the lack of consideration into strategic behavior.

Third, we find evidence that the usage of social networking features of the mobile app influences users' effort and strategic behavior. Our finding is consistent with the theories of social presence. Our result suggests that strategic behavior is less likely to occur in the users who were socially active. We suspect that social presence stands as additional monitoring source and may trigger spotlight effect, and participants are therefore less likely to engage in strategic behavior.. Moreover, consistent with previous studies, our research suggests that weight loss outcome is hard to sustain

(Charness and Gneezy 2009). In this regard, the results suggest that weight management platforms integrate socially-driven group incentives, in addition to financial incentives to encourage weight control behavior.

Finally, we explore the effects of financial incentives on behavioral intervention performance based on field data from a mobile app, using a quasi-experiment as the identification strategy. That is the significant difference between this study and previous studies that applied surveys or lab experiments to address research questions. Although survey studies allow researchers to measure personal perceptions of subjects and lab experiments enable researchers to randomize subjects to establish causal relationships, each has its own weaknesses. Observational studies have advantages in providing alternative evidence from the field. Notably, we provide the empirical evidence based on users' actual reaction to the financial incentives which may not reflect in the survey setting. Moreover, the field data allow exploring a real-world social network of users, which is hard to simulate in lab experiments.

3.6.2 Managerial Implications

This study has various managerial implications. First, we provide many insights to the practitioners about financial incentives for health intervention. According to our findings, the short-term return of the deployment of financial incentives is successful. However, it is very challenging to minimize strategic behavior while still achieving successful interventions with high long-term user engagement. As a result, the design of incentive programs is the key to this question. Given incentive contracts with self-reported performance, the level of strategic behavior depends on the way of recording their initial body weight, the schedule of campaigns, and the performance-pay scheme. In this context, the platform should use the participants' longer time body weight trend before the campaigns to calculate the "initial weight" to avoid any possible "data manipulation" before the campaigns. Moreover, the platform should try to lower the users' perceived difficulty toward the threshold, by highlighting progress to the targets and encouraging users to stick to daily fitness plans or breaking down one long campaign into several short ones.

Second, practitioners need to provide social networking features for users to build relationships, exchange information, and obtain social support, to trigger social pressure. Users appear in the social networks not only for information needs, such as to search for new diets and exercise tips but also for social needs, such as to communicate and track their friends' progress. We highlight the importance of social features in reducing strategic behavior. Therefore, the practitioners should help the users to build social connections and encourage social interactions. Moreover, the practitioners should come up with strategies to create social incentives to enhance the long-term performance. For example, the practitioners may design more socially-driven incentive campaigns, in which a small group of users (e.g., spouses or close friends) achieve the goal when everyone in the same group makes enough amount of progress. Furthermore, our empirical result has supported the idea that weight loss performance is difficult to sustain in the long run. Socially related incentives may potentially play an important role in maintaining motivation. The practitioners may provide reputational incentives, by developing rankings and badge systems (Anderson et al. 2013). Consequently, users receive a positive social image by obtaining high ranks or rare badges, which may lead to desirable behavior.

3.6.3 Future Research

Future research may explore the impact of the different amount of monetary rewards in mobile-based intervention apps. Except that, our data has no variations in the type of rewards. The weight loss campaigns we study always apply combined deposit and monetary reward. Therefore, it was difficult to distinguish the effect of monetary reward from the pure "deposit contract." Third, due to confounding factors such as homophily and simultaneity, we are not able to interpret the result of social activities as causality (Hartmann et al. 2008). Due to the complexity of social effect models, we believe that it could be a future study to quantify the effect of social interactions on weight loss performance.

There are several avenues for further research. First, future work could conduct a randomized field experiment on the weight intervention app to explore social effect in the social network. For example, future research can change "I bet I will be slimmer" setting into a group setting of "We bet we will be slimmer together." That is the setting that allows users to compete for reward as a group, which additionally involves social motivation. An interesting question is whether group financial incentives will lead to additionally higher performance compared to individual financial incentives. Since there are heterogeneous nodes in a social network, the second interesting question is whether the group formed by friends outperform the group formed by strangers. Users in the groups with friends are, ceteris paribus, presumably more motivated, since they do not want to reduce their reputation in front of their friends and they may have more interactions with their friends, compared to strangers.

3.7 Conclusion

We examine the impacts of financial incentives on weight loss performance under a mobile health app setting. Through the empirical study, we obtain the following key findings. First, financial incentive programs directly and positively affect the weight loss performance in the short-term, but there is a negative post-intervention effect. Second, financial incentives lead to strategic behavior, represented as a timing shift of weight loss performance. Third, the user's social activities and social network connections moderate the strategic behavior. Our study contributes to the literature on economics incentives for behavioral intervention under performance-contingent financial incentives with self-reporting of performance. We also provide several practical implications for mobile app developers to enhance the incentive programs and designing the features. The implications of this study are not only limited to mobile-based health management but also generalized to many IT-enabled behavioral interventions.

CHAPTER 4

PREDICTING FINANCIAL RISK USING NON-FINANCIAL DATA

4.1 Background

Using "big data" to predict consumer behavior is becoming increasingly important for modern firms. Methods such as data mining and machine learning have been widely used to achieve this task. This approach usually works well when the available big data is closely related to the outcome of interest. However, in many cases structured relevant data is deficient or totally unavailable. Therefore, practitioners and researchers are endeavoring to take advantage of unconventional or seemingly irrelevant data to achieve the same purpose (Zhang, et al., 2016). To solve this practical question and enlighten the design science theory of using unconventional data, this paper advances a design science approach to demonstrate how to build a predictive model which can utilize unconventional data in the context of online lending. We choose online lending as the problem domain because it is a long-lasting topic in the consumer finance industry and this industry is among the pioneer industries that utilize big data technologies. To predict online loan default risk, existing models mainly take advantage of highly relevant data, such as FICO score, payment history, income, default history, credit utilization, etc. (Lessmann et al. 2015). Although these models have been strengthened by advanced machine learning methods and remain widely used in the finance industry, they have two major issues: (1) structured credit data is not always available so they are not as useful when the loan applicants have thin credit data; (2) the predictive power of these models is capped because these models don't consider other data sources. One way to solve these two problems is to involve non-financial data/unconventional data into these models, which is the focus of this study. Several existing studies have examined this topic from different perspectives. For example, Lin et al. (2013) finds that friendship network can be used to infer loan default risk. However, our study proposes a unified design science framework on multiple data sources (rather than one factor or one area), details the data processing and model building procedures, and tests it with real business data.

The research objective of this study is to develop a predictive model to evaluate online lending default risk using non-financial data. The major design challenges are how to extract useful features from non-financial data and how to use these features to predict loan default risk. To solve these two problems, we adopt a design science approach and utilize predictive analytics as the kernel theory. Predictive analytics refer to “the building and assessment of a model aimed at making empirical predictions.” (Shmueli and Koppius 2011 P555). It includes two components, which are the empirical predictive models and the methods for evaluating predictive power. It also provides a well-defined procedure to create a predictive model, which exactly deals with the design challenge of this study. Based on predictive analytics framework, we first exact new features from three non-financial data sources, i.e. within-app browsing data, short message data, and social network data, and then build first-layer predictive model within each data source. As we will show later, specific first-layer classifier is selected to fit the unique characteristics of each data source and feature structure. We further combine individual predictive models into one second-layer predictive model and use it as the final design artifact. At last we evaluate our design choice at each step and the final artifact with real business data.

The theoretical contributions of this study are two-fold. First, it contributes to design science theory by demonstrating how to exact useful features and build predictive models from unconventional data. This study finds that theory-based features have better predictive power than raw features and that specific predictive method should be used to fit different data sources and feature structures. Second, it contributes to the theory of consumer finance by exploring new connections between non-financial features and loan default risk. Because the features used in the predictive model are generated from non-financial data, they are not likely to be identified in a traditional theory-building approach and may provide insights on understudied causal connections. This study finds that loan default risk is related with (1) how customers apply loans with smartphone apps, (2) how customers interact with financial institutions through short messages, and (3) how customers are connected with each other in a social network. The practical implication of this study is that it helps build a more powerful loan default risk predictive model, which can not only reduce credit risk of loan issuing institutions but also increase financial inclusion for customers with thin credit features. In the following sections, we organize this paper according to suggestions from Gregor and Hevner (2013) and Goes (2014).

4.2 Literature Review

4.2.1 Problem Domain: Predicting Financial Risk

Consumer finance is a long-lasting topic in the finance industry and it is important for financial health of individuals and households (Tufano 2009). The efficiency of consumer finance market largely depends on the prediction accuracy of consumer finance risk. Overestimation of consumer financial risk would lead to an undersupply of money while underestimation of consumer financial risk would lead to a high level of bad rate. Therefore, increasing prediction accuracy of consumer financial risk is important not only to financial institutions but also to customers.

This study focuses on online consumer loans, which are normally unsecured short-term loans. Online loans typically work in the following way: (1) customers first apply for a loan through websites or mobile apps by submitting their identity information and the requested loan amount; (2) online lenders then determine the loan default risk of each customer based on proprietary predictive models and automatically make underwriting decisions (Wang and Overby 2018a). Some online lending platforms may directly fund those qualified customers while in other platforms online investors would be involved to use their money to fund those customers. However, in either way the predictive model used in step 2 is essential for online lending industry to process loan applications. Because online loans are unsecured and lending decisions are made within seconds, how to increase prediction accuracy is a huge challenge to online lending platforms. According to American Bankers Association, online lending volume in the United States could reach $90 billion by 2020. However, the default rate of online loans usually ranges from 10% to 20%, which is far higher than traditional consumer loans. Therefore, loan default risk prediction is a significant topic for the whole industry. Improving loan default risk prediction can also provide insights to other financial products such as mortgage, auto loan, student loan, and etc.

4.2.2 Descriptive Knowledge: Factors That Influence Financial Risk

To predict loan default risk, using financial data is the most straightforward way. FICO score and vintage score are widely used among traditional lenders. Online lenders (e.g. Lending Club and Prosper.com) usually combine FICO score with other financial features to create their own lending algorithms (Jagtiani and Lemieux 2017). Some of these features come from credit bureaus, which include (but are not limited to) credit line, payment history, recent inquiries, utilization level, debt to income ratio, etc. (Wang and Overby 2018a). The other features are generated from various types of financial activities, such as home ownership, bankruptcy record, employment history, etc. (Wang and Overby 2018a). Existing studies show that online lenders are able to offer lower rate loans than do traditional lenders, partially because of that online lenders are taking advantage of a richer range of financial data (Jagtiani and Lemieux 2017). Because many of these financial features have a clear logical connection with loan default risk, putting them into predictive models (either simple or complicated one) can no doubt provide satisfactory predictive power.

Although financial data has been effectively used to predict loan default risk, the default rate of online loans is still a concern to regulators and practitioners. Another concern arises when customers with thin credit data come to these online lenders but they cannot make a decision on these customers due to insufficient financial data. As a result, online lenders keep looking for new data sources to further improve or complement their existing prediction algorithms. Some non-financial individual features, such as race, gender, appearance, political ideology have been believed to be correlated with loan default risk (Pope and Sydnor 2011, Wang and Overby 2018b). Some non-financial social features, such as number of friends, have been found to be correlated with loan default risk too (Lin et al. 2013). Although whether and to what extent these features are related with default risk are still inconclusive, some firms already start to incorporate non-financial data into their predictive models. Because it is largely unclear in the literature how these firms deal with non-financial data and how the idea works, we document this specific design science knowledge and the corresponding industrial practice based on the work and business of one Chinese firm who provides online lending loan risk predictive models. We aim to describe a design science approach of using non-financial data to predict loan default risk and investigate the performance of several related design principles.

4.3 Kernel Theory: Predictive Analytics

4.3.1 Predictive Analytics Framework

Kernel theory refers to “any descriptive theory that informs artifact construction” and it mainly “arises in disciplines outside of IS” (Gregor and Hevner 2013). We use Predictive Analytics as the kernel theory in this study because predictive analytics describe how to build a predictive model and how to evaluate a predictive model, which fits the design task of this paper. Predictive analytics include “statistical models and other empirical methods that are aimed at creating empirical predictions, as well as methods for assessing the quality of those predictions in practice.” (Shmueli and Koppius 2011 P554). Different from explanatory statistical models, predictive analytics focus on empirical prediction rather than theoretical prediction. Due to the difference in analysis goal, predictive analytics suggest specific ways to select variables of interest, to build model, and to validate model. We use predictive analytics theory to guide our design in several key steps and the final evaluation, which are elaborated as follows:

Raw data collection and data preparation. We focus on observed and measurable variables in this study and emphasize on measurement quality over explanatory quality. To guarantee closeness of the data, we make sure training data and real-time decision data are drawn in a similar fashion. We also pay special attention to missing values and data partitioning. We use data from three sources, which are within-app browsing behavioral data, short message data, and customer social network data. Because behavioral data, message data, and social network data are primary data sources in this study, we apply multiple tools to convert these unstructured data to numerical values.

Feature generation and feature selection. When raw data can be directly engineered into useful features, we tend to involve them directly. When raw data/features are not recorded at the customer level, we create new features based on expert knowledge and descriptive knowledge. Considering all data are non-financial data and most of time they don’t have an obvious connection with default rate, we draw theory from other domains such as graph theory for social network data and natural language processing theory for text data to generate potentially useful features. After creating all features, we use a combination of methods to pin down to a subset of features, including visualization, dimension reduction, etc.

Model selection. We rely on data-driven algorithms to capture complex relationships in the data and use different machine learning algorithms to build a model for each data source. We pay special attention to explore the algorithm that fits the data structure best. For flattened data/features we eventually choose to use ensemble algorithms while for matrix-like data/features we choose to use neural network algorithms. We apply a two-layer structure by creating a first-layer predictive model for each specific data source and a second-layer predictive model to include predictions from individual predictive models. The final ensemble predictive model serves as the design artifact and is used in real business.

Model evaluation. Following predictive analytics theory, we focus on predictive power rather than statistical significance in model evaluation. We evaluate the performance of two design components, which are individual models on each data domain and the final predictive model. Model evaluation and validation are conducted with testing data.

4.3.2 Within-APP Browsing Data Domain

Within-APP browsing data comes from customer smartphone usage behaviors, including app usage activities, device information, locations, app page content, etc. This data domain contains information about how customers use mobile APP to borrow money, what devices they use, and other information that is generated with mobile APP usage. The challenge to use this data is that all actions happen in a sequential order and the volume of data is large.

To solve the aforementioned challenge, we apply a Markov Transition Field (MTF) as the encoding layer to feed the raw data. The first advantage of MTF is it maintains the inter-relationships of actions. Because the order of actions is still missing so Convolutional Neural Network (CNN) is used to process MTF data. The second advantage of MTF is it prevents us generating too many features. There are around 200 states/domains and for each domain there are 5204 features. If we flatten all features then we would get more than 1,000,000 (200*5204) features. Instead of flattening all features, using MTF can represent both state transaction probability and feature transaction probability.

4.3.3 Short Message Data Domain

Short message information is generated during interactions between customers and financial institutions, either the online lender or other deposit institutions. Short message may contain action information such as withdraw money or deposit money and status information such as current credit line or current balance. The design challenge is how to process natural language embedded in short messages, especially when the formats of these messages is updated rapidly.

We add a hierarchical structure to the traditional knowledge base approach to handle pattern variety and rapid updates. Knowledge base is used to extract useful words and generate features from texts and messages. While traditional knowledge base contains word level and length level knowledge pattern, we in addition represent the knowledge base with a hierarchical vectorized storage (i.e. vector level knowledge pattern) and exact features using vector level pattern too. Features extracted from the three levels of knowledge base are feed into a LightGBM model (one type of gradient boosting model).

4.3.4 Social Network Data Domain

Social network data comes from the basic information of mobile devices and various types of information (e.g. location, internet access, social connection logs) which are stored in mobile devices. Based on the information, we can construct a social network of customers who are connected through multiple channels, such as work for the same company, use the same internet access, appear in each other's contact list, etc. Social network contains important information that can predict the loan default risk of each customer for two reasons: first, financial risk exhibits homophily effects--customers who finally default are generally more socially connected (Min et al. 2018); second, the position in a social network partially represents social-economic status of a consumer, which is correlated with loan default risk. According to the aforementioned two reasons, we generate new features from graph analysis of social network data. We eventually feed both basic features and graph-based features into a LightGBM model.

4.3.5 Predictive Analytics Framework Hypotheses

Based on the aforementioned design approach of the predictive analytics framework, we have two explicit hypotheses to test in this study.

H1: Combined with specific first-layer predictive models, new features generated from non-financial data outperform original features of non-financial data.

H2: Predictive models on non-financial data provide comparable prediction power to predictive models on financial data.

4.4 A Predictive Analytics Framework for Financial Risk Prediction

4.4.1 Framework Overview

In this section, we propose and introduce a framework for predictive analytics of loan default risk based on consumer within-app browsing data, short message conversation data, and social network data. Essentially, this framework combines behavioral language analysis and natural language processing within a specific finance scenario. This framework contains four parts from the bottom to the top (as shown in Figure 4.1):

(1) Raw data collection. We collect and clean data with different structures through ID-mapping, knowledge mapping, and other technologies to build a unified data input model. We skip detailed introduction of this step because it is quite common and doesn't bring in a design science contribution of this study.

(2) Feature generation. According to specific data structure and data meaning, we apply multiple data mining technologies to extract original features and generate new features. These features are used to construct a high-dimensional user financial image and feed into first-layer predictive models (classifiers/algorithms/learners). In addition to the previous section, we detail the feature generation process in this section and exhibit sample features in each data domain. Feature generation is the foundation for our contribution in design science theory and consumer finance theory.

(3) Individual predictive models. We first select machine learning algorithms and train individual predictive models within each data domain. Due to page limitation, we will just briefly introduce why we prefer some certain algorithms to the others and mainly introduce the winning algorithms.

(4) Ensemble predictive model. We eventually combine predictions from individual models into a final model by using an ensemble classifier. This ensemble predictive model is the final design artifact that is used in real business. We skip the design process of the ensemble predictive model because it simply follows common practice.

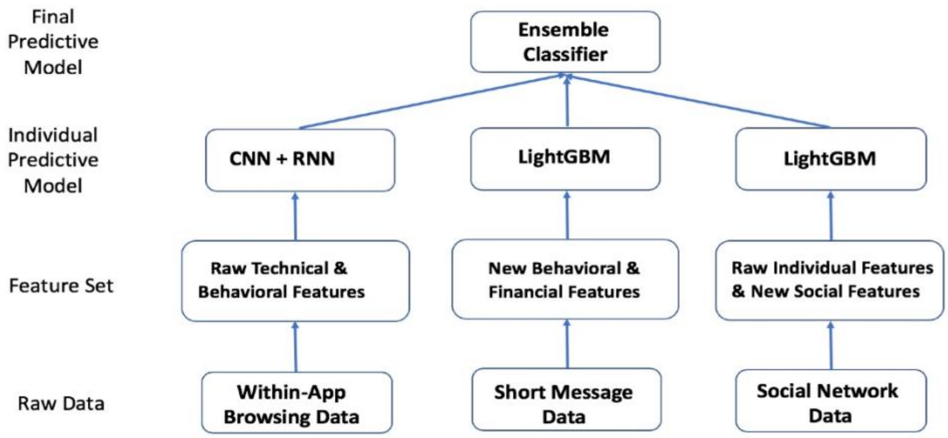


Figure 4.1: The Overview of the Predictive Analytics Model

4.4.2 Feature Generation and Individual Predictive Models

4.4.2.1 Within-APP Browsing Data Domain

As customers interact with websites or smart phone apps, firms can capture a series of actions and status which may implicitly contain customers' habits, motivations, and purposes. However, such behavioral flow data (such as the time customers stay on the page, the number of clicks on the screen, etc.) are mostly weakly-structured and high-dimensional data compared to traditional credit data (e.g. own a house or a car, etc.). How to extract features or patterns from this type of behavior flow data is the major design challenge in this data domain.

Traditional ways tend to flatten all features into the customer level but it is not feasible or efficient in this scenario because even one event type (e.g. register, ID verification, loan application) may have multiple actions. Flattening all features may dramatically increase data dimension and slowdown machine learning models. Several studies propose some other approaches, including a RNN-based approach (Recurrent Neural Network) and a CNN-based approach (Convolutional

Neural Network). The RNN approach follows an encoding layer which puts all events into a sequence and realized by a LSTM (Long-Term and Short-Term Memory Network) model. In addition to the RNN-based approach, extracting the Markov transition probabilities from the event sequence can preserve the information in the inter-relationships among sequential activities. We thus create a MTF (Markov Transition Field) to maintain this information and feed it into a CNN classifier. the This approach complements the main RNN-based approach and contribute to the final individual predictive model. Compared to the RNN-based sequential learning approach, the CNN approach can learn more global information at the MTF. See Zhang et al. (2018) for more details about the design framework. Figure 2 (which directly comes from Zhang et al. 2018) briefly illustrate the design.

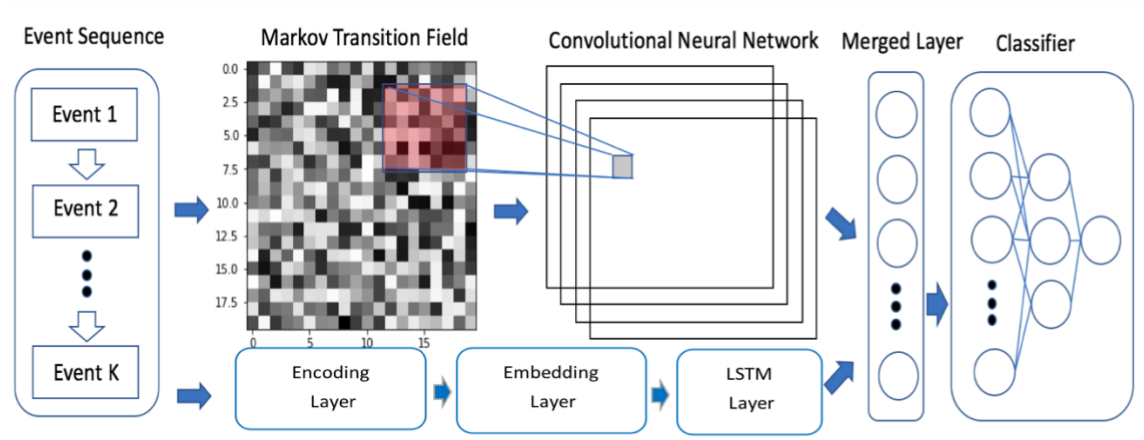


Figure 4.2: Predictive Model on Within-App Browsing Data

4.4.2.2 Short Message Data Domain

As customers interact with financial institutions, short messages are created to communicate financial activities such as money withdraw or money deposit and financial status such as credit line or current balance. Although short message data contain valuable financial information, they are not well-structured financial data in their original forms. In this context, they are stored as text data written by natural language. Therefore, the major challenge is how to extract useful features from massive text files.

Although natural language processing is not a new topic, in the field of financial risk, the value of text data has not been mined for a long time. Following common practice of natural language processing in other areas, an effective solution is to (1) extract templates/dictionaries from a sample of short texts, (2) use these templates to build a knowledge base, (3) use the knowledge base to exact/generate features from all short texts, and (4) use the features to train a predictive model. To focus on important contents, we actually borrow the idea of explicit representation and label key entities only. Similar as traditional approach, we keep word level and length level knowledge base to exact features. To deal with the variety and/or rapid change of text pattern, we in addition represent the knowledge base with an explicit hierarchical vector. The hierarchical design enables fuzzy searching, facilitating different levels of searching keys to give delineation of the short text from various aspects. The vectorized mechanism, serving as a specific layer in the hierarchical structure, extracts the pattern of the key entities from the short text and hence keeps semantical structure into consideration during searching. The hierarchical vectorized design thus enable us to exact more meaningful patterns. We summarize the feature generation approach in Figure 4.3 and exhibit some sample features in Table 1. See more design details in Chen et al. (2018). It is worth noting that although these features are closely related, they don't come from structured financial data and actually call for tremendous design effort to exact and generate.

Table 4.1: Sample Features on Short Message Data

Feature Name	**Feature Description**
Expenditure	The amount of money transferred or used.
Deposit	The amount of money received or deposited.
CreditLine	The credit line of a credit card.
Balance	The balance of a debit card.
MessageCount	The number of messages received from financial institutions.
WithdrawFail	The amount of a withdraw attempt which fails.

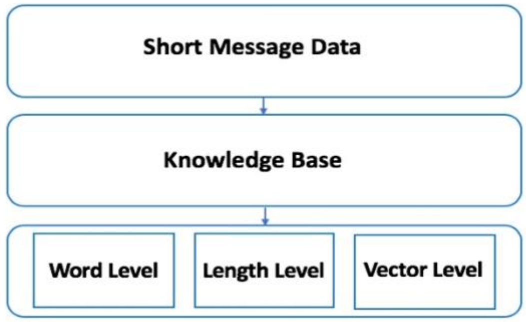


Figure 4.3: Short Message Data Feature

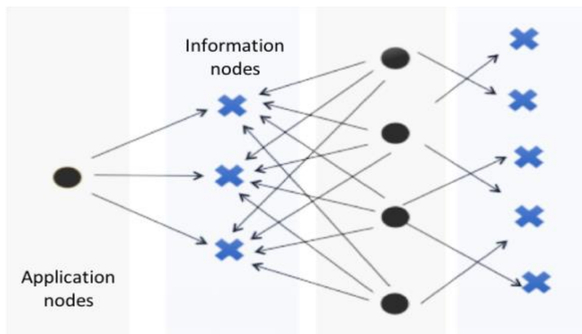


Figure 4.4: A Graph Analysis of Social Network

4.4.2.3 Social Network Data Domain

Social network data come from individual features that can be used to connect customers through multiple relationships. These individual features include customer-related information such as mobile number, home address, company address, emergent contacts and device-related information such as device id, wifi access, Mac address, GPS coordinates and so on. Customers can form a complicated social network through these connections and relationships. Although graph theory is mature enough to analyze a social network, the challenge in this study is that the social network here is built on multiple logics/connections. For example, customer A and B can be connected either because of sharing the same company address or the same wifi access.

To solve this problem, we apply bipartite graph to represent this customer social network. The bipartite graph contains two types of nodes: one is defined as the application node to represent application-related information such as application time, loan decision, loan performance, etc.; the other is defined as the information node to track information related to the applicant, such as email address, telephone number, address, equipment, etc. Nodes of the same type cannot be directly connected, but they are associated by the other type of node. The edge connecting application node to information node indicates a relationship type. For example, one application decision is associated with one applicant company address while the same company address may also be associated with another application decision. In addition, we can also define their own attributes on the nodes and edges to reflect the incidental information. For example, for the application node, the definable attributes include load decision (such as pass or reject), approval quota, and post-issuance performance (such as overdue or normal), fraud and non-fraud application time, etc. We can also define the time attribute of an edge to describe the effective time of the relationship and the weight to describe the strength of the relationship. Please see Min et al. (2018) for more details and Figure 4.4 (which is adopted from Min et al. 2018) for a brief summary.

Based on the aforementioned graph, we create three types of features, i.e. local features, global features, and mismatch features. Local features measure the statistical characteristics of n-order neighborhood around the application node. Given the ego network, we use three graph metrics to evaluate the local network structure, including degree, quadrangle, and density. Global features manage to take into account historical labeled fraudulent application nodes and use this knowledge to infer a primary default probability for the unlabeled application nodes. Personalized page rank algorithm is used to spread default from the labeled default application nodes to information nodes, and then to unlabeled application nodes again. This process is proportionally to the relationship strength while simultaneously assigned decaying weights on past defaults. After obtaining the default probability of each node through the graph mining algorithm, the following characteristics are calculated: the probability of default of the current application, the maximum value and average value of the probability of the neighbor node. This feature generation process makes theoretical sense because default exhibits homophily effect, which suggests that default customers are often more socially connected. Mismatch features are defined by anecdotal evidence and human experts. In risk management, finding leads for mismatch is an effective way to detect fraud and default. We consider two aspects of mismatch. One type is caused by inconsistent information collected from different channels. Jaccard distance is used to mathematically quantify the similarity of a given type of information from different data sources (similarity of two sub-graph). The other type is caused by conflicting information collected in the rest of network. We illustrate the three types of features in Table 4.2.

Table 4.2 Sample Features on Social Network Data

Feature Name	Feature Type	Feature Description
ClusteringCoefficient	local	How the applicant is clustered with his/her neighbors.
Quadrangles	local	How many quadrangles the applicant could form within two-degree neighborhood.
PageRankScore	global	How topologically important the applicant is in the network.
DefaultScore	global	How the applicant is affected by the default customers via semi-supervised propagation through the network.
PhoneMismatchLevel	mismatch	How much the phone number provided by the applicant is different from the one stored in financial institutions.
AddressMismatchLevel	mismatch	How much the address provided by the applicant is different from the one stored in financial institutions.

4.4.3. Ensemble Predictive Model

We applied various types of predictive models for each data domain, including boosting tree, deep learning classifier network, etc. At the aggregated level, we use a logistic regression model to integrate the results from individual predictive models to predict loan default risk. Logistic regression provides business insights in the importance of each data domain and thus helps explain the final lending decision.

4.5 Evaluation

4.5.1 Extracted Features versus Original Features

In this section we test Hypothesis 1, which is whether the new features generated by our proposed approach outperformance original features in each data domain. The experiment works in this way: we first train our predictive model with different sets of features and then check AUC and KS scores of each model with the testing data. As can be seen in Table 4.3, our proposed feature generation approach plus its specific machine learning algorithm always outperforms original features plus the same machine learning algorithm. This experiment confirms the predictive value of non-financial data and supports our design choice. Detailed experiment design is omitted due to space limitation but available upon request.

Table 4.3: Experiment Results on Individual Predictive Model

Individual Predictive Model	**Model Description**	**Test AUC**	**Test KS**
Data Domain: Within-App Browsing Data			
Baseline KNN + DTW	K Nearest Neighbor classifier based on features generated by Dynamic Time Warping	0.552	0.142
Baseline MLP	Multilayer Perceptron based on flattened features	0.560	0.167
Proposed RNN	Recurrent Neural Network classifier based on features encoded with sequential layers	0.602	0.203
Proposed RNN + CNN	Recurrent Neural Network classifier based on features encoded with sequential layers plus Convolutional Neural Network classifier based on features represented by Markov Transition Field	0.621	0.216
Data Domain: Short Message Data			
Baseline Regular Expression	LightGBM classifier based on features generated by Regular Expression of short messages	0.692	0.288
Proposed Hierarchical Vectorized Representation	LightGBM classifier based on features represented by hierarchical vectors	0.693	0.290
Data Domain: Social Network Data			
Baseline Individual Feature	LightGBM classifier based on individual features only	0.710	0.300
Proposed Graph Analysis	LightGBM classifier based on both individual features and graph-based three types of features	0.750	0.380

4.5.2 Non-financial Data-based Predictive Models

In this section we test Hypothesis 2, which is whether the performance of predictive models using non-financial data can be comparable to the performance of predictive models using financial data. Since we have both individual predictive model and the final ensemble model, we first test the performance of individual model respectively and then test the overall performance. For individual predictive models, we use the best configuration of feature generation and machine learning algorithm as shows in Table 4.3. The experiment covers the period from 2017/10/10 to 2017/10/31, including 8887 default customers and 38432 good customers. We train each model with 75% of the total population and measure model performance in testing data with AUC (area under the ROC curve) score, KS (Kolmogorov Smirnov) score and overall accuracy ((true positive + true negative) / total observations). Because for the same customer population we don't have strong financial data to build baseline predictive model (model 1), we use the value range of AUC, KS, and accuracy from existing studies that are investing loan default rate predictive models in similar context.

Table 4.4 Experiment Results on Ensemble Predictive Model

Model No.	Model Description	Test AUC	Test KS	Test Accuracy
1	Multiple predictive models on strong financial data	0.75-0.82	NA.	0.68-0.86
2	Proposed RNN+CNN predictive model on within-app browsing data	0.62	0.18	0.63
3	Proposed Hierarchical Vector model on short message data	0.69	0.28	0.68
4	Proposed Graph Analysis model on social network data	0.57	0.11	0.63
5	Combination of Models 2-4	0.71	0.31	0.69

Note: model 1 results come from Jin and Zhu (2015) and Jiang et al. (2018)

The performance of individual models is ranging from 0.57 to 0.69 in terms of AUC score and from 0.11 to 0.28 in terms of KS score. The ensemble model achieves the highest performance among all models, indicating that three individual predictive models don't perfectly overlap with each other. In other words, each individual model and the underlying data domain can provide unique contribution the final predictive power. To better compare the performance of models 2-5, we show the ROC (Receiver Operating Characteristics) curve in Figure 4.5 and Precision-Recall curve in Figure 4.6. Among three data domains, the predictive model on short message data has the best performance. However, combining all three data domains can further improve predictive power, which validates the power of "big data". The overall performance of our predictive analytics framework (model 5) is very close to predictive models based on strong financial data (model 1), which confirms the promise of using non-financial data to predict loan default risk. Although we cannot test the predictive power of models on both financial data and non-financial data, it is very likely non-financial data can complement financial data in predicting loan default.

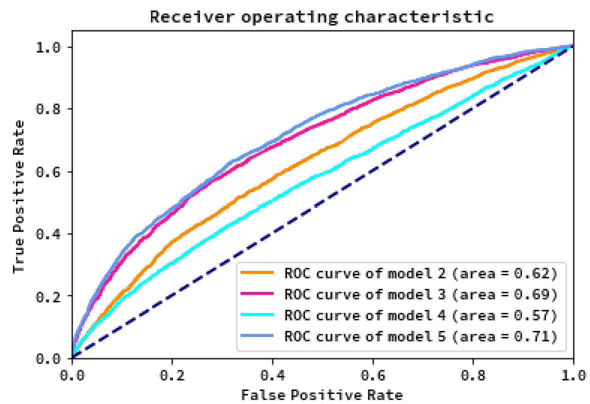


Figure 4.5 ROC Curve

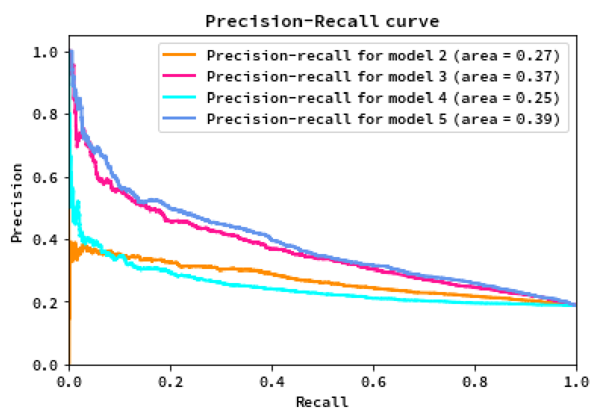


Figure 4.6 Precision-Recall Curve

4.6 Conclusions

We propose a predictive analytics framework to predict loan default risk using non-financial data. We build individual predictive model on three data domains, i.e. within-app browsing data, short message data, and social network data, and combine individual prediction results into an ensemble predictive model. A series of experiments validate the feature generation approach we proposed to use on non-financial data and the predictive power of the final model.

This predictive analytics framework contributes to consumer finance literature by exploring potential connections between non-financial features with loan default risk, indicating directions for future causality analysis. The framework also contributes to design science theory by illustrating how to generate useful predictive features from unconventional data and how to find specific predictive model for these features. The proposed framework can not only increase loan default risk prediction accuracy but also be applied to areas of credit card application, identity fraud control, as well as financial product marketing, and other financial services.

CHAPTER 5

CONCLUSION

In the beginning of this dissertation, I motivated the research by a speculation that understanding user's status and needs by understanding user's behavior. Specifically, this study presented serval different investigations to derive new insights into the empirical study of mobile user behaviors. This concluding chapter highlights the main contributions of the work. In particular, this thesis sought out to address the following questions:

- Is user's culture an antecedent of user's behavior?
- Does user's culture influence deviation behavior?
- Does user's culture influence emotional expression (suppression)?
- Do financial incentives induce strategic behavior?
- Assume the existence of strategic behavior, what is the posterior behavior whem financial incentive removed?
- Can online social network features mitigate strategic behavior?
- Do predictive models on non-financial data provide comparable prediction power to predictive models on financial data.

The novel insights and discoveries presented are:

(1) Research perspective innovation. Traditional mobile health research studies the effects of mobile applications from the perspective of a single task, while ignoring the entire behavior process. This thesis attempts to re-examine mobile applications from the perspective of user behavior, combining user generated content, in app activities, social network etc.

The second study in this Thesis conducts research on financial incentives, summarizes the relationship between user behavior characteristics and financial incentives, explores the mechanism of financial incentives, and illustrate the design of user incentives and the influencing factors of weakening strategy behaviors. This study has an interdisciplinary thinking on user activity, health performance, social interaction and other issues in health applications; and it breaks through the limitations of the single-disciplinary perspective to broaden the scope of research and enrich the research theme.

(2) Research method innovation. Traditional research has used multivariate statistical analysis methods to explain the relationship between the psychological and behavioral influencing factors of users and the variables in the theoretical model. These methods have great limitations in the research of user behavior data in mobile applications, mainly because of the amount and complexity of the data (usually it contain large amounts of unstructured data). Traditional methods are difficult to cope with such high dimensional data. Thus, a more rapid and robust feature extraction design is introduced by the third study in this research--the "Behavioral Language Framework".

This study initially find that features based on theory have better predictive power than original features. Then, this study also identify useful features from unconventional data. Our framework show that specific methods should be used to fit different data sources and feature structures. The result show that such intelligent and disruptive research frameworks are able to obtain more accurate predictions results than traditional research methods.

REFERENCES

Chapter 2

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. Information Systems Research, 24(4), 956-975.

Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. International Journal of Electronic Commerce, 20(1), 76-111.

Asch, S. E. (1955). Opinions and social pressure. Scientific American, 193, 35-35.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. MIS Quarterly, 26(3), 243-268.

Bond, R., & Smith, P. B. (1996). Culture and conformity: A Meta-analysis of studies using Asch's (1952b, 1956) line judgment task. Psychological Bulletin, 119(1), 111-137.

Brehm, S., & Brehm, J. W. (1981). Psychological reactance: A theory of freedom and control. New York, NY: Academic Press.

Burtch, G., Ghose, A., & Wattal, S. (2014). Cultural differences and geography as determinants of online pro-social lending. MIS Quarterly, 38(3), 773-794.

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (Forthcoming). Stimulating online reviews by combining financial incentives and social norms. Management Science.

Butler, E. A., Lee, T. L., & Gross, J. J. (2007). Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific? Emotion, 7, 30-48.

Cao, Q., Duan, W. J., & Gan, Q. W. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. Decision Support Systems, 50(2),511-521.

Chen, Z., & Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. Journal of Marketing Research, 50(4), 463-476.

Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth online: Online book reviews. Journal of Marketing Research, 43(3), 345-354.

Chung, C. M., & Darke, P. R. (2006). The consumer as advocate: Self-relevance, culture, and word-of- mouth. Marketing Letters, 17(4), 269-279.

Cialdini, R., Wosinska, W., Barrett, D. W., Butner, J., & Gornik-Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. Personality and Social Psychology Bulletin, 25(10), 1242-1253.

Dai, W., Jin, G., Lee, J., & Luca, M. (2012). Optimal review aggregation of consumer ratings: An application to Yelp.com (NBER working paper).

Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009). How opinions are received by online communities: A case study on Amazon.com helpfulness votes. In Proceedings of the 18th international Conference on World Wide Web (pp. 141-150). ACM.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management Science, 49(10), 1407-1424.

Dimoka, A., Hong, Y., & Pavlou, P. A. (2012). On product uncertainty in online markets: Theory and evidence. MIS Quarterly, 36(2), 395-426

Fang, H., Zhang, J., Bao, Y., & Zhu, Q. (2013). Towards effective online review systems in the chinese context: A cross-cultural empirical study. Electronic Commerce Research and Applications, 12(3), 08- 220.

Friedlmeier, W., Corapci, F., & Cole, P. M. (2011). Socialization of emotions in cross-cultural perspective. Social and Personality Psychology Compass, 5(7), 410-427.

Giannetti, M., & Yafeh, Y. (2012). Do cultural differences between contracting parties matter? Evidence from syndicated bank loans. Management Science, 58(2), 365-383.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Transactions on Knowledge and Data Engineering, 23(10), 1498-1512.

Godes, D., & Silva, J. C. (2012). Sequential and temporal dynamics of online opinion. Marketing Science, 31(3), 448-473.

Goes, P. B., Lin, M., & Yeung, C. M. A. (2014). “Popularity effect” in user-generated content: Evidence from online reviews. Information Systems Research, 25(2), 222-238.

Goleman, D. (1990). The group and the self: New focus on a cultural rift. The New York Times. Retrieved from http://www.nytimes.com/1990/12/25/science/the-group-and-the-self-new-focus-on-a-cultural- rift.html

Hofstede, G. H. (2001). Culture's consequences: International differences in work-related values. Thousand Oaks, CA: Sage.

Holtgraves, T. (1997). Styles of language use: Individual and cultural variability in conversational indirectness. Journal of Personality and Social Psychology, 73(3), 624-637.

Hong, Y., Chen, P., & Hitt, L. (2013). Measuring product type with dynamics of online review variance. In Proceedings of International Conference on Information Systems.

Hong, Y., & Pavlou, P. A. (2014). Product fit uncertainty in online markets: Nature, effects, and antecedents. Information Systems Research, 25(2), 328-344.

Hong, Y., & Pavlou, P. A. (2014). Is the world truly “flat”? Empirical evidence from online labor markets (Research paper no. 15-045). Fox School of Business.

House R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). Culture, leadership, and organizations: The GLOBE study of 62 societies. Thousand Oaks, CA: Sage.
Huang, H. (2005). A cross-cultural test of the spiral of silence. International Journal of Public Opinion Research, 17(3), 324-345.

Huang, N., Burtch, G., Hong, Y., & Polman, E. (2016). Effects of multiple psychological distances on construal and consumer evaluation: A field study of online reviews. Journal of Consumer Psychology, 26(4), 474-482.
Huang, N., Hong, Y., & Burtch, G. (2015). Anonymity and language usage: A natural experiment of social network integration. In Proceedings of the 35th International Conference on Information Systems.

Liu, Y., Chen, P. Y., & Hong, Y. (2014). Value of multi-dimensional rating systems: An information transfer view. In Proceedings of the 35th International Conference on Information Systems.
Inglehart, R., & Welzel, C. (2010). Changing mass priorities: The link between modernization and democracy. Perspectives on Politics, 8(2), 551–567.

Inglehart, R., & Oyserman, D. (2004). Individualism, autonomy, self-expression. The human development syndrome. In International studies in sociology and social anthropology (pp. 74-96). Institute for Research on Intercultural Cooperation.

King, R. A., Racherla, P., & Bush, V. D. (2014). What we know and don't know about online word-of-mouth: A review and synthesis of the literature. Journal of Interactive Marketing, 28(3), 167-183.

Koh, N. S., Hu, N., & Clemons, E. K. (2010). Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. Electronic Commerce Research and Applications, 9(5), 374-385.

Kwark, Y., Chen, J., & Raghunathan, S. (2014). Online product reviews: Implications for retailers and competing manufacturers. Information Systems Research, 25(1), 93-110.

Lam, D., Lee, A., & Mizerski, R. (2009). The effects of cultural values in word-of-mouth communication. Journal of International Marketing, 17(3), 55-70.

Lee, D., Hosanagar, K., & Nair, H. (2013). The effect of advertising content on consumer engagement: Evidence from Facebook.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. Journal of Marketing Research, 48(5), 881-894.

Lee, Y. J., Hosanagar, K., & Tan, Y. (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. Management Science, 61(9), 2241-2258.

Leidner, D. E., & Kayworth, T. (2006). Review: A review of culture in information systems research: Toward a theory of information technology culture Conflict. MIS Quarterly, 30(2), 357-399.

Li, X., & Hitt, L. (2008). Self-selection and information role of online reviews. Information Systems Research, 19(4), 456-474.

Liu, Y., Chen, P., & Hong, Y. (2014). Value of multi-dimensional rating systems: An information transfer view. In Proceedings of the 35th International Conference on Information Systems.

Lu, X., Ba, S., Huang, L., & Feng, Y. (2013). Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. Information Systems Research, 24(3), 596-612.

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. Journal of Marketing, 77(1), 87-103.

Moe, W., & Schweidel, D. (2012). Online product opinions: Incidence, evaluation, and eevolution. Marketing Science, 31(3), 372-386.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. Science, 341(6146), 647-651.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quarterly, 34(1), 185-200.

Nagle, F., & Riedl, C. (2014). Online word of mouth and product quality disagreement. In Proceedings of the Academy of Management Annual Meeting.

Niedenthal, P., Krauth-Gruber, S., & Ric, F. (2006). Psychology of emotion interpersonal, experimental, and cognitive approaches. New York, NY: Psychology Press.

Pavlou, P., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. Information Systems Research, 17(4), 392-414.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001.Mahway: Lawrence Erlbaum Associates.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. Academy of Management Review, 32(2), 344-354.

Spiller, S. A., Fitzsimons, G. J., Lynch Jr, J. G., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. Journal of Marketing Research, 50(2), 277-288.

Sridhar, S., & Srinivasan, R. (2012). Social influence effects in online product ratings. Journal of Marketing, 76(5), 70-88.

Sun, M. (2012). How does the variance of product ratings matter? Management Science, 58(4), 696-707.

Takahashi, K., Ohara, N., Antonucci, T., & Akiyama, H. (2002). Commonalities and differences in close relationships among the Americans and Japanese: A comparison by the individualism/collectivism concept."International Journal of Behavioral Development, 26(5), 453-465.

Triandis, H. C. (1995). Individualism and collectivism. Boulder, CO: Westview Press.

Tsai J. L., Miao F. F., Seppala E., Fung H. H., & Yeung D. Y. (2007). Influence and adjustment goals: Sources of cultural differences in ideal affect. Journal of Personality and Social Psychology, 92(6), 1102-1117

Wang, C., Zhang, X., & Hann, I. H. (Forthcoming). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. Information Systems Research.

Wu, F., & Huberman, B. (2008). How public opinion forms. In Proceedings of the 4th International Workshop on Internet and Network Economics (pp. 334-341) Berlin: Springer.

Yaveroglu, I. S., & Donthu, N. (2002). Cultural influences on the diffusion of new products. Journal of International Consumer Marketing, 14(4), 49-63.

Yin, D., Bond, S., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. MIS Quarterly, 38(2), 539-560.

Chapter 3

Acland, D., & Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. Management Science, 61(1), 146-160.

Adar, E., Teevan, J., & Dumais, S. T. (2008). Large scale analysis of web revisitation patterns. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems(pp. 1197-1206). ACM.

Agarwal R, Gao G, DesRoches C, Jha AK (2010) The digital transformation of healthcare: Current status and the road ahead. Inform. Systems Res. 21(4):796–809.

Angrist, J., & Lavy, V. (2009). The effects of high-stakes high school achievement awards: Evidence from a randomized trial. The American Economic Review, 1384-1414.
Artyom Dogtiev 2018 "App Download and Usage Statistics" http://www.businessofapps.com/data/app-statistics/

Ashraf, N., Karlan, D., Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. The Quarterly Journal of Economics, 635-672.

Babar, Y., Chan, J., & Choi, B. (2018). Run Forrest Run!: Measuring the Impact of App-Enabled Social and Performance Feedback on Running Performance.

Barrera, M., Glasgow, R. E., McKay, H. G., Boles, S. M., & Feil, E. G. (2002). Do Internet‐Based Support Interventions Change Perceptions of Social Support? An Experimental Trial of Approaches for Supporting Diabetes Self‐Management. American journal of community psychology, 30(5), 637-654.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., & Perez-Calle, F. (2008). Conditional cash transfers in education design features, peer and sibling effects evidence from a randomized experiment in Colombia. NBER Working Paper.

Barrow, L., Richburg-Hayes, L., Rouse, C. E., & Brock, T. (2014). Paying for performance: The education impacts of a community college scholarship program for low-income adults. Journal of Labor Economics, 32(3), 563-599.

Behrman, J. R., Sengupta, P., Todd, P. (2005). Progressing through PROGRESA: An impact assessment of a school subsidy experiment in rural Mexico. Economic development and cultural change, 54(1), 237-275.

F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. 2009. Characterizing User Behavior in Online Social Networks. In Proc. of IMC.

Beshears, J. L., Choi, J. J., Laibson, D., Madrian, B. C., Sakong, J. (2015). Self control and commitment: Can decreasing the liquidity of a savings account increase deposits? NBER Working Paper.

Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. Review of Economics and Statistics, 94(3), 686-698.

Blackburn, G. (1995). Effect of degree of weight loss on health benefits. Obesity Research, 3(2), 211-216.

Bonner, S. E., Hastie, R., Sprinkle, G. B., Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. Journal of Management Accounting Research, 12(1), 19-64.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., ... & Shekelle, P. G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Annals of internal medicine, 144(10), 742-752.

Cavallo, D. N., Tate, D. F., Ries, A. V., Brown, J. D., DeVellis, R. F., Ammerman, A. S. (2012). A social media–based physical activity intervention: a randomized controlled trial. American journal of preventive medicine, 43(5), 527-532.

Charness, G., Gneezy, U. (2009). Incentives to exercise. Econometrica, 77(3), 909-931.
G. Chittaranjan, J. Blom, and D. Gatica-Perez, “Mining large-scale smartphone data for personality studies,” Pers. Ubiquitous Comput., vol. 17, pp. 433–450, 2012.

Cohen-Cole, E., Fletcher, J. M. (2008). Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. Journal of health economics, 27(5), 1382-1387.

Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium puzzle. Journal of Political Economy, 519-543.

Courty P, Marschke G (2004) An empirical investigation of gaming responses to explicit performance incentives. Journal of Labor Economics 22(1): 23-56.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology, 18(1), 105.

T. M. T. Do, J. Blom, and D. Gatica-Perez, “Smartphone usage in the wild: A large-scale analysis of applications and context,” in Proc. 13th Int. Conf. Multimodal Interfaces, 2011, pp. 353–360.

Donatelle, R. J., Hudson, D., Dobie, S., Goodall, A., Hunsberger, M., & Oswald, K. (2004). Incentives in smoking cessation: status of the field and implications for research and practice with pregnant smokers. Nicotine & Tobacco Research, 6(2), 163-179.

Eysenbach, G. (2011). CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. Journal of medical Internet research, 13(4), 126.

H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, “Diversity in smartphone usage,” in Proc. 8th Int. Conf. Mobile Syst., Appl.Services, 2010, pp. 179–194.

Fox, S., & Duggan, M. (2010). Mobile health 2010. Washington, DC: Pew Internet & American Life Project.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. Journal of economic surveys, 15(5), 589-611.

R. S. Geiger and A. Halfaker. 2013. Using Edit Sessions to Measure Participation in Wikipedia. In Proc. Of CSCW.

Ghose, A., Goldfarb A., and Han, S.P. 2013. “How is the Mobile Internet Different? Search Costs and Local Activities,” Information Systems Research (24:3), pp. 613-631

Ghose, A., and Han, S.P. 2011. “An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet,” Management Science (57:9), pp. 1671-1691.

Gine, X., Karlan, D., Zinman, J. (2008). Put your money where your butt is: a commitment savings account for smoking cessation. Working paper, Yale University.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. The Journal of Economic Perspectives, 25(4), 191-209.

S. Gündüz and M. T. Özsu. 2003. A Web page prediction model based on click-stream tree representation of user behavior. In Proc. of SIGKDD.

N. Henze, M. Pielot, B. Poppinga, T. Schinke, and S. Boll, “My app is an experiment: Experience from user studies in mobile app stores,” Int. J. Mobile Human Comput. Interaction, vol. 3, pp. 71–91, 2011.

Jackson, C. K. (2010). A little now for a lot later a look at a Texas advanced placement incentive program. Journal of Human Resources, 45(3), 591-639.

Jeffery, R. W., Hellerstedt, W. L., & Schmid, T. L. (1990). Correspondence programs for smoking cessation and weight control: a comparison of two strategies in the Minnesota Heart Health Program. Health Psychology, 9(5), 585.

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., & Volpp, K. G. (2011). Financial incentives for extended weight loss: a randomized, controlled trial. Journal of general internal medicine, 26(6), 621-626.

Kapeller, J. (2010). Citation metrics: Serious drawbacks, perverse incentives, and strategic options for heterodox economics. American Journal of Economics and Sociology, 69(5), 1376-1408.

Kwon, H. E., So, H., Han, S. P., & Oh, W. (2016). Excessive dependence on mobile social apps: A rational addiction perspective. Information Systems Research, 27(4), 919-939.

Lin P-H, Wang Y, Levine E, Askew S, Lin S, Chang C, Sun J, Foley P, Wang H, Li X, Bennet GG (2014) A text message-assistant randomized lifestyle weight loss clinical trial among overweight adults in Beijing. Obesity 22(5): 29-37.

L. Lu, M. Dunham, and Y. Meng. 2005. Mining significant usage patterns from clickstream data. In Proc. of WebKDD .

Mayer, C., Morrison, E., Piskorski, T., & Gupta, A. (2014). Mortgage modification and strategic behavior: evidence from a legal settlement with Countrywide. The American Economic Review, 104(9), 2830-2857.

Meier, S. (2007). Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. Journal of the European Economic Association, 5(6), 1203-1222.

Milkman, K. L., Minson, J. A., Volpp, K. G. (2013). Holding the Hunger Games hostage at the gym: An evaluation of temptation bundling. Management Science, 60(2), 283-299.

Obendorf, H. Weinreich, H. Herder, E. and Mayer, M.. 2007. Web page revisitation revisited: implications of a long-term click-stream study of browser usage. In Proc. of CHI

Ogden, C. L., Carroll, M. D., Kit, B. K., Flegal, K. M. (2014). Prevalence of childhood and adult obesity in the United States, 2011-2012. Journal of the American Medical Association, 311(8), 806-814.

H. Oktay, B.J. Taylor, and D.D. Jensen. Causal discovery in social media using quasi-experimental designs. In Proceedings of the First Workshop on Social Media Analytics, pages 1–9. ACM, 2010.

Oyer P. (1998) Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. Quarterly Journal of Economics 113: 149-185.

Pantelopoulos A, Bourbakis N (2010) A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Transactions on Systems, Man, and Cybernetics—Part C, 40(1): 1-12.

Park A. (2015) The New Science of How to Quit Smoking. TIME Health.

J. Y. Park, N. O'Hare, R. Schifanella, A. Jaimes, and C. Chung. (2015) A Large-Scale Study of User Image Search Behavior on the Web. In Proc. of CHI .

Pasanisi, F., Contaldo, F., De Simone, G., & Mancini, M. (2001). Benefits of sustained moderate weight loss in obesity. Nutrition, metabolism, and cardiovascular diseases: NMCD, 11(6), 401-406.

Pantelopoulos, A., & Bourbakis, N. G. (2010). A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(1), 1-12.

Patel D., Lambert E.V., da Silva R., Greyling M., Nossel C., Noach A., Derman W., Gaziano T. 2010. "The Association between Medical Costs and Participation in the Vitality Health Promotion Program among 948,974 Members of a South African Health Insurance Company." American Journal of Health Promotion 24: 199–204

A. Rahmati, C. Tossell, C. Shepard, P. Kortum, and L. Zhong, "Exploring iPhone usage: The influence of socioeconomic differences on smartphone adoption, usage and usability," in Proc. ACM Int. Conf. Human Comput. Interaction Mobile Devices Services, 2012, pp. 11–20.

Ritterband, L. M., Andersson, G., Christensen, H. M., Carlbring, P., & Cuijpers, P. (2006). Directions for the international society for research on internet interventions (ISRII). Journal of Medical Internet Research, 8(3), e23.

J. M. Rzeszotarski and A. Kittur. 2011. Instrumenting the Crowd: Using Implicit Behavioral Measures to Predict Task Performance. In Proc. of UIST .
N. Sadagopan and J. Li. 2008. Characterizing Typical and Atypical User Sessions in Clickstreams. In Proc. Of WWW.

Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa Poverty Program. Journal of Development Economics, 74(1), 199-250

W.R. Shadish, T.D. Cook, and D.T. Campbell. Experimental and quasi-experimental designs for generalized causal inference. Houghton, Mifflin and Company, 2002

Siopis G, Chey T, Allman-Farinelli M (2015) A systematic review and meta-analysis of interventions for weight management using text messaging. Journal of Human Nutrition and Dietetics 28(2): 1-15.

Statista (2017) https://www.statista.com/statistics/387867/value-of-worldwide-digital-health-market-forecast-by-segment/

Svetkey, L.P., Batch, B.C., Lin, P.H., Intille, S.S., Corsino, L., Tyson, C.C., Bosworth, H.B., Grambow, S.C., Voils, C., Loria, C., Gallis, J.A., Schwager, J., Bennet, G.B. (2015). Cell phone intervention for you (CITY): A randomized, controlled trial of behavioral weight loss intervention for young adults using mobile technology. Obesity, 23(11), 2133-2141.

Thaler, R. H., Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. Journal of Political Economy, 112(1), 164-187.

Ting, C. Kimble, and D. Kudenko. 2005. UBB Mining: Finding Unexpected Browsing Behaviour in Clickstream Data to Improve a Web Site's Design. In Proc. of ICWI.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive-based approaches for weight loss: a randomized trial. Journal of the American Medical Association, 300(22), 2631-2637.

Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., Galvin, R., Zhu, J., Wan, F., DeGuzman, J. and Corbett, E. (2009). A randomized, controlled trial of financial incentives for smoking cessation. New England Journal of Medicine, 360(7): 699-709.

G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao. 2013b. You Are How You Click: Clickstream Analysis for Sybil Detection. In Proc. Of USENIX Security .

Xu K, Chan J, Ghose A, Han SP (2016) Battle of the channels: The impact of tablets on digital commerce. Management Science, forthcoming.

Yan L, Tan Y (2014) Feeling blue? Go online: An empirical study of social support among patients. Information Systems Research, 25(4): 690-709.

Yan L, Peng J, Tan Y (2015) Network dynamics: how can we find patients like us? Information Systems Research, 26(3): 496-512

Chapter 4

Chen, J.L., Zhou, M.X., Li, R.X., and Wei, M. 2018. "A Hierarchical Vectorized Representation of Knowledge Base," in Proceedings of KDD Fintech Workshop, New York, NY.

Goes, P.B. 2014. "Design Science Research in Top Information Systems Journals," MIS Quarterly (38:1), pp. iii-viii.

Gregor, S., and Hevner, A.R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," MIS Quarterly (37:2), pp. 337-355.

Jagtiani, J., and Lemieux, C. 2017. "Fintech Lending: Financial Inclusion, Risk Pricing, and Alternative Information," SSRN working paper available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3005260.

Jiang, C.Q., Wang, Z., and Wang, R.Y. 2018. "Loan Default Prediction by Combining Soft Information Extracted from Descriptive Text in Online Peer-to-peer Lending," Annuals of Operations Research (266;1-2), pp. 511-529.

Jin, Y., and Zhu, Y.D. 2015. "A Data-driven Approach to Predict Default Risk of Loan for Online Peer-to-peer Lending," in International Conference on Communication Systems and Network Technologies.

Lessmann, S., Baesens, B., Seow, H. V., and Thomas, L. C. 2015. "Benchmarking State-of-the-art Classification Algorithms for Credit scoring: An Update of Research," European Journal of Operational Research 247(1), pp. 124–136.

Lin, M.F., Prabhala, N.R., and Viswanathan, S. 2013. "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending," Management Science (59:1), pp. 17-35.

Min, W., Tang, Z.Y., Zhu, M., Dai, Y.X., and Wei, Y. 2018. "Behavior Language Processing with Graph Based Feature Generation for Fraud Detection in Online Lending," in Proceedings of Workshop on Misinformation and Misbehavior Mining on the Web, Marina Del Rey, CA.

Pope, D.G., and Sydnor, J.R. 2011. "What's in A Picture? Evidence of Discrimination from Prosper.com," Journal of Human Resources (46:1), pp. 53-92.

Shmueli, G., and Koppius, O.R. 2011. "Predictive Analytics in Information Systems Research," MIS Quarterly (35:3), pp. 553-572.

Tufano, P. 2009. "Consumer Finance," Annual Review of Financial Economics, pp. 227-247.

Wang, H.C., and Overby, E. 2018. "How Does Online Lending Influence Bankruptcy Filings?" SSRN Working Paper available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958916.

Wang, H.C., and Overby, E. 2018. "Do Political Differences Decrease Market Efficiency? An Investigation in the Context of Online Lending," in Proceedings of the 39th International Conference on Information San Francisco, CA.

Zhang, R.N., Zheng, F.L., and Min, W. 2018. "Sequential Behavioral Data Processing Using Deep Learning and the Markov Transition Field in Online Fraud Detection," KDD 2018 Data Science in Fintech.

Zhang, Y., Jia, H., Diao, Y., Hai, M., and Li, H. 2016. "Research on Credit Scoring by Fusing Social Media Information in Online Peer-to-peer Lending," Procedia Computer Science (91), pp. 168-174.