Distribution Agreement

Essays on Online Word of Mouth

By

Hulya Karaman
Doctor of Philosophy

Business

Jagdish Sheth
Advisor

Anand Swaminathan
Advisor

Douglas Bowman
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Essays on Online Word of Mouth

By

Hulya Karaman
B.Sc., Istanbul Technical University, 2004
M.Sc., University of Florida, 2006
M.Sc., Washington University in St. Louis, 2012

Advisor: Jagdish Sheth
Advisor: Anand Swaminathan

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Business
2018

Abstract

Essays on Online Word of Mouth
By Hulya Karaman

The significance of online word of mouth (WOM) in customer decision making is increasing. Thus, it is becoming more and more important for marketers to understand online WOM. In my dissertation work, I study two distinct aspects of online WOM. More specifically, the first essay of my dissertation focuses on identifying the extent to which social influence impacts online reviews; the second essay addresses how online reviews of competitors shape consumer decision making. Additionally, I discuss the importance of how managers should respond when confronted with negative (positive) online reviews posted by highly dissatisfied (satisfied) reviewers in order to retain them as customers. Overall, my dissertation work provides managerial insights into how online reviews evolve over time, how they impact customer decision making, and how managers should respond to them.

Essays on Online Word of Mouth

By

Hulya Karaman
B.Sc., Istanbul Technical University, 2004
M.Sc., University of Florida, 2006
M.Sc., Washington University in St. Louis, 2012

Advisor: Jagdish Sheth
Advisor: Anand Swaminathan

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Business
2018

Acknowledgments

I am especially indebted to Dr. Jagdish Sheth and Dr. Anand Swaminathan, who have provided me extensive personal and professional guidance over the years. I could not have asked for better mentors and friends.

I would like to thank my mother Nursen Karaman. I would not be able to complete this work without her encouragement. I am forever indebted to my husband Iban Auzmendi Juiz for his endless love and support throughout this endeavor.

I gratefully acknowledge the financial support of the Goizueta Business School, the Sheth Foundation, and the Laney Graduate School.

Atlanta, Georgia H. K.

April, 2018

Table of Contents

FIGURES

TABLES

CHAPTER I: Introduction

The importance of online word of mouth (WOM) in customer decision making is significant. Online WOM now spans a significant swath of the Internet, and online consumer reviews constitute one of the most predominant types of online WOM. Individuals use online reviews to resolve uncertainty about a product's quality. According to a study, for more than 80% of customers, online reviews play a very important or somewhat important role when deciding on a product to purchase. Further, online reviews have become the most trusted source for accurate product information for customers. Internet users trust online reviews more than they would trust friends, family, colleagues or a message from the brand itself. Consequently, it is becoming more and more important for marketers to understand online reviews.

In this dissertation work, I study two distinct aspects of online WOM in two chapters. The first chapter of my dissertation focuses on identifying the extent to which social influence impacts online reviews; the second chapter addresses how online reviews of competitors shape consumer decision making. Overall, my dissertation work aims at providing managerial insights into how online reviews evolve over time and how they impact customer decision making. Additionally, I outline my plan to study how managers should respond to online reviews in the third chapter.

CHAPTER II: The Effect of Social Influence on Online Reviews

1. INTRODUCTION

The Internet is increasingly changing not only the way customers shop, but also the way they gather and share information. The majority of customers trust online reviews as much as personal recommendations. According to one survey, as many as 88% of customers state that they rely on online customer service reviews when making a buying decision (Zendesk 2013). In line with this report, previous research consistently documents that a product's online review characteristics (e.g., valence, volume and variance of its reviews[1]) are significantly associated with its demand in various contexts (Archak, Ghose, and Ipeirotis 2011; Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Luca 2016a; Moe and Schweidel 2012). According to Yelp (2016), the leading review website that hosts crowd-sourced reviews about local businesses, such as restaurants, the number of average monthly visitors to its website was 167 million during the 1st quarter of 2016, and the number of cumulative reviews contributed since its inception reached 102 million by March of 2016. Additionally, the rate at which new reviews arrive follows an exponential pattern, and the increase in the number of average monthly visitors has not yet dwindled. Hence, the importance of online reviews in customer decision making is projected to increase. As such, from a managerial standpoint, it is becoming increasingly important to understand the dynamics that influence customer opinions posted on online review websites.

This study primarily investigates the dynamics in online product ratings due to social influence. In this paper, social influence is specifically defined as a reviewer's tendency to report an online review rating that differs from his private satisfaction rating due to exposure to others' review ratings. Our main variable of interest is the online review rating posted by a reviewer. We argue

[1] Research on online customer reviews mainly focuses on the impact of three metrics derived from a product's online reviews on its demand (King, Racherla, and Bush 2014). These metrics include valence, volume and variance. In general, valence is measured by the average review rating, volume is measured by the number of total reviews, and variance is measured by the variance of ratings.

that a customer's online review rating is a function of two important factors: (1) purchase experience, i.e., satisfaction level with the product in the purchase occasion for which the review is submitted, and (2) social influence, i.e., opinions previously expressed by other people. The first factor represents an individual's private opinion about the product, whereas the second one influences his or her publicly expressed opinion about the product in an online review. In this paper, we investigate (1) whether social influence in online reviews exists, and (2) if so, how social influence shapes online ratings posted by reviewers.

Two seminal studies that specifically focus on quantifying the amount of social influence on online reviews conclude that the magnitude of social influence on online product ratings is substantial. First, (Wendy W Moe and Trusov 2011) decompose ratings into a baseline rating, the contribution of social influence and an idiosyncratic error. They report that, on average, social effects decrease ratings valence and volume while increasing ratings variance. The second study takes the literature of social influence in online reviews a step further by accounting for a reviewer's current experience characteristics while investigating the extent of social influence. Sridhar and Srinivasan (2012) infer a customer's current experience from the textual elements in his review and show that other customers' online ratings moderate the effect of a reviewer's own product experience on the online rating that he eventually posts. Similar to (Sridhar and Srinivasan 2012), we specifically focus on the impact of social influence on review valence.

This study expands earlier work (Wendy W Moe and Trusov 2011; Sridhar and Srinivasan 2012) by directly incorporating a reviewer's observed private evaluation and by constructing control and treatment groups that allow us to establish both the existence and the direction of social influence with precision. The current study contributes to research on social influence in four important ways. First, we take into account the private[2] satisfaction level with the reviewed purchase occasion whereas previous research treats current purchase experience as a latent variable (Lee, Hosanagar, and Tan 2015; Moe and Schweidel 2012; Wendy W Moe and Trusov 2011) or infers it from review content (Sridhar and Srinivasan 2012). Controlling for the actual satisfaction level of customers before they post allows us to identify the extent of social influence more precisely. Additionally, even though constructing a customer's current experience through the textual elements in his review is innovative, the review content itself could actually be subject to social influence and thus may not completely reflect his private opinion.

Second, the exogenous intervention employed by the firm in the design of customer review solicitation allows us to construct a control group against which a treatment group is compared. The control group consists of randomly selected individuals who are invited to post a review after completing the customer satisfaction survey. For these survey takers, posting takes only a few clicks, and throughout the process, they are not shown any other review on the website. These individuals form the control group for which the effect of social influence is expected to be minimal or none because they are not likely to read others' opinions already posted on the website. On the other hand, the treatment group consists of individuals who are not invited to post a review through random selection, but still post an online review. These individuals belong to the treatment group because they are subject to social influence as they are likely to be exposed to others' already existing opinions when posting their reviews. After matching the two groups on numerous bias-inducing factors, such as private satisfaction levels, a comparison between posting behavior of the matched control and the matched treatment groups establishes the causal effect of social influence.

[2] In this study an individual's satisfaction score is private in the sense that it is only observed by the firm and the researcher. Only his or her online review rating is made public.

Third, this study shows how people resolve potential divergences between their private evaluations and the average opinions of others. Extant research on online reviews ignores this issue and primarily focuses on the impact of disagreements amongst existing reviewers (measured by the variance of posted review ratings) on subsequent review posting behavior. We show that reviewers conform to the average opinions expressed in previously posted reviews. To our knowledge, this type of conformity has not been documented previously in the context of online reviews.

Finally, we adopt a methodology called coarsened exact matching (CEM) to discover causal associations (Iacus, King, and Porro 2012). This methodology allows us to match the control and treatment groups on variables that play a significant role in determining the online rating posted by an individual. Consequently, the differences in online ratings posted between these two groups can be attributed to social influence. Moreover, this methodology allows us to overcome self-selection concerns. Thus, a secondary contribution of this paper is to illustrate the use of CEM for research in marketing.

We carry out this study based on a unique data set from a multi-chain hotel group that conducts satisfaction surveys and hosts online reviews on its own website. The dataset consists of individual level satisfaction survey response and online review activity of all individuals who took a survey or wrote a review at any time between January of 2012 and May of 2015. Not all satisfaction survey takers post online reviews nor do all online reviewers take satisfaction surveys. For that reason, we match each satisfaction survey with a corresponding online review (provided that it exists) using a unique combination of customer and reservation identification codes and construct the treatment and control groups using only matched pairs. In doing so, we ensure that both the private satisfaction score and the public online review rating posted for a given purchase occasion are known for the entire population used in the analysis.

In total, we have access to approximately 443,000 satisfaction surveys that are matched with a corresponding online review. The vast majority (95%) of these reviews were collected through solicitations made by the hotel group at the end of randomly selected satisfaction surveys and constitute the control group. Consequently, the control group has approximately 420,000 observations and the treatment group has approximately 23,000 observations that consist of pairs of completed satisfaction surveys and posted online reviews. We do not use the entire dataset to estimate our model because there exists a slight amount of imbalance in pre-treatment covariates between our control and treatment groups. Instead, in order to eliminate the imbalance in pre-treatment covariates across the two groups, we employ CEM methodology and match the two groups on these pre-treatment covariates, such as the exact hotel, check-in month and year, and individual's private satisfaction score. Consequently, we estimate the model using matched observations. The final number of observations used from each group in model estimation vary depending on the matching variables/specification.

Our analysis shows that social influence operates on online rating behavior in systematic ways. Our first finding indicates that when there is no considerable difference between an individual's satisfaction score and the average rating of previous reviews before his review, an individual who is subject to social influence is more likely to post an online review rating lower than an individual who is not subject to social influence. This treatment effect is consistent with (Schlosser 2005) finding, which shows that reviewers post less favorable product ratings due to their desire to appear more knowledgeable or to have higher standards. Our second finding indicates that reviewers in the treatment group are significantly more likely to adjust their online review ratings in order to conform to the majority opinion already existing on the website. We label this effect as conformity effect. We argue that the conformity effect arises because reviewers are uncertain that they can objectively and fully evaluate a hotel on their own and therefore incorporate others' already expressed opinions in their public online review rating.

We organize the remainder of the article as follows: First, we discuss the role that social influence plays in customer decision making in general and in the context of online reviews. Next, we provide a detailed description of the data and describe the empirical analysis. Finally, we present our results and conclude with a discussion of this study's contributions to our understanding of social influence in online reviews, its managerial implications, its limitations and the opportunities for future research.

Social Influence on Customer Decisions and Evaluations

While this study focuses on the role of social influence in an online review environment, it builds on research showing that social influence operates more broadly in consumer decision making. Research in consumer behavior and social psychology has generated ample evidence that the decisions and judgments of individuals are influenced by the decisions and judgments of others. Consequently, changes in customer choice and opinion can be induced as a result of social influence. For instance, Venkatesan (1966) showed that individuals follow the majority in deciding which of three identical suits is of better quality. Similarly, Asch's (1955) classic experiment documents that individuals' perceptions depend upon others' evaluations despite the existence of a single correct answer and that individuals go along with the majority renouncing the evidence of their own senses. Furthermore, his work emphasizes that social influence is substantial when individuals are forming judgments in the presence and visibility of others' opinions. A reviewer who is composing his online review more often than not does so in the presence and visibility of others' views expressed in previous reviews. Therefore, Asch's (1955) study provides an additional basis for why social influence may exist in the context of online reviews. In general, this form of social interaction where others' expressed opinions influence individuals' product perceptions and choice is termed word of mouth (WOM).

On the other hand, individuals' perceptions and decisions are not only influenced by others' expressed opinions but also by others' actions. This form of social interaction is usually called observational learning. Previous studies provide substantial evidence for observational learning. For instance, Salganik, Dodds, and Watts (2006) establish that an individual's decision to listen to and download a song is affected by the number of times the song was previously downloaded. Similarly, Tucker and Zhang (2011) show that an individual's decision to visit a wedding vendor's website is driven by the number of previous clicks it received. By the same token, Cai, Chen, and Fang (2009) demonstrate that when customers are given ranking information of the five most popular dishes, the demand for those dishes increases by 13 to 20 percent due to observational learning.

Chen, Wang, and Xie (2011) emphasize that the distinctive feature of observational learning is the amount of information it contains compared with WOM. Consistent with previous research (Bikhchandani, Hirshleifer, and Welch 1998), they argue that observational learning contains less information because observational learning information contains only the actions of others but not the reasons behind their actions, whereas WOM information contains both others' opinions (e.g., ratings) and the reasons behind their assessments (e.g., review text). Thus, our study examines consumer behavior that is closer to the first form of social interaction, namely WOM. WOM is shown to influence customers' trial and adoption of a product (Arndt 1967), evaluations of and attitudes toward a product (Bone 1995) and switching behavior (Wangenheim and Bayón 2004).

Based on previous WOM research, we expect a great deal of social influence to occur in an online reviews environment.

One of the main consequences of social influence is conformity (see Ariely and Levav (2000) for an exception). For instance, Burnkrant and Cousineau (1975) show that after observing favorable evaluations of others, individuals perceive the coffee product more favorably themselves than they would have in the absence of this observation. Furthermore, conformity is more broadly observed in various other domains, such as political voting (Cukierman 1991; Muchnik, Aral, and Taylor 2013), trading in financial markets (Alevy, Haigh, and List 2007), stock market participation (Hong, Kubik, and Stein 2004), and even an individual's decision to become a criminal (Glaeser, Sacerdote, and Scheinkman 1996).

Cialdini and Goldstein (2004) define conformity as "the act of changing one's behavior to match the responses of others". They argue that an individual's motivation to conform may stem from (1) his desire to form an accurate interpretation of reality and behave correctly or (2) his goal of obtaining social approval from others. This argument is based on Deutsch and Gerard (1955) who distinguish normative conformity (e.g., Bernheim 1994) from informational conformity (e.g., Burnkrant and Cousineau 1975; Salganik and Watts 2008). They argue that normative conformity is based on individuals' goal of obtaining social approval from others, whereas informational conformity is based on individuals' desire to accurately interpret reality by incorporating information obtained from others. In the online reviews context, we expect informational conformity to play a greater role. In our empirical context, regardless of a reviewer's own experience with the hotel, she is likely to be uncertain about all the true quality dimensions of the hotel. For instance, a reviewer may have doubts about her expertise in evaluating the hotel's quality with respect to dining. For that reason, she may be influenced by others' reviews in the sense that she may accept information obtained from other reviewers as evidence about the objective reality. As a result, conformity may arise in an online reviews setting.

Social Influence on Online Reviews

The preceding discussion on the role of social influence in customer decisions and evaluations provides a foundation for why one might expect social influence to play a significant role in the context of online reviews. In this research, we are primarily interested in understanding the role of others' expressed opinions, i.e., social influence, in shaping subsequent reviewers' online rating behavior. The term "social influence" is broadly used in online reviews research to capture various social dynamics (Moe and Schweidel 2012). On the one hand, social influence may refer to the impact of previous reviews on encouraging or discouraging subsequent positive or negative review posting, i.e., individuals' decision of whether to post. On the other hand, social influence may refer to the impact of previous reviews on individuals' decision of what to post conditional on having already decided to post. This study uses the term "social influence" in line with the latter description. Although the decision of whether to post could be affected by previous reviews, we abstract away from that effect and specifically explore how a reviewer's observed private and public (possibly subject to social influence) opinions differ because of others' previously posted reviews.

To the best of our knowledge, the only other study that focuses on potential differences between public and private evaluations by observing both is Schlosser's (2005) experimental work. In an early effort in examining the existence of such social influence in online posting behavior, Schlosser (2005) shows that reviewers post less favorable product ratings after seeing a negative review than a positive review or no review. She argues that such a negative adjustment stems from the poster's desire to appear more knowledgeable or to have higher standards. Consequently, her work reinforces the notion that previously posted reviews may affect ratings posted by future reviewers.

Amabile and Glazebrook (1982) showed that individuals lean toward negative criticism as a strategy of impression management and labeled this tendency as negativity bias. They argued that the negativity bias stems from an individual's desire to be perceived as more intelligent. Amabile (1983) found that a negative reviewer was seen as more intelligent and competent than a positive reviewer with an equivalent quality of writing. These two findings imply that not only individuals use negativity to demonstrate their intelligence, but also they find others who express negatively critical evaluations to be more intelligent. Consequently, these works coupled with Schlosser's (2005) suggest that posters who are subject to social influence may have a tendency to post lower product ratings in order to appear more knowledgeable or competent.

Additionally, we expect to detect conformity amongst reviewers. Conformity stems from the fact that a reviewer shapes his review in the visible presence of others' reviews and opinions. There is ample evidence that an individual who is forming his opinion after having heard or read about others' opinions is influenced by those judgements (Asch 1955; Burnkrant and Cousineau 1975). We expect this pressure to conform to be informational in nature rather than normative. It is difficult for an individual to completely and accurately evaluate a hotel's objective quality from his direct experience. For instance, he may not have observed or experienced all characteristics of the hotel, or his evaluation standards may not represent the complete or correct set of standards. For that reason, he may deem others' opinions as evidence about the "true" nature of the product and adjust his private evaluations to reflect the product's "true" quality.

Conformity is usually an easy decision path because generating a defensible and divergent position requires a certain level of cognitive effort. In their review article, Lerner and Tetlock

(1999) report that individuals are inclined to conform especially when they must explain their opinions to an audience with known views. In general, a reviewer is expected to provide evidence that justifies his rating in his review text. For that reason, we expect to discover that reviewers adjust their posted ratings in order to reduce the discrepancy between their private evaluations and the average opinion of others. In order to uncover such adjustments, we construct a variable labeled as *opinion deviation*. This variable simply captures the difference between a reviewer's private evaluation and the average rating of online reviews posted before his review.

While research (Moe and Schweidel 2012; Wendy W Moe and Trusov 2011; Nagle and Riedl 2014) has investigated the impact of disagreement (measured by the variance of ratings) among previous reviewers on subsequent posting behavior, this study is novel in that it takes into consideration a different kind of disagreement: the disagreement between a poster's private evaluation and the average opinion of other reviewers. Focusing on how consumers handle their disagreements with the average opinion of others enables us to study conformity in online review posting behavior[3].

In summary, we expect individuals who are exposed to previous reviews, i.e., subject to social influence, to be more likely to post online ratings that are lower than their control group counterparts who are not exposed to others' reviews. We argue that this treatment effect stems from an individual's desire to be perceived as more knowledgeable or capable of incisive criticism. Furthermore, in light of previous research on conformity, we expect to detect that reviewers try to

[3] Previous research has produced mixed results in terms of the impact of consensus on conformity. Asch's (1955) seminal work suggests that consensus may strengthen social influence. However, Burnkrant and Cousineau (1975) find that the uniformity of previous evaluations does not significantly impact an individual's evaluation. Further, Bone (1995) reports that consensus is not needed for WOM to have an influence. In our context, we do not examine the impact of consensus because in this study, individuals who are subject to social influence only see the average rating of previously posted online reviews. The rating distribution of previously posted online reviews (hence information on consensus) for each hotel is not readily available on the website.

conform to the average opinions of others, i.e., they adjust their private satisfaction levels considering others' opinions before posting their online review ratings publicly.

2. DATA

The data consist of individual level satisfaction survey responses and online reviews for all individuals who took a survey or wrote a review sometime between January of 2012 and May of 2015 for all hotels that belong to a multi-chain hotel group. The hotel group, which wishes to remain anonymous, provided the survey data, and we collected the online review data from the hotel group's website. Satisfaction surveys provided by the hotel group were matched to online reviews collected from the hotel group's website by using unique combinations of customer and reservation identification codes. It has been shown that some online reviews could be submitted by customers with no record of ever purchasing the product they are reviewing (Anderson and Simester 2014). However, online reviews without a purchase do not exist in the current study because the hotel group makes sure that all reviews posted on its website are submitted by actual hotel guests. In total, there are 443,385 surveys that are matched with online reviews collected from the company's website. The treatment and control groups are constructed from these 443,385 matched pairs of satisfaction surveys and online reviews.

In order to track the performance of hotels in its portfolio, the hotel group conducts surveys. Survey solicitations are made via emails, and the average response rate is approximately 2%. The majority (80%) of these survey takers are loyalty program members. In this study, we focus on loyalty program members because matching survey responses to online reviews is not possible for nonmembers. The hotel group has been hosting online reviews on its website since 2012. In an attempt to increase the number of reviews posted on its website, in 2014, the hotel group started inviting randomly selected survey takers to post their guest experience as an online review. 48% of these invited survey takers accepted the invitation to post an online review. Consequently, these randomly selected survey takers could easily post their opinions as online reviews with a few easy steps. In order to streamline the process of posting an online review, the hotel group automatically generates review ratings based on reviewers' survey responses. Individuals report satisfaction score on a scale from 1 to 10 in the survey[4]. When converting these satisfaction scores into online review ratings out of 5, the hotel group simply divides the survey satisfaction score by two and rounds it up to the nearest integer value[5]. Next, the hotel group presents converted ratings to its surveyors, and surveyors are allowed to change these ratings if they disagree with the converted ratings provided by the hotel group. Additionally, surveyors are assured that their edits will only affect their online review, not the original survey answers.

One of the prominent features of this solicitation strategy is that a hotel guest, who posts a review following these steps after completing his or her survey, is not exposed to other reviews already posted on the website. We do not claim that these hotel guests are completely unaware of reviews that are already posted online because they may have seen these reviews at other instances, such as at the time of booking. However, they are less likely to read others' opinions during posting because such information is not readily provided to them in the process of review creation. All online reviews posted in this fashion form the control group of this study. We expect the effect of social influence to be minimal for these individuals.

All other reviewers have to first navigate to the website of the specific hotel for which they want to post an online review. The website displays summary statistics of previously posted reviews right next to the "write a review" button. These summary statistics include the average rating of previously posted reviews and the total number of reviews posted to date. Additionally,

[4] In this study, satisfaction scores are only observed by the firm and the researcher. These satisfaction scores are considered to be private evaluations of survey takers. On the other hand, online review ratings are publicly posted.
[5] For example, a satisfaction score of 5 (or 6) is automatically converted into an online rating of 3.

the same webpage exhibits ratings and full texts of the last eight reviews posted on the website. Therefore, hotel guests who post a review following this conventional process are more likely to be influenced by social factors and form the treatment group of this study.

In total, we have 443,385 observations that consist of pairs of completed surveys and online reviews. The majority (94.8%) of these reviews are collected through the above explained review solicitation strategy. For that reason, the size of the control group is much larger than that of the treatment group. We have 420,503 observations in the control group and 22,882 observations in the treatment group. Figure 2.1 summarizes how these two groups are created. Having a large reservoir of control observations helps reduce the bias problem. The final number of observations used from each group in model estimation varies depending on the matching specification as explained in Section 3.3.

[Insert Figure 2.1 here]

The way in which the hotel group selects which individual to invite for converting their filled survey to an online review is random. One might suspect that the hotel group may be interested in soliciting reviews from hotel guests who fill in surveys with higher satisfaction scores, but we do not observe that in our data. Figure 2.2 shows that regardless of the satisfaction score, on average, 30% of surveys are selected for review invitation. Even though we make sure to reduce the imbalance of pre-treatment covariates between our control and treatment groups by employing CEM methodology, it is reassuring to know that the hotel group is not engaging in strategic behavior.

[Insert Figure 2.2 here]

Evidence of Social Influence

In this section, we contrast the online review posting behavior of hotel guests in the control and treatment groups. We argue that hotel guests in the control group are less affected by social factors than their counterparts in the treatment group. Therefore, we expect hotel guests in the control group to post online ratings that are very similar to what their satisfaction scores imply. Figure 2.3 lends support to this argument and shows that the distribution of online ratings (conditional on survey satisfaction scores) for hotel guests in the control group diverges much less from the distribution of their satisfaction scores compared to the distribution of online ratings for hotel guests in the treatment group.

More specifically, Figure 2.3 demonstrates that hotel guests who belong to the control group are very unlikely to post an online rating different from their satisfaction scores. However, hotel guests in the treatment group exhibit big deviations from their satisfaction scores when posting their online reviews. For instance, consider individuals whose satisfaction score is 7. As seen in Figure 2.3, those in the control group are observed to post an online review rating of 4 (no change) 88% of the time, whereas the same probability is 52% for treated individuals. There is a 33% chance that individuals in the treatment group will post an online rating of 3 (downgrade), whereas the same is observed only 11% of the time for individuals in the control condition. In other words, compared to the control group, the treatment group is three times as likely to downgrade while posting an online rating. It is evident that hotel guests in the treatment condition are changing their online ratings significantly after presumably reading other people's reviews and opinions. The impact of social influence is, therefore, primarily identified from the online rating behavior of treated reviewers. Next, we provide details of the empirical analysis implemented.

[Insert Figure 2.3 here]

3. EMPIRICAL ANALYSIS

3.1. Variables

Dependent variable. The main dependent variable of this study is the online review rating posted by a reviewer. The hotel group's website allows reviewers to give a rating between 1 and 5 for the hotel that they are evaluating. Since the dependent variable has five categories with a meaningful sequential order, we use ordered logit to model the online review ratings posted.

Independent variables. In this research, we argue that online ratings are shaped by two primary factors: (1) the satisfaction level with the current stay experience and (2) social influence. The first factor is measured by the satisfaction score reported by an individual on his survey response, and consequently, we label this variable as *satisfaction score*. In order to capture the impact of the second factor, social influence on online reviews, we use two independent variables: (1) a dummy variable for treatment and (2) a variable that measures the difference between an individual's private satisfaction score and the average review rating of previous reviews posted on the hotel's website. First, the coefficient estimate for the treatment dummy variable captures the treatment/exposure effect of social influence. This coefficient represents the difference between the average online ratings posted by control and treatment groups holding everything else constant and conditional on there being no significant difference between an individual's private satisfaction score and the average review rating of previously posted reviews.

Second, we label the variable that measures the difference between a reviewer's satisfaction score and the average online rating of previously posted reviews as *opinion deviation.* The opinion deviation variable measures how a reviewer's satisfaction level compares with the average rating of reviews posted before his own review. If his opinion deviation is positive, it means that his current experience with the specific hotel is more positive than the average experience reported on the hotel website. On the other hand, if his opinion deviation is negative, it implies that his current experience is worse than the overall experience reported by others on the hotel website. More precisely, opinion deviation of an individual, *i*, for a hotel, *h*, at time, *t*, is calculated as follows.

$$Opinion\ deviation_{ith} = \frac{Satisfaction\ score_{ith}}{2} - Online\ Average\ Rating_{ith} \qquad (1)$$

In order to re-scale the satisfaction score from a scale of 1 to 10 to a scale of 1 to 5, we divide it by 2[6]. Thus, both terms on either side of the minus sign on the right hand side of the equation (1) are measured out of 5.

The coefficient estimate of opinion deviation determines whether reviewers engage in conforming behavior. For instance, a positive coefficient estimate of opinion deviation suggests polarizing/differentiating behavior. In other words, it implies that individuals whose satisfaction score indicates a more positive experience than the online average rating are more likely to post an online review rating that is even higher than their satisfaction score, and individuals whose satisfaction score indicates a more negative experience than the online average rating are more likely to give an online review rating that is even lower than their satisfaction score. On the other hand, a negative coefficient estimate of opinion deviation means that individuals attempt to conform to the majority opinion already expressed on the website. A negative estimate of opinion deviation implies that individuals whose satisfaction score is lower than the average online rating are more likely to post an online rating that is higher than their satisfaction score and that individuals whose satisfaction score is greater than the average online rating are more likely to give an online rating that is lower than their satisfaction score.

[6] Figure 2.3 provides justification for dividing the satisfaction score by 2. It shows that 96% of control group individuals whose satisfaction score is 2 post an online rating of 1, 93% of control group individuals whose satisfaction score is 4 post an online rating of 2, 91% of control group individuals whose satisfaction score is 6 post an online rating of 3, 96% of control group individuals whose satisfaction score is 8 post an online rating of 4 and finally, 100% of control group individuals whose satisfaction score is 10 post an online rating of 5.

3.2. Empirical Strategy

The fact that the hotel group randomly invites survey takers to post their opinions as online reviews greatly reduces the potential for bias resulting from individual differences or self-selection. However, some characteristics, either hotel-specific or individual-specific, could still cause bias. In order to reduce bias from these pre-treatment covariates, we use CEM algorithm (Iacus, King, and Porro 2011, 2012) to match an individual in the treatment group to individuals from the control group on certain attributes that may play a significant role in determining online ratings. By doing so, we create a subset of observations that matches on all sample moments of the treatment and control groups, and we use this subset of observations to carry out model estimation. In Section 3.3, we discuss these matching specifications in greater detail.

The main objective of matching is to increase the similarity between the empirical distributions of covariates in the treatment and control groups. The idea is that the more similar the treatment and the control groups are in their observed characteristics, the more likely they are to be identical in their unobserved characteristics. CEM entails "coarsening" observed attributes and performing exact matching on the coarsened variables (Iacus, King, and Porro 2012). Widely used current matching methods, such as propensity score and Mahalanobis matching, belong to the class of matching methods called "equal percent bias reducing" (EPBR). Iacus, King, and Porro (2012) note that EPBR does not guarantee any level of imbalance reduction in any given data set and ignores imbalance due to differences in variances, ranges, covariances, and higher order interactions. In order to avoid these problems with EPBR, Iacus, King, and Porro (2011) propose CEM that belongs to a broader class of "monotone imbalance bounding" (MIB) matching methods, which guarantee that the imbalance between matched treatment and control groups will not be larger than the pre-matching levels. Because the matching between the two groups is exact, no propensity score is estimated.

More specifically, the implementation of CEM consists of two steps. First, online reviews are classified into strata based on pre-determined matching variables. Second, strata containing only treated or control observations are discarded and normalized CEM weights are calculated for each observation. Finally, we estimate the model using CEM weights and clustering standard errors at the strata level. Next, we turn to explain the matching variables used in CEM.

3.3. Matching variables

3.3.1. Reducing bias from satisfaction level differences between the control and the treatment groups. We match reviewers in the control and treatment groups on their survey satisfaction scores for all matching specifications. This ensures that both groups are identical in terms of reviewers' private satisfaction scores. Reviewers who come to the website to post a review on their own, i.e. who are in the treatment group, may be driven by a certain level of either positive or negative experiences that differ from those who are in the control group. For instance, reviewers who choose to go to the website on their own to write a review could be more negative than those who accept to post a review after being invited by the company. By matching the two groups on survey satisfaction scores, we make sure that the pre-treatment satisfaction scores of both groups are the same. Consequently, we can attribute differences in their online ratings to social influence and eliminate the possibility of biasing the social influence effect due to differences in their private satisfaction levels of the control and treatment groups.

3.3.2. Reducing bias from fixed hotel attributes. An important variable that may play a significant role in online ratings is the hotel itself. For that reason, we match the control and treatment groups either on the exact hotel or on certain hotel attributes that include hotel chain, hotel location, hotel state, hotel country and hotel star level.

3.3.3. Reducing bias from time trends. Online ratings are shown to evolve systematically over time and sequence (Godes and Silva 2012). Previous work shows that the average ratings follow a declining trend both over time and order (Godes and Silva 2012; Li and Hitt 2008; Wu and Huberman 2008). Several explanations are provided for this tendency. On the one hand, Li and Hitt (2008) argue that this trend can be attributed to a changing customer base as it evolves from early adopters with stronger preferences for the product to late adopters with weaker preferences. On the other hand, Wu and Huberman (2008) attribute this trend to the tradeoff between the benefit and cost associated with posting. They argue that because online average ratings are positive-leaning due to purchasing bias, reviewers can more easily justify negative reviews given the cost of writing a review. Findings from Godes and Silva (2012) lend additional support to this argument. However, this systematic trend should not be confused with the social dynamic identified in the current study. Because we match the control and treatment groups on the year and month of the consumption experience[7], the social influence identified in this study is not driven by any time trends.

By conditioning on check-in month and year, we are also reducing any bias that may result from differences amongst service staff across time. Typically, check-in experience plays a significant role in overall stay evaluation, and conditioning on check-in month and year allows us to reduce any bias that may result from individuals' check-in experiences, either in terms of having longer/shorter wait times or in terms of having more/less attentive front desk personnel.

3.3.4. Reducing bias from reviewer's previous stay experience. Previous research on social influence in online reviews often overlooks the possibility that posters may have repeated encounters with the same firm. For instance, Bowman and Narayandas (2001) show that customers, who initiate communications with manufacturers, are among a brand's most loyal customers. In line with Hirschman (1970), they argue that loyal customers want the manufacturer

[7] On average individuals post an online review within 9 days from their check-in date.

to do better because they have more at stake with the focal brand than their less loyal counterparts, and for that reason, they are more likely to voice their opinions. Thus, loyal customers may exhibit different posting behaviors and be less likely to be swayed by social pressure. Consequently, it is important to condition on the relationship history between the reviewer and the firm. We accomplish this by matching the control and treatment groups on the number of cumulative stays a reviewer had until the review date. Therefore, the rating difference between the two groups cannot be attributed to customer loyalty.

3.3.5. Reducing bias from reviewer's previous review posting experience. In a similar vein, we condition on an individual's previous review posting behavior because Moe and Schweidel (2012) show that previously posted opinions impact individuals differently depending on their intensity of posting behavior. They find that the least active posters are more positive, whereas the most active posters tend to be more negative in an attempt to be perceived as "experts" by differentiating themselves through posting negatively skewed opinions. Thus, we match the control and treatment groups on reviewers' previous posting behavior in order to make sure that the rating difference between the two groups is not driven by posting experience.

Additional variables that we match on as part of alternative matching specifications include the individual's membership level and whether the individual booked his stay online. Table 2.1 shows the reduction of imbalance across the treatment and control groups after the implementation of CEM methodology.

[Insert Table 2.1 here]

We matched the treatment and control groups using three different specifications where the matching variables were determined in alternative ways. First, we matched observations on four variables that include exact hotel, check-in month and year and the satisfaction score reported on the survey, i.e. the private evaluation. This first matching strategy shrinks the dataset to 19,453 observations. Second, we matched observations using four additional variables to further account for individual differences. These four additional variables are membership program status (e.g. club, gold or platinum), a dummy variable indicating whether the individual booked his stay online, the number of previous online reviews written on the company website and the number of cumulative bookings at the current hotel brand. Since this second strategy involves additional matching variables, it further shrinks the dataset to 5,042 observations for analysis. Finally, in our third alternative matching specification, we loosen the matching criteria from exact hotel to a set of hotel specific variables that includes hotel chain, location, state, country and star level. Consequently, we matched the two groups on twelve variables that include hotel chain, hotel location, hotel state, hotel country, hotel star level, check-in month and year, satisfaction score reported on the survey, membership program status (e.g. club, gold or platinum), a dummy variable indicating whether the individual booked his or her stay online, the number of previous online reviews written on the company website and the number of cumulative bookings at the current hotel brand. This third matching specification reduces the dataset to 31,342 observations for analysis. These three matching specifications highlight the trade-off between matching granularity and the number of dropped observations. For instance, the second matching specification imposes the most stringent restrictions for units within each stratum to be similar and reduces the observation size to 5,042 whereas the third matching specification has the most lenient constraints for them to be similar and reduces the observation size to only 31,342. In the next section, we present our estimation results for all three matching specifications.

4. RESULTS

Table 2.2 reports our estimation results. The dependent variable for all models is the online review rating posted. Consistent with the premise of this research, independent variables include survey satisfaction score and social influence variables. The following discussion is based on Models 1.3, 2.3 and 3.3 in Table 2.2. First, the coefficient estimate of the satisfaction score indicates that the online review rating posted by a reviewer is primarily driven by his or her satisfaction level (i.e., private evaluation) with the stay experience. Second, the coefficient for the treatment dummy variable is estimated to be negative and statistically significant. This result indicates that reviewers who are subject to social influence are, on average, more likely to post an online review rating that is lower than reviewers in the control group, conditional on there being no difference between their satisfaction score and the average rating of previous reviews before their review, i.e., their opinion deviation being zero. This result is conditional on opinion deviation being zero because the model includes an interaction term between the opinion deviation and the treatment dummy variables. This finding is consistent with the idea that individuals may post lower ratings in order to be perceived as knowledgeable or discriminating (Schlosser 2005).

Third, the conforming behavior is greatly pronounced for reviewers in the treatment group. This finding suggests that treated individuals whose experience is more positive than the average rating of previously posted reviews are more likely to post an online rating that is lower than their private opinion in order to conform to the majority opinion and are more likely to post an online rating that is lower than their counterparts in the control group. Similarly, treated individuals whose satisfaction scores are lower than the average online rating of previously posted reviews adjust their online rating to conform to the majority opinion and are more likely to post an online rating that is higher than their counterparts in the control group. Finally, if there is no difference between a reviewer's satisfaction score and the average rating of previous reviews then conformity pressure does not exist. In such instances, social influence only operates through the main treatment effect identified with the treatment dummy variable. Compared to the conformity effect for the treatment group, the conformity effect for the control group is considerably smaller[8] and statistically insignificant at 5% significance level[9]. As evident in Table 2.2, these three main findings are consistent across all models estimated using three different matching specifications.

[Insert Table 2.2 here]

4.1. Social Influence

The model coefficient estimates imply two social forces that are at play: conditional on the same private evaluation, (1) individuals who are subject to social influence are more likely to post a lower online rating than their counterparts in the control group when there is no considerable difference between their private evaluation and the average rating of previous reviews before their review, and (2) individuals who are subject to social influence are more likely to adjust their online ratings up or down while posting their opinions publicly in order to conform to the average rating of previously posted reviews.

The coefficient estimates of an ordered logit model are hard to interpret due to the nonlinear nature of the model in consideration. Therefore, in order to illustrate the extent of these social effects, we simulated 10,000 independent sets of model parameters drawn from their estimated distribution based on Model 1.3 in Table 2.2 (Tomz, Wittenberg, and King 2003). This simulation

---

[8] χ^2 statistic for testing the equivalence of coefficients of opinion deviation and opinion deviation*treated variables is 20.04 using Model 1.3. *P*-value is 0.000, suggesting that these two coefficients are significantly different from each other.

[9] The non-significant effect of opinion deviation on review ratings within the control group alleviates concerns that some individuals in this group (1) may have deliberately navigated to the hotel website in order to check opinions of other reviewers before posting their survey as a review or (2) may have read previous reviews at any other point in time, for example at the time of booking.

exercise takes into account the uncertainty associated with each parameter that is estimated. We used these simulated parameters to create Figure 2.4. Figure 2.4 demonstrates the behavioral difference between the treatment and control groups in posting a particular online rating conditional on a particular satisfaction score[10], and thus illustrates the impact of social influence on online ratings. In our data, the majority of survey takers reported satisfaction scores of either 9 (25%) or 10 (33%). For that reason, Figure 2.4 is drawn conditional on a satisfaction score of either 9 or 10.

[Insert Figure 2.4 here]

First, the x-axis in each figure denotes opinion deviation. Opinion deviation variable measures the difference between a reviewer's satisfaction score (i.e., private evaluation) and the average rating of all reviews posted before her review (i.e., the average opinion of others). For that reason, a positive opinion deviation implies that the reviewer's satisfaction level is higher than the average rating of all reviews posted up until her review. Similarly, a negative opinion deviation suggests that the reviewer's satisfaction level is lower than the average rating of all reviews posted up until her review. Conditional on a specific satisfaction score, the observed range of opinion deviation changes. For instance, Figure 2.4.1 is drawn conditional on a satisfaction score of 10. Given a satisfaction score of 10, the observed opinion deviation range in our dataset is for the most part confined to [0, 1.5]. This limitation is due to two factors: (1) the highest average rating of all previous reviews cannot exceed 5, thus the lowest value of opinion deviation is zero (0) ($\frac{10}{2} - 5$), and (2) it is highly unlikely for the average rating of all previous reviews to be lower than 3.5, and as a result, the highest value of opinion deviation is 1.5 ($\frac{10}{2} - 3.5$). By the same token, the observed

[10] Additional figures for all other possible combinations of posted online ratings and survey satisfaction scores are available from authors upon request.

range for opinion deviation is [-0.5, 1] conditional on a satisfaction score of 9. The lower bound is calculated by ($\frac{9}{2} - 5$), whereas the upper bound is calculated by ($\frac{9}{2} - 3.5$). Accordingly, the x-axis of Figures 2.4.2.1 and 2.4.2.2 is restricted to [-0.5, 1] because both are drawn conditional on a satisfaction score of 9.

Second, the y-axis denotes the 95% confidence interval for the probability difference between the treatment and control groups in posting a particular online rating. For example, Figure 2.4.1 shows that conditional on a survey satisfaction score of 10, on average, an individual in the treatment group is approximately 4% less likely to post an online rating of 5 compared to an individual in the control group given a positive opinion deviation of 1. The 95% confidence interval for this differential probability is calculated to be [3.1%, 4.6%] as indicated by the vertical bar corresponding to the opinion deviation of 1. Further, at higher levels of opinion deviation, for instance, given an opinion deviation of 1.5, on average, an individual in the treatment group is approximately 8% less likely to post an online rating of 5 than an individual in the control group. The 95% confidence interval for this differential probability is calculated to be [6.2%, 10.5%] as shown in Figure 2.4.1. The two social forces that are identified earlier work in the same direction and reduce the probability of posting an online rating of 5 for treated individuals whose satisfaction score is 10. These two forces are (1) the main negative effect of social influence that increases the likelihood of posting a lower rating for treated individuals, and (2) conformity that increases the likelihood of posting a lower rating for treated individuals in an attempt to give a score that is more in line with the average rating of reviews by previous reviewers. The second force, i.e. the conformity pressure, gets stronger as the opinion deviation increases. Therefore, we observe further differentiation in posting behavior between the two groups at higher levels of opinion deviation as evident in Figure 2.4.1.

Figures 2.4.2.1 and 2.4.2.2 display a pattern consistent with Figure 2.4.1. Figure 2.4.2.1 shows the probability difference in posting an online rating of 5 between the treatment and control groups conditional on a survey satisfaction score of 9, whereas Figure 2.4.2.2 shows the same difference in posting an online rating of 4. For instance, according to Figure 2.4.2.1, conditional on a survey satisfaction score of 9, on average, an individual in the treatment group is 26% less likely to post an online rating of 5 compared to an individual in the control group given a positive opinion deviation of 0.5. Because an individual with a survey satisfaction score of 9 predominantly posts either an online rating of 4 or 5, Figure 2.4.2.2 shows that, on average, an individual in the treatment group is 26% more likely to post an online rating of 4 compared to an individual in the control group given a positive opinion deviation of 0.5. In other words, Figure 2.4.2.2 is just the mirror image of Figure 2.4.2.1. It merely shows that as the probability of posting an online rating of 5 decreases for individuals in the treatment group due to social influence, the probability of posting an online rating of 4 increases.

Finally, two observations are worth noting. First, since the 95% confidence intervals reported in Figure 2.4 do not include zero (0), the behavioral difference in online posting between the treatment and control groups is deemed to be statistically significant at 5% significance level. Second, we argue that individuals whose satisfaction score is 9 are on the border of posting either a score of 4 or 5. A comparison of the magnitude of behavioral differences portrayed between Figures 2.4.1 and 2.4.2.1 suggests that the effect of social influence is exacerbated for borderline cases. Given an opinion deviation of 1, on average, a treated individual with a survey satisfaction score of 10 is approximately 4% less likely to post an online review rating of 5 compared to his counterpart in the control group, whereas a treated individual with a survey satisfaction score of 9 is 43% less likely to post an online review rating of 5 compared to his counterpart in the control group.

4.2. Robustness Checks

Our findings reported in Table 2.2 are consistent across three different matching specifications. In this section, we present results from additional tests that we performed to further establish the robustness of our findings. Models 1.1, 1.2 and 1.3 presented in Table 2.2 are estimated after performing a one-to-many CEM matching. The estimation results in Table 2.3 shows the same models estimated after performing a one-to-one matching under 1.1', 1.2' and 1.3' columns, respectively. As evident from Table 2.3, results are identical.

According to one survey, 45% of shoppers prefer to read the most recent reviews first and 74% of them want to read a minimum of between 2 and 7 customer reviews per product to have sufficient confidence in their product judgment (Powerreviews 2008). Therefore, the final robustness check that we performed is based on the assumption that reviewers may be paying more attention to the most recent reviews as well. In order to investigate whether recency effects exist and to demonstrate robustness of our findings, we carried out the same analyses using the most recent reviews displayed on the hotel's first review page in calculating opinion deviation as opposed to using all reviews.

For each hotel, the hotel group's website displays reviews by sorting them in the order of recency. The first webpage shows the most recent reviews, whereas the last webpage shows the oldest ones. Each webpage has a certain number of reviews that it holds. For instance, any given hotel's first review webpage lists only the most recent eight reviews. Therefore, we use the average of these eight[11] reviews (as opposed to all reviews) to construct the individual level opinion deviation measures. Table 2.3 displays coefficient estimates when opinion deviation is calculated using only the most recent eight reviews under 1.1'', 1.2'' and 1.3'' columns. Results are fairly similar and consistent across all three matching specifications, except for the second one. The second matching specification yields a significantly positive coefficient for the opinion deviation variable for individuals in the control group. We are not able to explain this odd result; however, the most crucial findings hold: (1) an online rating posted by an individual is primarily driven by his satisfaction level, (2) the main effect of social influence (treatment) is negative when there is no difference between his satisfaction score and the average rating of previous reviews, and (3) an individual who belongs to the treatment group adjusts his online review rating in order to conform to the majority opinion expressed previously if there exists a difference between his private satisfaction score and the average rating of reviews before his review.

[Insert Table 2.3 here]

In sum, results highlight two social forces at play in shaping a reviewer's online rating. First, individuals in the treatment group are more likely to post a lower online rating than their counterparts with an opinion deviation of zero and identical satisfaction scores in the control group. Second, an individual in the treatment group is inclined to post an online rating higher than her counterpart in the control group if her opinion deviation is negative, whereas she is inclined to post an online rating lower than her counterpart in the control group if her opinion deviation is positive. These two forces sometimes work in the same direction (e.g., when an individual's satisfaction score is 10/10 and the average rating of previous reviews is 4/5) and sometimes work in opposite

[11] Aside from being listed on the first review webpage, these 8 customer reviews would be sufficient for 74% of shoppers to have confidence in their product judgment (Powerreviews 2008).

directions (e.g., when an individual's satisfaction score is 7/10 and the average rating of previous reviews is 4/5). First, consider an individual with a satisfaction score of 10/10 writing an online review for a hotel whose average rating is 4/5. His opinion deviation is positive (5-4). Both the main treatment effect and the conformity effect reduces this individual's probability of posting an online review rating of 5/5 under treatment condition. Now, consider an individual with a satisfaction score of 7/10 under the same scenario. His opinion deviation is negative (3.5-4). Therefore, the main treatment effect and the conformity effect work in opposite directions. The treatment effect works in the direction of reducing his probability to post an online rating of 3 whereas the conformity effect works in the direction of increasing his probability to post an online rating of 3 or even 4.

5. DISCUSSION AND CONCLUSION

The main objective of this study is twofold: (1) to investigate whether social influence on online reviews exists, and (2) to understand how social influence shapes review posting behavior, provided that it exists. Research to date has shown that social dynamics influence an individual's decision of not only whether to post a review, but also what to post (Moe and Schweidel 2012). This study focuses on the latter dynamic and shows that (1) online opinions previously posted by others influence the rating that a reviewer gives even after accounting for his own private evaluation; (2) individuals who are subject to social influence tend to post ratings that are lower than those who are not subject to social influence; and (3) individuals exhibit conformity when adjusting their ratings taking into account the information provided in previously posted reviews. For instance, an individual, whose private evaluation is 5 out of 5, is more likely to post an online review rating of 5 out of 5 if the individual faces an average online rating of 4.8 rather than an average online rating of 4.2. And conversely, the same individual is more likely to post a rating of 4 when confronted with an average online rating of 4.2 than an average online rating of 4.8.

Online WOM plays a crucial role in consumer decision making, and its influence is projected to further intensify. This research deepens our understanding of how online reviews evolve due to social influence and contributes to research on social influence in online reviews in four ways. First, observing both individuals' private evaluations (through surveys) and public opinions (through reviews) is a key feature that allows us to assert that individuals' posted ratings are affected by others' opinions. Additionally, this key feature allows us to examine how individuals reconcile potential disagreements that may exist between their private evaluations and the average opinions of others. Extant research has ignored this issue focusing mainly on the impact of disagreements between existing reviews (i.e., variance of online ratings) on subsequent review ratings. Second, the random nature of how the company selects individuals for soliciting reviews allows us to create treatment and control groups that are essential in establishing causal inference with high external validity. Third, we document that conformity exists in online reviews. Finally, we use a matching technique to overcome selection concerns that may arise due to various observed factors and to establish the causal impact of social influence. One secondary contribution of the paper is that it illustrates the use of CEM for research in marketing.

From a managerial perspective, this study underscores the importance of maintaining a positive review environment. The significance of a positive review environment stems not only from its positive impact on product demand as documented in previous research (Chevalier and Mayzlin 2006; Wendy W Moe and Trusov 2011), but also from the social pressure that it exerts on ratings of subsequent reviewers, as demonstrated in this study. Nonetheless, our results indicate that review valence of products wind up decreasing due to social influence. In other words, our results suggest that the average rating of a product is lower when reviewers, who post an online review for it, are subject to social influence (i.e., presented with others' reviews) than when they are not.

For that reason, managers should consider eliciting online reviews in alternative ways which hinder reviewers' ability to read others' evaluations. Alternatively, managers may consider hiding online reviews of a product until the total number of reviews submitted for it reaches a certain quantity.

The idea that online reviews suffer from reporting biases is not new (Godes and Silva 2012; Li and Hitt 2008; Moe and Schweidel 2012). However, the extant literature attributes these biases to self-selection. In other words, they argue that an existing set of online reviews attracts or repels a certain type of subsequent review posting behavior. For instance, Moe and Schweidel (2012) report that positive environments increase posting incidence, whereas negative environments decrease it. That is, depending on the direction of the online opinions, some individuals are encouraged to contribute or discouraged from contributing to online conversations. Consequently, online reviews no longer represent the opinions of the general customer base. The current study reveals an additional source of reporting bias in that individuals' publicly expressed opinions may differ from their private evaluations due to the social influence from already existing online reviews. This finding further strengthens the argument that managers should approach online reviews with caution and with an understanding that online reviews may not represent the "unbiased" opinions of the overall customer base.

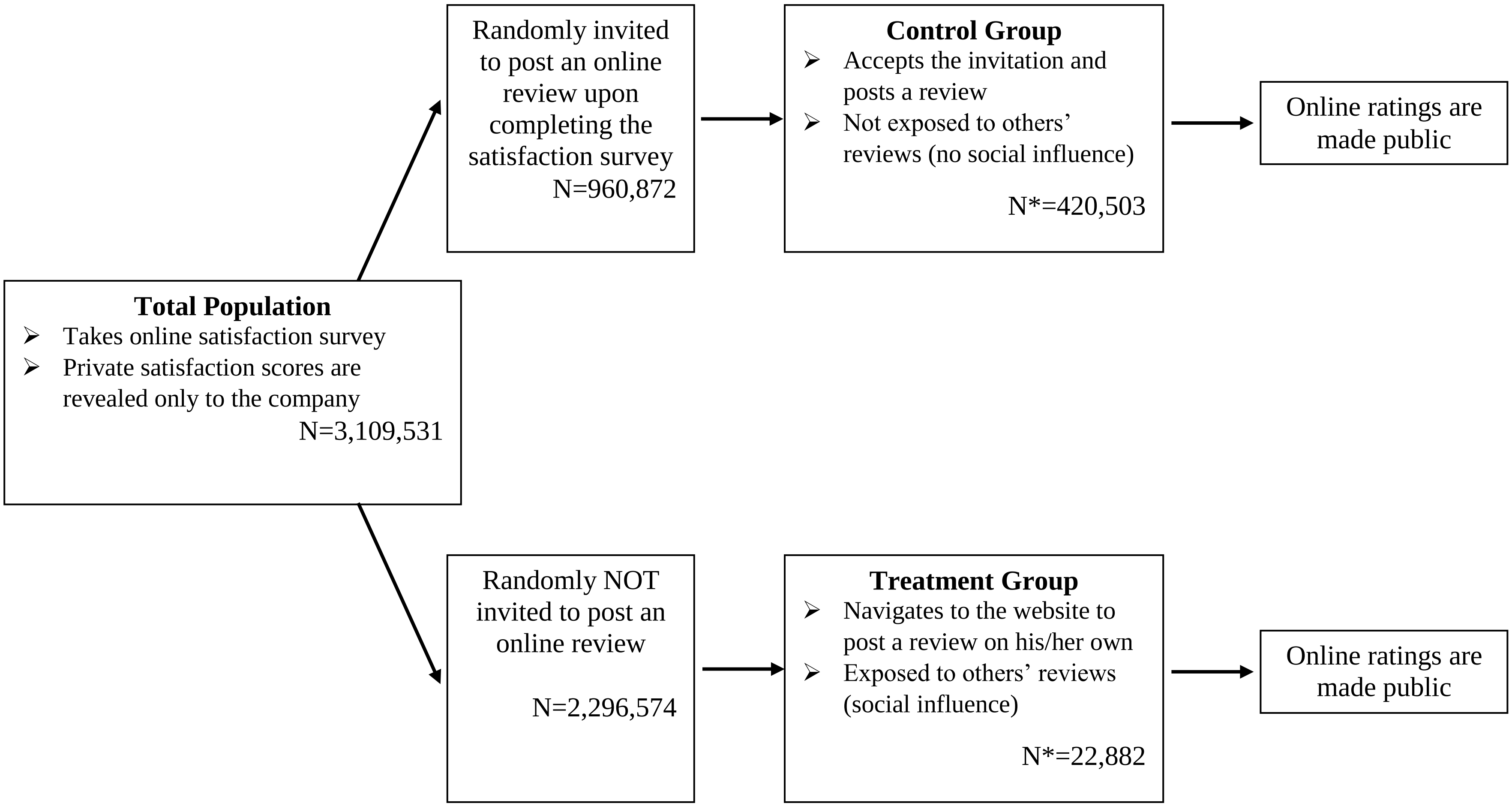
To date, companies, such as the one that provided the data for this research, continue to conduct surveys to track customer satisfaction. They do so in the hopes of monitoring "unbiased" opinions of their customer base. Customer satisfaction surveys are costly endeavors in terms of not only money, but also time. This study illustrates that it is possible to back out private evaluations of online reviewers. In order to do so, a firm still needs to conduct a small set of surveys and use them to estimate the effect of social influence on online reviews as demonstrated in this study. After estimating the effect of social influence, the firm can determine the “unbiased” satisfaction score that its online reviewers would have reported had they taken the survey.

Bourne (1957) introduces the concept of reference-group influence and argues that an individual’s behavior is influenced in varying degrees by other people and perhaps in different ways as well. For instance, some textual elements of reviews can induce stronger or weaker social influence. Consider two reviewers: one who compares his stay experience to his previous experiences within the same hotel chain and the other who compares her stay experience to her previous experiences with competing hotel chains. The first reviewer demonstrates that he has in depth experience with the same hotel chain, whereas the second reviewer demonstrates that she has a variety of experiences. Perhaps the strength of social influence induced by these two distinct reviewers and reviews differs. Identifying such characteristics from review text requires advanced text mining skills; however, identifying reference-group effects in the context of online reviews could prove fruitful for future research.

Finally, this study shows that individuals’ privately held opinions may differ from their publicly expressed evaluations. Analysis presented here suggests that these two evaluations will be positively correlated. However, it does not address which one of these two judgments (i.e., private and public) plays a more significant role in shaping consumer purchase behavior in subsequent periods. This issue remains open for future research.

TABLES AND FIGURES

Figure 2.1. Control and Treatment Groups



*Actual number of observations used in the analysis differs depending on the matching specification. Control and treatment groups are always matched on the satisfaction score they reported on the survey.

Figure 2.2. There is no bias in review invitation

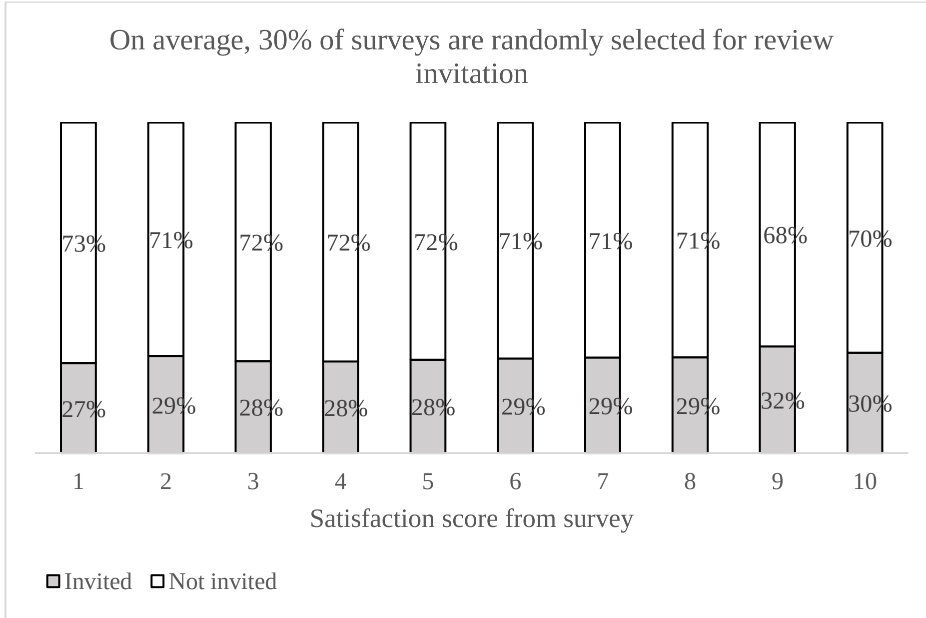
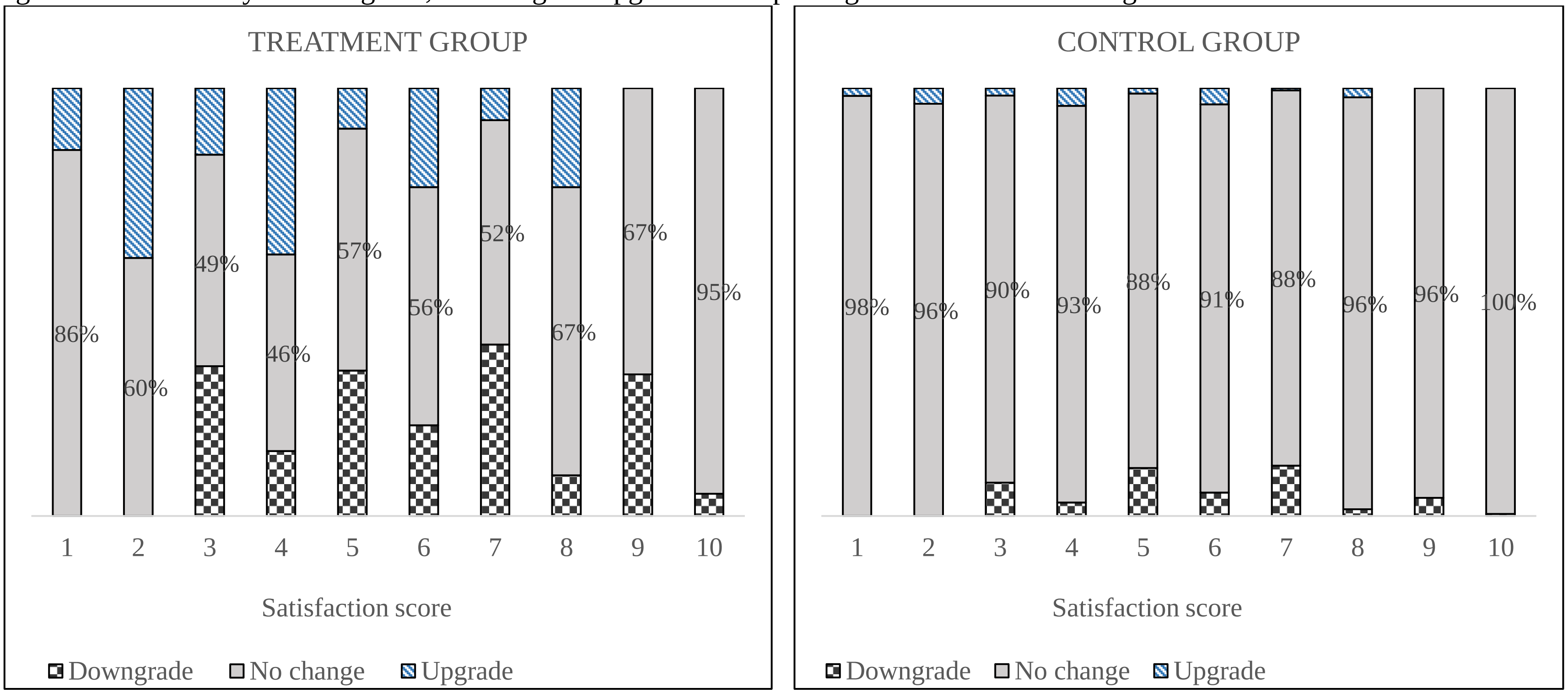


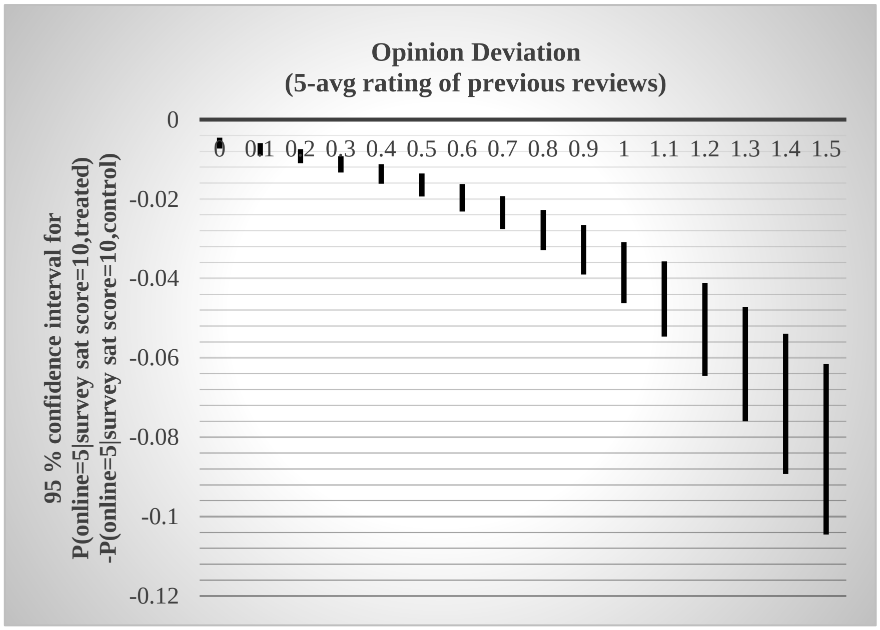
Figure 2.3. Probability of downgrade, no change or upgrade while posting an online review rating



Notes. Individuals in the treatment (social influence) group are more likely to post an online rating that differs from what their satisfaction scores imply. On the other hand, individuals in the control group post online ratings that are very similar to what their satisfaction scores imply. For instance, conditional on a satisfaction score of 7, 88% of control group individuals post an online rating of 4 (no change) whereas the same probability is 52% for treated individuals. There is a 33% chance that individuals in the treatment group will post an online rating of 3 (downgrade), whereas the same is observed only 11% of the time for individuals in the control condition.

Figure 2.4. Illustration of social influence (based on Model 1.3 in Table 2.2)

Figure 2.4.1. Probability of posting an online rating of 5 conditional on a satisfaction score of 10.



Notes. The x-axis denotes opinion deviation calculated as the difference between 5 (10/2) and the average rating of previous reviews. Conditional on a satisfaction score of 10, the observed range of opinion deviation is [0, 1.5]. Given that the average rating of all other reviews cannot exceed 5, opinion deviation cannot be less than 0. The y-axis denotes the 95% confidence interval for the probability difference between the treatment and control groups in posting an online review rating of 5. When opinion deviation is 0, the only social influence results from the negative main effect of treatment. Correspondingly, the figure indicates that individuals in the treatment group is slightly less likely (about 1%) to post an online rating of 5 compared to individuals in the control group. As opinion deviation increases, i.e., as the average rating of previous reviews decreases, the individuals in the treatment group becomes much less likely to post an online rating of 5 compared to individuals in the control group. For instance, when opinion deviation is 1, i.e., the average rating of previous reviews is 4, treated individuals are, on average, 4% less likely to post an online rating of 5 than control individuals. The 95% confidence interval for this differential probability is calculated to be [3.1%, 4.6%] as shown in the figure. Furthermore, when opinion deviation is 1.5, i.e., the average rating of previous reviews is 3.5, treated individuals are, on average, 8% less likely to post an online rating of 5 than control individuals.

Figure 2.4.2.1. Probability of posting an online rating of 5 conditional on a satisfaction score of 9.

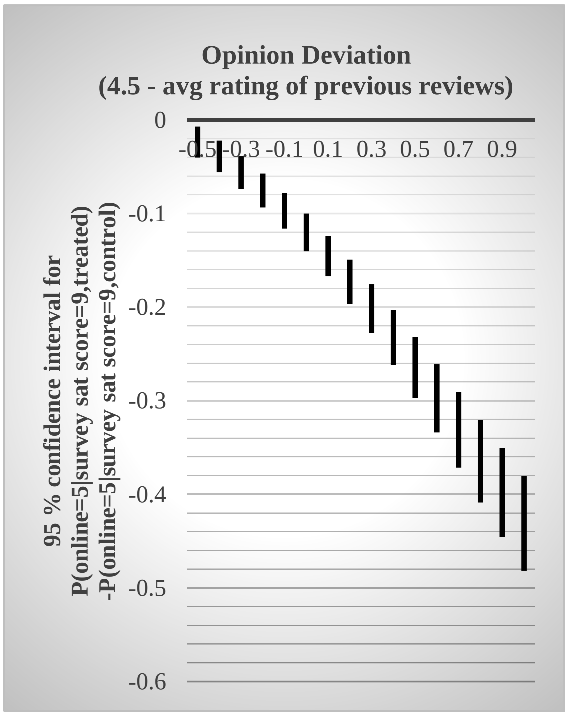
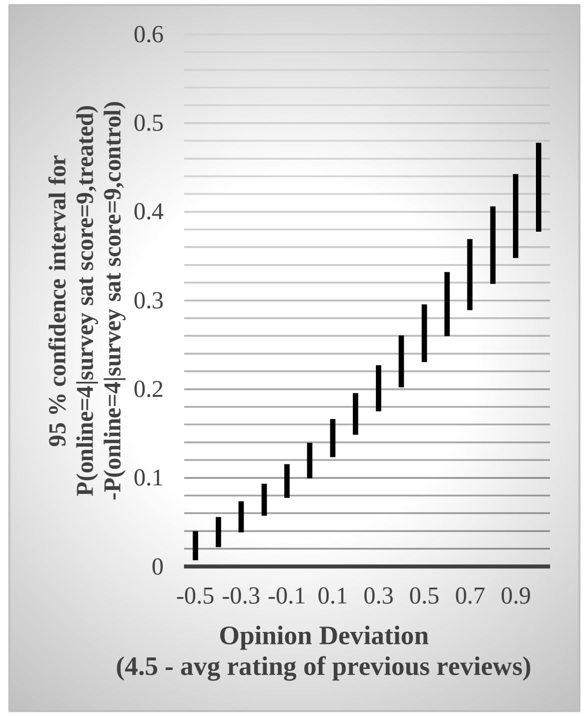


Figure 2.4.2.2. Probability of posting an online rating of 4 conditional on a satisfaction score of 9.



Notes. The x-axis denotes opinion deviation calculated as the difference between 4.5 (9/2) and the average rating of previous reviews. Conditional on a satisfaction score of 9, the observed range of opinion deviation is [-0.5, 1]. The y-axis denotes the 95% confidence interval for the probability difference between the treatment and control groups in posting an online review rating of 5 in Figure 2.4.2.1 and 4 in Figure 2.4.2.2. These two figures are mirror images of each other because as individuals become less likely to post an online review rating of 5 they become more likely to post an online review rating of 4 and vice versa. When opinion deviation is 0, i.e., the average online ratings of previous reviews is 4.5, individuals in the treatment group is, on average, 12% less likely to post an online review rating of 5 than individuals in the control group. This effect is a result of the negative main effect of treatment identified in this study. As opinion deviation decreases, i.e., as the average rating of previous reviews increase from 4.5, this differential probability decreases as individuals in the treatment group exhibit conformity. For instance, when opinion deviation is -0.3, i.e., the average rating of previous reviews is 4.8, individuals in the treatment group is, on average only 2% less likely to post an online review rating of 5 than individuals in the control group. On the other hand, as opinion deviation increases, i.e., as the average rating of previous reviews decreases, the individuals in the treatment group becomes much less likely to post an online rating of 5 compared to individuals in the control group. For instance, when opinion deviation is 1, i.e., the average rating of previous reviews is 3.5, treated individuals are, on average, 43% less likely to post an online rating of 5 than control individuals. The 95% confidence interval for this differential probability is calculated to be [38%, 48%] as shown in the figure.

Table 2.1. Dataset Characteristics Before and After Coarsened Exact Matching Procedure

	Matching 1 Pre-CEM	Matching 1 Post-CEM	Matching 2 Pre-CEM	Matching 2 Post-CEM	Matching 3 Pre-CEM	Matching 3 Post-CEM
L1 – multivariate distance	0.745	0	0.917	0.472	0.974	0.585
L – exact hotel	0.073	0	0.076	0		
L – check-in month	0.154	0	0.127	0	0.128	0
L – check-in year	0.607	0	0.624	0	0.615	0
L – satisfaction score from survey	0.080	0	0.079	0	0.082	0
L – membership level			0.032	0	0.030	0
L – online booking dummy variable			0.117	0	0.116	0
L – # of previous reviews written			0.017	0.031	0.017	0.036
L – # of cumulative bookings at current brand			0.056	0.023	0.055	0.029
L – hotel Chain					0.024	0
L – hotel Location					0.029	0
L – hotel State					0.069	0
L – hotel Country					0.076	0
L – hotel star level					0.040	0

*Pre-CEM imbalance measures of variables differ slightly across different matching specifications because observations that involve missing values of matching variables are excluded in each pre-CEM imbalance calculations.

*An imbalance of 0 implies that all variables are perfectly matched.

Table 2.2. Ordered logit models of posted online review ratings

	MATCHING SPECIFICATION 1			MATCHING SPECIFICATION 2			MATCHING SPECIFICATION 3		
Variables	**Model 1.1**	**Model 1.2**	**Model 1.3**	**Model 2.1**	**Model 2.2**	**Model 2.3**	**Model 3.1**	**Model 3.2**	**Model 3.3**
Treatment dummy	-0.75***	-0.76***	-0.84***	-0.90***	-0.91***	-0.90***	-1.05***	-1.05***	-1.08***
(1=treatment, 0=control)	(0.08)	(0.08)	(0.07)	(0.12)	(0.12)	(0.12)	(0.10)	(0.10)	(0.09)
Satisfaction score	3.11***	3.38***	3.41***	2.92***	3.12***	3.13***	3.59***	3.82***	3.85***
(measured on a scale from 1 to 10)	(0.08)	(0.10)	(0.10)	(0.12)	(0.15)	(0.15)	(0.09)	(0.11)	(0.11)
Opinion deviation		-0.60***	-0.24		-0.44*	0.17		-0.50***	-0.24
(Satisfaction score/2-online avg rating)		(0.12)	(0.13)		(0.20)	(0.25)		(0.12)	(0.13)
Opinion deviation*Treatment dummy			-1.23***			-1.26***			-1.37***
			(0.14)			(0.29)			(0.15)
# of observations	19,453	19,448	19,448	5,042	5,041	5,041	31,342	31,335	31,335
Log likelihood	-4,623	-4,604	-4,522	-1,103	-1,101	-1,082	-6,209	-6,190	-6,089
	List of matching variables								
	Exact hotel			Exact hotel			Hotel Chain		
	Check-in month			Check-in month			Hotel Location (airport, resort etc…)		
	Check-in year			Check-in year			Hotel State		
	Satisfaction score			Satisfaction score			Hotel Country		
				Membership level			Hotel star level		
				Online booking			Check-in month		
				# of previous reviews written			Check-in year		
				# of cumulative bookings at current brand			Satisfaction score		
							Membership level		
							Online booking dummy		
							# of previous reviews written		
							# of cumulative bookings at current brand		

Notes. Results: (1) online rating posted by an individual is primarily driven by his satisfaction level, (2) the main effect of social influence (treatment) is negative when there is no difference between an individual's satisfaction score and the average rating of previous reviews before his review, i.e., his opinion deviation is zero, and (3) an individual who belongs to the treatment group adjust his online review rating in order to conform to the majority opinion expressed previously if there exists a difference between his satisfaction score and the average rating of reviews before his review, i.e., his opinion deviation is either positive or negative.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 2.3. Robustness checks

Variables	MATCHING SPECIFICATION 1		MATCHING SPECIFICATION 2		MATCHING SPECIFICATION 3	
	One-to-one Matching 1.3’	**Using the most recent 8 reviews 1.3’’**	**One-to-one Matching 2.3’**	**Using the most recent 8 reviews 2.3’’**	**One-to-one Matching 3.3’**	**Using the most recent 8 reviews 3.3’’**
Treatment dummy (1=treatment, 0=control)	-0.67*** (0.06)	-0.94*** (0.08)	-0.87*** (0.11)	-0.99*** (0.12)	-0.69*** (0.07)	-1.17*** (0.10)
Satisfaction score (measured on a scale from 1 to 10)	2.71*** (0.09)	3.23*** (0.09)	2.85*** (0.15)	2.95*** (0.13)	2.69*** (0.09)	3.69*** (0.10)
Opinion deviation (Satisfaction score/2-online avg rating)	-0.13 (0.14)		0.30 (0.26)		-0.038 (0.14)	
Opinion deviation*Treatment dummy	-0.92*** (0.12)		-1.21*** (0.27)		-0.94*** (0.12)	
Opinion deviation_last8 (Satisfaction score/2-online avg rating_last8)		0.09 (0.09)		0.45** (0.16)		0.003 (0.10)
Opinion deviation_last8*Treatment dummy		-1.08*** (0.12)		-1.03*** (0.22)		-1.07*** (0.14)
# of observations	8,914	19,382	3,663	5,024	7,510	31,259

Notes. Models 1.3’, 2.3’ and 3.3’ are estimated after performing one-to-one matching. Models 1.3’’, 2.3’’ and 3.3’’ are estimated after performing one-to-many matching and using only the most recent 8 reviews displayed on the hotel’s website to calculate opinion deviation. These most recent 8 reviews are the only reviews that are displayed on the same page as the “write a review” link. The main findings of Models 1.3, 2.3 and 3.3 presented in Table 2.2 hold.
***p<0.001;**p<0.01;*p<0.05

CHAPTER III: Competition and the Impact of Online Reviews on Product Financial Performance: Evidence from the Hotel Industry

The Internet has had a profound impact on not only how consumers shop, but also how they gather information about products and services. User-generated content (UGC) now spans a significant swath of the Internet, and online consumer reviews constitute one of the most predominant types of UGC (Luca 2016a). Online review platforms, such as Yelp and TripAdvisor, are heavily visited by consumers. For instance, the number of unique monthly visitors at Yelp and TripAdvisor was 167 million and 350 million, respectively, during the first quarter of 2016 (Yelp 2016, TripAdvisor 2016). It is evident that many consumers seek information about other consumers' experiences and recommendations through online reviews. This article addresses the question of how online reviews shape consumer decision-making. A substantial body of research in marketing and information systems focuses on this issue. This article builds on previous research in three important respects: (1) it incorporates competitor reviews, (2) it highlights the significance of relative online review metrics, and (3) it demonstrates the impact of online consumer reviews on actual revenue data.

The majority of the published research to date focuses on the impact of a product's absolute online review characteristics on its demand (Archak, Ghose, and Ipeirotis 2011; Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008; Forman, Ghose, and Wiesenfeld 2008; Ghose, Ipeirotis, and Li 2012). Previous work often characterizes a product's online review environment by three distinct metrics; namely, valence, volume, and variance (King, Racherla, and Bush 2014). While these studies have significantly improved our understanding of how the online review characteristics of a product impact its demand, the current paper complements them by incorporating online review characteristics of competing products. In doing so, we respond to the call for additional research investigating the effect of online WOM about a competing product

(Babić Rosario et al. 2016). We study this effect in the context of the hotel industry. One of the major challenges associated with incorporating the impact of competitors' online reviews stems from the difficulty in identifying a product's competitor set. An important aspect of the unique data set used in this study is that it elicits the set of a hotel's local competitors from the general manager of the focal hotel. This is a real advantage because the way in which hotel specific competitor sets are constructed could potentially alter empirical results.

Second, this study considers the possibility that an individual's judgments may be relative. A growing consensus suggests that preferences are often constructed at the time purchasing decisions are made (Yoon and Simonson 2008), and behavioral research has repeatedly demonstrated that individuals are regularly influenced by elements in the environment in which preferences are constructed (Adomavicius et al. 2013). In an online review environment, the most readily available elements that could affect consumers' preferences for a product are characteristics of its own reviews as well as its competitors' reviews. We argue that competitors' review characteristics serve as a reference point against which a product's own review characteristics are compared, and consequently, relative review metrics, e.g., a product's own review valence minus its competitors' review valence, play a significant role in shaping preferences of consumers. Our decision to incorporate competitor review characteristics through relative metrics is consistent with several psychological theories, which advocate that perceptions and judgments are relative (Helson 1964; Huber, Payne, and Puto 1982; Kahneman and Tversky 1979; Sherif, Taub, and Hovland 1958; Simonson 1989; Simonson and Tversky 1992).

Third, previous research on online reviews (Archak, Ghose, and Ipeirotis 2011; Chevalier and Mayzlin 2006; Dhar and Chang 2009; Forman, Ghose, and Wiesenfeld 2008; Jabr and Zheng

2014) uses sales rank as a proxy for demand due to data availability constraints. In their meta-analysis, (Babić Rosario et al. 2016) report that 60.5% of studies operationalize product demand using sales rank. In this article, we measure a product's financial performance using actual total revenue data.

We specifically address the following research questions: (1) What is the impact of relative online review metrics on a product's financial performance? (2) Is the impact of relative online review metrics symmetric or asymmetric? and (3) Do these effects depend on competitor characteristics? To address these questions, we constructed a unique data set containing financial performance and online consumer review data. A multi-chain hotel group provided monthly financial information and competitor sets for hotels within its portfolio, and we scraped the online consumer review information for both the focal hotels and their competitors from TripAdvisor.com. The financial data set consists of monthly revenues and number of rooms available for each focal hotel from January of 2010 to September of 2015 for a total of 69 months. The online review data contain every TripAdvisor review written for all focal hotels and their competitors from the very first review written for each hotel to March of 2016. In addition to hotel specific online review histories, the data set contains TripAdvisor review histories of every reviewer in the data set. These review histories of reviewers are used to construct instruments for potentially endogenous independent variables of interest. We analyze 69 months of time series data using an instrumental variables approach and investigate the relationship between a hotel's financial performance and its relative online review metrics.

Our findings are unique in several ways. To our knowledge, this study is the first to demonstrate the significance of relative online review characteristics. From a managerial standpoint, this result highlights the importance of monitoring online reviews submitted not only for the focal product, but also for its competitors. Remarkably, many businesses continue to overlook online consumer reviews (CMO 2016). Moreover, we demonstrate that the impact of relative metrics is asymmetric. The magnitude of increase in a product's financial performance due to improvements in its online reviews depends on how its online reviews compare to those of its competitors. For instance, we show that increasing a product's average online rating while catching up to its competitor's average online rating is more beneficial than after having already surpassed its competitor's average online rating. Managers should be mindful of these asymmetric effects, especially when they are evaluating whether or not to invest in improving their product's online review metrics. In addition, we provide evidence that the strength of these effects depends on competitor characteristics.

We organize the rest of the article as follows: First, we provide a review of the research literature. Then, we provide a detailed description of the dataset and describe the empirical analysis. Finally, we present our results and conclude with a discussion of this study's contributions to our understanding of the impact of online reviews, its managerial implications, its limitations, and the opportunities for future research.

1. LITERATURE REVIEW

As summarized in Table 3.1, most of the research on online consumer reviews to date focuses on the impact of a product's own online reviews on its demand. Additionally, we refer the reader to (Babić Rosario et al. 2016) for a comprehensive meta-analysis on the effect of online WOM on sales. Following earlier work (Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008; Forman, Ghose, and Wiesenfeld 2008) investigating the impact of online reviews on demand, we primarily focus on two online review metrics: (1) review valence and (2) review volume. Our decision to limit ourselves to these two metrics is mainly driven by the desire to stay close to the empirical literature on online WOM. In their meta-analysis, (Babić Rosario et al. 2016) report that most researchers captured online WOM mainly by volume (used in 88% of the studies) and valence (used in 81% of the studies). Recent developments in text mining techniques have sparked interest in understanding the impact of the textual content of online consumer reviews (Archak, Ghose, and Ipeirotis 2011; Ghose, Ipeirotis, and Li 2012; Sonnier, McAlister, and Rutz 2011); however, textual content of online reviews is outside the scope of this study.

[Insert Table 3.1 about here]

In line with Chevalier and Mayzlin (2006) and Forman, Ghose, and Wiesenfeld (2008), we argue that consumers use online consumer reviews as an additional signal of underlying product quality. Findings of an event study suggest that published online ratings are indeed a valid source of information in assessing underlying product quality (Tellis and Johnson 2007). Although rising popularity is shown to have a negative effect on product evaluations for products with identity-signaling value, such as fashion items, music, or hairstyle (Berger and Heath 2007), positive effects of popularity on product evaluations are widely documented. Previous research on herding (e.g., Banerjee 1992; Salganik, Dodds, and Watts 2006) suggests that individuals may infer product quality based on product popularity, which could also be measured by the number of online reviews it receives. Thus, we argue that both the average online rating and the number of online reviews are used by consumers to infer a product's quality.

The increased availability of online reviews provides consumers an opportunity to easily compare and contrast reviews of multiple competing products at a time. Therefore, it is reasonable to expect not only a product's absolute online review characteristics, but also its competitor products' review characteristics to have an impact on its demand. Moreover, Gupta and Zeithaml (2006) pointed out the importance of accounting for competition more than a decade ago. They argued that ignoring competition yields not only an incomplete picture, but also a misleading one. For example, online average ratings of a product may be increasing over time with no impact on the given product's financial performance if its competitors' online average ratings are increasing at an even faster pace. The current study recognizes the role of competition and investigates the impact of online review metrics of a product on its demand in conjunction with how these review metrics may be changing for its competitors.

Individuals tend to evaluate alternatives relative to reference points (Hardie, Johnson, and Fader 1993; Hsee 1996; Hsee and Zhang 2010; Kahneman and Tversky 1979) because they do not have the ability to make absolute judgments (Prelec, Wernerfelt, and Zettelmeyer 1997). On a daily basis, consumers are constantly exposed to prices. Despite this fact, individuals tend to be influenced by external reference points even when evaluating prices. Research has shown that these reference prices have a consistent and significant impact on consumer demand (Kalyanaram and Winer 1995; Mazumdar, Raj, and Sinha 2005). Theoretically, reference points can come from anywhere. For example, external (contextual) reference prices can be constructed based on prices of competitive products in the store. Dickson and Sawyer (1990) showed that 93.1% of shoppers offered a response to the question about the relative price of the chosen brand (compared to other brands) whereas only 78.9% of shoppers could provide a response to the question about absolute prices. This finding is also consistent with the idea that consumers pay more attention to relative prices than absolute prices. Similarly, in an online review environment, the most readily available elements that could serve as reference points are competitors' online reviews. Consistent with these ideas, we operationalize online WOM characteristics of a product and its competitors by constructing relative online review metrics. For instance, relative valence is defined to be the difference between a product's own review valence and its competitors' review valence. This operationalization relies on the notion that competitor review characteristics serve as reference points (Kahneman and Tversky 1979).

Relative online WOM metrics used in this study can additionally be interpreted as an approximation of anticipated decision regret by consumers. Not only do consumers often experience postpurchase regret if they appear, after the fact, to have chosen the wrong option amongst possible alternatives, but also they often anticipate such feelings of regret at the time of choice. Further, Simonson (1992) demonstrated that the anticipated regret influences the choices made. Bell (1982) measures regret "as the difference in value between the assets actually received and the highest level of assets produced by other alternatives". Similarly, Inman, Dyer, and Jia (1997) captures regret as "the difference between the performance of the chosen product/service and the performance of a forgone product/service". Both of these measures of regret are similar to the relative online reviews metrics constructed in this study. Thus, one possible interpretation of relative online reviews metrics is that they approximate anticipated regret, which has been shown to impact consumers' purchase decisions (Bell 1982; Bolton, Kannan, and Bramlett 2000; Inman, Dyer, and Jia 1997; Simonson 1992). Given the nature of the data used in this study, the interpretation of the effects of relative online WOM is consistent with both the reference point and anticipated regret theories. Both of these theories suggest that the effects of online WOM are better captured through relative review metrics.

Researchers have long recognized that research on online consumer reviews can be enhanced by incorporating the impact of competition (Dellarocas, Zhang, and Awad 2007). However, as recently as last year, in their meta-analysis, (Babić Rosario et al. 2016) reiterated the lack of empirical studies investigating the effect of online WOM about a competing product. One exception is the study by Borah and Tellis (2016). They specifically focus on the effect of product recalls on online chatter of recalled products' competitors in the context of the automobile industry. They show that negative chatter resulting from a product recall of a focal brand may increase negative chatter about competing brands. Product recalls are prevalent in various industries, such as food, toys, and automobiles. In our context, a hotel rarely experiences a crisis as catastrophic as a product recall.

Two main reasons that contribute to the tendency of ignoring competition have been (1) the lack of online review data on competition, and (2) the difficulty in identifying a product's competitor set. The second reason has been more prohibitive than the first. In order to study competition amongst firms, researchers must identify who competes with whom (Porac et al. 1995). Mayzlin, Dover, and Chevalier (2014) make a notable attempt to overcome this second hurdle. They define two hotels as competitors if the distance between them is less than half a kilometer. Hence, one simple heuristic is to deem two hotels competitors if they are located within a fixed threshold distance of each other.

According to Porac and Thomas (1994), researchers define rivalry based on either of the two most commonly used criteria: (1) a technological criterion that categorizes firms to the extent that they have similar technologies and competencies, and (2) a market criterion that categorizes firms to the extent that they draw from the same resource pools (e.g., of customers and suppliers). The criterion used by Mayzlin, Dover, and Chevalier (2014) is more in line with the second criterion. Moreover, Porac et al. (1995) show that rivalry is socially constructed and that managers subjectively define the competitive environment around their firm through a cognitive process. As a result, there is often a difference between competitor sets constructed based on researcher imposed criterions and those identified by managers. Porac and Thomas (1990) prescribe that it is meaningful to describe competitive boundaries from an insider's subjective point of view rather than a researcher's objective or analytical point of view. In line with this prescription, the current study elicits the local competitor set of a hotel directly from its general manager and, therefore, relies on managers' own judgments about their local competitive environment. Next, we turn to describe the data in detail.

2. DATA

The data set used here consists of three separate parts: (1) hotel-level monthly financial performance information on all focal hotels operated by the hotel group, (2) entire TripAdvisor.com consumer review histories of all focal hotels and their competitors, and (3) entire TripAdvisor.com consumer review histories of all reviewers in the data set. First, the hotel-level financial performance data contain monthly revenues and number of rooms available for each focal hotel that operates as a franchisee of the multi-chain hotel group from January of 2010 to September of 2015 for a total of 69 months. The financial data encompass all transactions through all channels, i.e., both online and offline channels. Therefore, the relationships identified in this study are not restricted to the impact of online WOM on hotel performance from online channels. Even though using aggregate (i.e., sum of online and offline sales) product performance is a novel (only 12.5% of 96 studies examined by (Babić Rosario et al. 2016) used sum of online and offline sales) feature of the current study, it does introduce a potential for bias. First, online WOM may not be representative of offline WOM. Moreover, online WOM measured from the TripAdvisor website may not be representative of all online

WOM. However, these potential biases would, if anything, diminish the estimated relationship between online WOM and hotel performance in this study.

Second, TripAdvisor is by far the most popular online review platform for hotels (Luca 2016a), reaching 350 million average monthly unique visitors by the second quarter of 2016 (TripAdvisor 2016). For that reason, the online WOM data for this study were constructed from the TripAdvisor website and contain all consumer reviews written for all focal hotels and their competitors, from the very first review written for each hotel to March of 2016. Specifically, we collected the following information for each consumer review: reviewer ID, review ID, review date, and review rating. Third, in addition to hotel specific online consumer review histories, the online WOM data contain review histories of every reviewer in the data set. Particularly, we collected the following information for each consumer review written by each reviewer who wrote at least one review for any hotel (i.e., focal or competitor) in the data set: reviewer ID, review ID, review type, review date, and review rating. TripAdvisor is primarily known for hosting reviews for hotels (Luca 2016a); however, it also hosts reviews for other travel related categories, such as restaurants and attractions. Hence, the online WOM dataset contains all types of reviews written by each reviewer on the TripAdvisor website, from the very first review they wrote to March of 2016.

For each focal and competitor hotel, we constructed two online review metrics: (1) valence and (2) volume using review histories collected. On the TripAdvisor website, the review ratings range from 1 to 5. In line with previous work on online WOM, we measured review valence by the average rating of a hotel at the beginning of a given month, whereas we measured review volume by the number of reviews received up to the beginning of the given month. Additionally, by observing individual level purchase behaviors of hotel guests, we are able to rationalize using online review metrics at the beginning of a given month as independent variables. From January 2011 until May 2015, we observe 148M bookings across the entire portfolio of hotels of the multi-chain hotel group. On average, corporate travelers book their room 11 days ahead of their trip, whereas non-corporate travelers book 15 days ahead of their trip. As expected, on average, leisure travelers make reservations further in advance compared to corporate travelers. Further, 92% of corporate travelers book their trip within 30 days of their stay, whereas 86% of non-corporate travelers book their trip within 30 days of their stay. This observation substantiates addressing the lag effect of online reviews on revenues using online review metrics at the beginning of a given month as independent variables.

The data used for the analysis include 1,992 focal hotels constructed sometime between 1998 and 2011 across 7 different chains. All focal hotels are located in the US. For each hotel, observations pertaining to its first year of operations are excluded from the analysis in order to ensure that the analysis is carried out after all hotels in the data have reached their stabilization period.

Table 3.2 summarizes the breakdown of these brands and provides online review characteristics of each chain. For each hotel in the data, we extracted star rating information for the hotel and its competitors from the TripAdvisor website. The hotel class star ratings are provided to TripAdvisor from third-party partners, such as Expedia. Focal brand 1 is a luxury hotel chain with 4 star or above hotels, and it is comparable to brands such as Lowes, Ritz-Carlton, and Waldorf Astoria. Focal brands 2, 3, and 4 are upscale hotel chains with 3 or 3.5 star hotels, and they are comparable to brands such as Best Western Premier, DoubleTree, and Crowne Plaza. Focal brands 5 and 6 are upper midscale hotel chains with 2.5 or 3 star hotels, and they are comparable to brands such as Comfort Inn, Fairfield Inn, and Holiday Inn. Finally, focal brand 7 is a midscale hotel chain with 2 star hotels, and it is comparable to brands such as Quality Inn, Ramada, and Wingate by Wyndham. By the end of the data collection period in August 2015, across the portfolio of 1,992 hotels, the average rating was 4.06 out of 5 and the average number of reviews was 179.

[Insert Table 3.2 about here]

Competitor Sets and Classification

A novel feature of this study is that it elicits a focal hotel's competitor set from the general manager of the focal hotel. In doing so, we rely on the reasonable assumption that the general manager of a focal hotel correctly identifies local competitors of the focal hotel. Unlike previous researchers (Jabr and Zheng 2014; Luca and Zervas forthcoming; Mayzlin, Dover, and Chevalier 2014), we do not resort to any additional assumptions in determining a product's competitor set. We geocoded hotels using their addresses listed on TripAdvisor.com. The average distance between a hotel and its competitor is 3.36km.

Furthermore, hotels can be considered as vertically differentiated products that can very easily be classified into different quality tiers (e.g., Kalnins 2017). As mentioned earlier, we extracted star rating information for the hotel and its competitors from the TripAdvisor website. Therefore, each hotel in the data can be classified into different quality tiers independent of its online review characteristics. Overall, hotels with higher star ratings have superior amenities, such as improved room furniture, lobby, onsite restaurants, lounge facilities, meeting space, and spa, compared to hotels with lower star ratings. Therefore, using competing hotels' star ratings, we can categorize each of a focal hotel's competitors into groups of lower quality, same quality, or higher quality competitors. For instance, if a competitor has the same star rating as the focal hotel, then it is considered to be a same quality competitor. Similarly, if a competitor has a higher star rating than the focal hotel, then it is considered to be a higher quality competitor. Finally, if a competitor has a lower star rating than the focal hotel, then it is considered to be a lower quality competitor. Consider a 3.5 star Crowne Plaza hotel. In this study, a competing 4.5 star Lowes hotel is deemed to be a higher quality competitor irrespective of how the average online rating of the Crowne Plaza hotel compares to that of the Lowes hotel. Moreover, a competing 3.5 star DoubleTree hotel is considered to be a same quality competitor, whereas a competing 2.5 star Ramada hotel is regarded as a lower quality competitor.

[Insert Table 3.3 about here]

Table 3.3 summarizes the competitive landscape of all focal hotels. It shows that almost all (93%) of the focal hotels have a same quality competitor. Likewise, it is very likely (61%) for a focal hotel to have a lower quality competitor. Approximately half of all focal hotels have a higher quality competitor.

3. EMPIRICAL ANALYSIS

Variables

Dependent Variable. In this study, we focus on revenue per available room (RevPAR) to measure a hotel's financial performance. RevPAR is calculated by dividing the total revenue of a hotel by the number of available rooms in the same period. It is computed on a monthly basis at the individual hotel level. We use RevPAR as a dependent variable because it is one of the most prevalent performance metrics used in the hotel industry (Ismail, Dalbor, and Mills 2002). We take the natural logarithm of a hotel's RevPAR and use this transformed value as our dependent variable.

Independent Variables. The main variables of interest for the current study are relative review metrics. More specifically, we focus on relative valence and relative volume. Relative review metrics for a given hotel are calculated separately for each competitor quality level. For instance, relative valence of a focal hotel with respect to its same quality competitor, i.e., $Relative\_valence_{same,it}$, is calculated by subtracting the focal hotel i's average online rating at the beginning of month *t* from its same quality competitor's average online rating at the beginning of month *t*. If a particular focal hotel has more than one same quality competitor, we use the maximum average rating across average ratings of all same quality competing hotels. Natural logarithm of the cumulative number of reviews is used to measure a hotel's review volume. In similar fashion, relative volume of a focal hotel with respect to its same quality competitor, i.e., $Relative\_volume_{same,it}$, is calculated by subtracting the natural logarithm of the focal hotel *i*'s cumulative number of reviews at the beginning of month *t* from the natural logarithm of its same quality competitor's cumulative number of reviews at the beginning of month *t*. This specification is identical to taking the natural logarithm of the ratio of the focal hotel's cumulative number of reviews to its same quality competitor's cumulative number of reviews. Additionally, we construct separate measures of positive and negative relative metrics (e.g., $Relative\_valence_{same,pos,it}$ and $Relative\_valence_{same,neg,it}$) to allow for asymmetrical effects on hotel performance. Table 3.4 provides a summary of how these key independent variables of interest are calculated, and Table 3.5 presents descriptive summary statistics.

[Insert Tables 3.4 and 3.5 about here]

Based on Tables 3.2 and 3.5, it is worth noting that even though both a hotel's star rating and its review valence are indicative of a hotel's quality, these two measures are not identical to one another. In other words, hotels with higher star ratings are not necessarily attaining greater review valence than hotels with lower star ratings. Table 3.5 shows that there are focal hotels whose review valence is greater than that of their higher quality competitors and whose review valence is smaller than that of their lower quality competitors. One explanation for this is that online reviews of a hotel take into account consumer expectations given the star rating of the hotel. In other words, before going into a 3 star hotel, consumers form expectations, and if their expectations are exceeded, they do not refrain from posting a review rating of 4 or higher.

Importance of Accounting for Competition

Consider two hotels depicted in Figure 3.1. Hotel A, presented on the left, has been experiencing a constant decline in its review valence, whereas its same quality competitor's review valence has been increasing over the 69-month period. Suppose that we observe a decline in Hotel A's RevPAR over the same time period and that we consider the impact of its own review valence, ignoring its competitor's reviews. We may attribute the entire decline in Hotel A's RevPAR to its decreasing review valence. However, the decline in Hotel A's RevPAR is presumably driven by a combination of two factors: (1) the decrease in Hotel A's review valence, and (2) the increase in its competitor's review valence. On the other hand, Hotel B, presented on the right, has been experiencing a constant increase in its review valence, whereas its same quality competitor's review valence has been increasing, as well. Assume that Hotel B's RevPAR has been constant over the same period. If we ignore its competitor's reviews and only focus on Hotel B's own review valence, we may conclude that online reviews do not impact a hotel's RevPAR because Hotel B's own review valence has been increasing over time, but its RevPAR has been constant. Nevertheless, Hotel B's RevPAR has been constant even though its review valence has been increasing probably because its competitor's review valence has been increasing at the same rate. Both of these examples highlight the importance of accounting for online reviews of competitors in order to uncover the true impact of online reviews on product performance.

[Insert Figure 3.1 about here]

Empirical Strategy

In this section, we discuss potential econometric issues that need to be addressed in this study. At the outset, we begin with testing for the presence of unit roots in time series of the dependent variable. In order to do so, we use a Fisher-type test that combines the *p*-values of augmented Dickey-Fuller (Dickey and Fuller 1979) tests, which are performed at each hotel's series separately, into an overall test for the entire panel series (test statistic= 226, *p*-value<0.0000), and therefore, the presence of unit roots is overruled.

Next, we test for serial correlation because generally, the error terms of individual units in a panel structure are serially correlated. In order to do so, we use the Woolridge test to detect whether the series followed a first-order autoregressive process AR(1). The test statistic calculated as $F_{(1,1991)}=3{,}978$ (at *p*-value<0.0001) implied that the null hypothesis of no first order autocorrelation is rejected for the dependent variable. Additionally, a modified Wald test performed to test the existence of heteroskedasticity rejected the null hypothesis of homoskedasticity (at *p*-value<0.0000). In order to account for both serial correlation and heteroskedasticity, we use heteroskedasticity- and autocorrelation-consistent (Newey and West 1987) estimators for the IV estimation.

Further, in order to control for hotel-level unobserved heterogeneity, we use hotel fixed effects. The random effects model assumes that the unobserved (time-invariant) hotel fixed effects are uncorrelated with the included independent variables, such as relative metrics, in the model. This assumption is not realistic. Therefore, we resort to a fixed-effects model that does not rely on this restrictive assumption. Additionally, the Hausman specification test (statistic=89, *p*-value=0.0024) performed rejects the null hypothesis that the difference in coefficient estimates of FE and RE models is not systematic and provides additional justification for the fixed effects specification. In addition to hotel fixed effects, time fixed effects for each month are incorporated in the model. Formally, the baseline model can be expressed as follows:

$$\log(RevPAR_{it}) = \alpha + \beta * \mathcal{R} + \alpha_i + \alpha_t + \mu_{it} \quad (1)$$

where <*i*,*t*> indexes a hotel-month combination and $\mathcal{R}$ represents a vector of relative review metrics calculated as of the beginning of month *t*. Correlations among these relative review measures are reported in Table 3.6 and do not suggest that multicollinearity is an issue.

Finally, using fixed effects lessens endogeneity concerns because any relationship between revenues and relative review metrics will be identified from changes over time in revenues and relative review metrics, weakening the possibility that our findings reflect differences in average quality across hotels. Still, we take further measures to alleviate endogeneity concerns, which we discuss next.

Endogeneity Concerns

It is well known that consumer reviews may be simultaneously determined with product revenue. Both review valence and review volume metrics are subject to this endogeneity problem. For instance, even though review volume may increase future demand for a product, it can also be an outcome of previous and current demand (Godes and Mayzlin 2004). Further, while higher review valence may result in increased revenue for a product, it could also be determined by some unobserved factors that influence product revenue, such as unobserved product quality, advertising, or positive critical acclaim (Archak, Ghose, and Ipeirotis 2011; Godes and Mayzlin 2004; Zhu and Zhang 2010).

As we mentioned earlier, the inclusion of hotel fixed effects in our analysis enable us to account for time-invariant hotel quality, further decreasing the possibility that our results reflect differences in average hotel quality rather than relative metrics themselves (Forman, Ghose, and Wiesenfeld 2008). However, to further alleviate endogeneity concerns, we use the method of instrumental variables (IV), a common approach to correct for the endogeneity bias. In our context, instrumental variables must be correlated with a hotel's relative review metrics, but must not be correlated with hotel revenue. We explain how these instrumental variables are constructed next.

Construction of Instrumental Variables

Using lagged endogenous variables as instruments is a common approach in dealing with endogeneity. However, IV methods are not recommended if the only instruments available are lagged endogeneous variables (Rossi 2014). In this study, using review histories of reviewers, we construct two different sets of instruments: one for relative review valence and the other for relative review volume. We use individuals' review histories until the time they post a review on the focal hotel to construct IVs. The intuition for using review histories of reviewers is similar to Jabr and Zheng (2014): a review is a reflection of a reviewer's attitude toward a hotel as well as his or her idiosyncratic tendency to post or to rate. An individual's review history reveals a reviewer's individual tendency to write a review and to rate a hotel. Such individual tendencies are not driven by focal hotel's characteristics, and therefore, they are exogenous. Consider the review valence of a focal hotel at the beginning of month *t* and the average rating of all reviews written, including those for *other* product categories on TripAdvisor until month *t* by the same set of reviewers. These two average ratings will be correlated because even though these reviews are written for different products, they are written by the same set of reviewers whose idiosyncratic tendency to rate in a certain way (either because in general they tend to be harsh/lenient reviewers or because they tend to use comparable criteria to rate different products) is the same. However, it is hard to imagine the average rating of all reviews written for *other* products by the same set of reviewers to be driven by characteristics of the focal hotel.

More specifically, the IV for relative valence of the focal hotel *i* at month *t* for the same quality competitor is constructed by first identifying all reviewers for the focal hotel and its same quality competitor until month *t*, and then identifying all other product reviews posted by these same reviewers until month *t-1*. Then, we calculate two distinct averages: (1) the average of all *other* product reviews posted by all reviewers who posted a review for the focal hotel until month *t-1* (labeled as $valence\_other_{it}$) and (2) the average of all *other* product reviews posted by all reviewers who posted a review for the same quality competitor hotel until month *t-1* (labeled as $valence\_other_{same,it}$). Finally, the IV for $Relative\_valence_{same,it}$ is calculated by subtracting $valence\_other_{same,it}$ from $valence\_other_{it}$. The same procedure is followed for lower and higher quality competitors.

The IV for $Relative\_volume_{same,it}$ is constructed following the same logic. This time, instead of calculating the average rating of all *other* product reviews, we computed the total number of *other* product reviews and took the natural logarithm of it to be consistent with the relative volume measure. The IVs created in this manner are closely related to the relative valence and relative volume of the focal hotel, whereas they are not related to the focal hotel's RevPAR. The correlation between $Relative\_volume_{same,it}$ and its IV is 0.59, whereas the correlation between $Relative\_valence_{same,it}$ and its IV is 0.16. In order to avoid weak identification, we supplement these IVs with $Relative\_valence_{same,it-4}$. We believe that this is reasonable given that only a negligible portion of individuals book their stay 5 months or more ahead of time. Both underidentification and weak identification tests are rejected confirming that the IVs are strong. Finally, a test of overidentification is carried out. The test statistic, $\chi^2_{(2)} = 1.3$ with *p*-value=0.52 fails to reject the null hypothesis and confirms that the instruments are valid.

Estimation

We use the generalized method of moments (GMM) approach adopted by Chintagunta, Gopinath, and Venkataraman (2010) and Jabr and Zheng (2014). The GMM procedure not only allows us to address the endogeneity of relative metrics through the use of IVs, but also enables us to obtain heteroskedasticity- and autocorrelation-consistent coefficient estimates (Baum, Schaffer, and Stillman 2007; Chintagunta, Gopinath, and Venkataraman 2010). The estimation procedure consists of following steps: (1) estimate the regression model using standard IV methods, (2) use residuals of the regression in the first step to obtain the optimal GMM weighting matrix, and (3) allow for heteroskedasticity and correlation between error terms. We refer the reader to Chintagunta, Gopinath, and Venkataraman (2010) for a detailed description of the estimation procedure. The GMM estimator and its asymptotic variance are

$$\hat{\beta}_{GMM} = [(X'Z)W(Z'X)]^{-1}(X'Z)W(Z'y),$$

$$V(\hat{\beta}_{GMM}) = [(X'Z)W(Z'X)]^{-1},$$

Where X is the matrix of regressors in Equation 1, Z is the matrix of instruments, W is the optimal weighting matrix, and y is the dependent variable, i.e., $\log(RevPAR_{it})$.

4. RESULTS

Before we move onto our main results pertaining to the impact of relative reviews, we start with describing results in relation to the impact of absolute review metrics. Consistent with previous work, our findings reported in the first and fourth columns of Table 3.7 suggest that both review valence and review volume improve product revenue.

The Impact of Relative Reviews with Respect to Same Quality Competitor

All models in Table 3.7 include both individual hotel and time fixed effects even though we do not report these fixed effects due to the large number of estimates. Comparing the coefficient estimates between the "No endogeneity correction" and the "GMM Instrumental variables correction heteroskedasticity+autocorrelation" columns in Table 3.7 indicates that after accounting for endogeneity, parameter estimates, for the most part, do not change dramatically. As explained earlier, once hotel fixed effects are introduced, any relationship between revenues and relative review metrics will be identified from changes over time in revenues and relative review metrics, weakening the possibility that the findings reflect differences in average quality across hotels. Rossi (2014) argues that a simple product-specific fixed effects would be sufficient to remove endogeneity problem and there is no need to use instruments as long as the unobserved product characteristics do not vary across time. Our estimation results suggest that using hotel fixed effects is indeed sufficient for our analysis because the coefficient estimates do not differ drastically between results from standard regression methods and IV methods.

Further, adding absolute review metrics as independent variables in the model does not significantly alter our findings, as can be seen in the third and sixth columns, and therefore, the following discussion is based on the results presented in the fifth column in Table 3.7.

[Insert Table 3.7 about here]

Impact of Relative Valence

The results presented in Table 3.7 highlight the significance of a hotel's relative valence with respect to its same quality competitor in determining its financial performance. Increasing a hotel's relative review valence with respect to its same quality competitor has a positive impact on its RevPAR. However, the magnitude of this impact changes depending on whether the hotel's relative review valence is positive or negative. For instance, increasing a hotel's review valence is most rewarding when its review valence is lower than that of its same quality competitor. Once a hotel achieves a high level of review valence that is above the review valence of its same quality competitor, then further increasing its valence is not as beneficial. In fact, such an increase is found to have no statistically significant impact on the hotel's financial performance. The absolute value of coefficient estimates of $Relative\_valence_{same,pos}$ is smaller than and statistically different from (p=0.008) that of $Relative\_valence_{same,neg}$. These results support the notion that consumers weigh negative relative valence more than positive relative valence in evaluating alternative hotels.

More specifically, if a hotel's review valence is below that of its same quality competitor, then increasing its relative valence with respect to its same quality competitor by 0.1 or 0.05, on average, results in an increase of RevPAR in the amount of $0.26 or $0.13, respectively.

Impact of Relative Volume

Table 3.7 suggests that a hotel's financial performance is significantly impacted by its relative volume with respect to its same quality competitor. The asymmetric effect identified for relative valence is replicated for relative volume, suggesting that increasing a hotel's review volume is most beneficial when its review volume is smaller than that of its same quality competitor. Once a hotel's number of reviews surpasses that of its same quality competitor, then further increasing its relative volume with respect to its same quality competitor is found to have no statistically significant impact on the hotel's financial performance. Similarly, these results confirm the notion that consumers, on average, weigh negative relative volume more than positive relative volume in evaluating alternative hotels. The absolute value of coefficient estimates of $Relative\_volume_{same,pos}$ is smaller than that of $Relative\_volume_{same,neg}$ even though they are not statistically different from each other (p=0.36).

By August 2015, the end of our data collection period, the average number of reviews per hotel across 1,992 hotels was approximately 170 online reviews. Holding everything else constant, a five percent increase in relative volume of a hotel (which equates to 8.5 additional online reviews), on average, results in an increase of RevPAR in the amount of $0.18 when its relative volume with respect to its same quality competitor is negative. If the particular hotel's number of reviews is already greater than that of its same quality competitor, then this effect no longer holds. Next, we describe results from an extended model that incorporates the relative reviews of other competitors, including lower and higher quality competitors.

Extended Model: the Impact of Relative Reviews with Respect to Lower, Same and Higher Quality Competitors

The parameter estimates of the extended model are reported in Table 3.8. Coefficient estimates of the same quality competitor's relative review metrics are very similar to those presented in Table 3.7 and described earlier. Similarly, adding absolute review metrics as independent variables in the model does not significantly alter our findings, as can be seen in the third and fourth columns in Table 3.8.

[Insert Table 3.8 about here]

Impact of relative valence. Relative valence with respect to lower quality competitor does not seem to have a statistically significant impact, whereas the asymmetric impact of positive and negative relative valence holds for the higher quality competitor; i.e., the coefficient estimate of negative relative valence is much greater than that of positive relative valence for the higher quality competitor. Further, the former effect is statistically significant, whereas the latter is not.

More specifically, if a hotel's review valence is below that of its higher quality competitor, then increasing its relative valence with respect to its higher quality competitor by 0.1 or 0.05, on average, results in an increase of RevPAR in the amount of $0.18 or $0.09, respectively. If the particular hotel's review valence is already greater than that of its higher quality competitor, then this effect no longer holds.

Impact of relative volume. Our findings indicate that relative volume of higher quality competitors does not play a significant role in shaping a hotel's financial performance. However, a five percent increase in relative volume of a hotel in our data, on average, results in an increase of RevPAR in the amount of $0.16 when its relative volume with respect to its lower quality competitor is positive. If the same hotel's number of reviews is smaller than that of its lower quality competitor, then this effect no longer holds.

Robustness Checks

The fact that our coefficient estimates are robust to inclusion of absolute review metrics significantly increases our confidence in the reliability of our findings. In this section, we further investigate the robustness of our results. An alternative specification that can be used in constructing relative review metrics is using the *average* valence or volume across all competing hotels within corresponding quality levels. This specification implies that the reference valence or volume is the *average* valence or volume as opposed to the *maximum* valence or volume amongst competing hotels. Table 3.9 presents results from this alternative specification. First, we observe that the R-squared of this alternative specification (0.6856) is lower than that of the preceding specification (0.6858). This observation suggests that the reference is more likely to be the *maximum* valence or volume amongst competing hotels. Second, results from either of the specifications are very similar, giving us additional assurance that our results are robust.

An alternative specification where an interaction term between review valence and review volume is also considered. This specification relies on the idea that the higher the number of reviews and the average review rating are, the higher is the impact on hotel performance. Consistent with Chintagunta, Gopinath, and Venkataraman (2010), we found that the interaction term is not statistically significant.

Moreover, 1,575 of 1,992 focal hotels carry either the focal brand name 5 or the focal brand name 6. These two brands are perceived to be very similar by hotel guests. Estimating the same models using observations from these 1,575 hotels yields qualitatively similar results.

[Insert Table 3.9 about here]

5. DISCUSSION

In a recent meta-analytic study, (Babić Rosario et al. 2016) call for research investigating the impact of online WOM about a competing product. We respond to their call in this article. We argue that consumers use online consumer reviews as an additional signal of underlying product quality, and advocate that consumers rely on relative reviews in their decision making. Therefore, an improvement in a product's online relative review metrics is expected to result in increased consumer preferences towards the product. Our findings are in line with this expectation. Coefficient estimates suggest that improving a hotel's relative valence or volume with respect to its competitors may significantly improve its financial performance measured by revenue. However, trying to increase a hotel's relative valence or volume can be less productive under certain conditions and may not be justified given the cost of doing so.

More specifically, online consumer reviews seem to be a credible signal of superiority amongst hotels with the same star rating, i.e., same quality hotels. However, our findings reveal that consumers still weigh negative relative valence and volume more than positive relative valence and volume in evaluating alternative hotels with the same star rating. Consider a 3.5 star Crowne Plaza hotel, which competes against a 3.5 star DoubleTree hotel. Increasing the relative valence with respect to its same quality competitor, DoubleTree hotel, is more valuable if the review valence of the Crowne Plaza hotel is lower than the review valence of the DoubleTree hotel. Similarly, increasing the relative volume is more worthwhile if the review volume of the Crowne Plaza hotel is lower than that of the DoubleTree hotel.

We did not find a statistically significant impact of relative valence with respect to lower quality competitors. However, catching up to the average online rating of higher quality competitors is shown to have a positive impact on revenue. This finding is consistent with the notion that consumers aspire to attain higher quality products. The closer a hotel's average rating is to its higher quality competitor, the higher its perceived quality. And consequently, it is able to achieve higher revenues.

Previous work on pricing literature consistently documents that price promotions induce asymmetric switching patterns between lower- and higher-tier brands. When a high quality brand promotes, it is able to steal market share from its low quality competitor, whereas when a low quality brand promotes, there is very little switching down by consumers of high quality brand (Blattberg and Wisniewski 1989). Our article offers some insights into how these patterns may change if a lower quality competitor has a credible signal that suggests that it is of higher quality, i.e., a lower quality competitor enhances its perceived quality. Our findings advocate that having such a credible signal of product quality is beneficial if indeed the signal confirms better quality. The closer the average online rating of a 3.5 star Crowne Plaza to its 4.5 star competitor, Lowes, the higher its revenue. Probably, the increase in revenue is resulting from consumers switching down from a higher quality competitor. Even though price promotions by low quality brands induce little switching down by consumers of high quality brands, our result suggests that an increase in quality perceptions of low quality brands can cause switching down by consumers of high quality brands.

Moreover, once a hotel achieves a high level of review valence that is above the review valence of its higher quality competitors, then holding everything else constant, further increasing its relative valence with respect to its higher quality competitors is found to be statistically not beneficial. A hotel's review valence is not anticipated to exceed that of its higher quality competitor. However, it is theoretically possible for a hotel to beat review valence of its higher quality competitor. For example, if the review valence of a 3.5 star Crowne Plaza hotel is already higher than that of a 4.5 star Lowes hotel, our findings indicate that further increasing its relative valence with respect to the Lowes may not credibly alter consumers' quality perceptions of the Crowne Plaza formed from both hotels' star ratings (3.5 vs 4.5), and hence, may not be a worthy action to take. In sum, outstanding online consumer reviews, evidenced by a hotel's online average rating that is very close to its higher quality competitors, do provide a credible signal to enhance a hotel's quality perceptions. However, if the hotel's review valence exceeds that of its higher quality competitors, then our results indicate that it is much less worthwhile for a hotel to increase its review valence beyond that level.

We did not find a statistically significant impact of relative volume with respect to higher quality competitors. However, surpassing the number of reviews of lower quality competitors is shown to have a positive impact on revenue when a hotel's number of reviews is already more than its lower quality competitors. This finding diverges from the notion that consumers weight negative relative volume more than positive relative volume with respect to lower quality competitors. One possible explanation is that a hotel's review volume is not anticipated to exceed that of its lower quality competitor. However, if it does, then this popularity measure could enhance a hotel's quality perceptions and result in increased revenue.

Interestingly, our findings suggest that consumers may use different attributes in forming reference points while evaluating products. We found evidence that when evaluating a product with respect to its lower quality alternative, consumers form a reference point based on online review volume, whereas when evaluating a product with respect to its higher quality alternative, consumers form a reference point based on online review valence. Reference groups seem to change depending on the product attribute. The reference point literature mainly focuses on construction of a reference point for a single attribute (e.g., price) and does not consider reference points with two attributes or the possibility that reference group may change depending on the product attribute in question. Future experimental work could further explore this idea.

6. CONCLUSION

In conclusion, our objective here was to extend current understanding of how online WOM shapes consumer decision making by incorporating competitor online WOM information. It is not possible to establish true causality of the effects mentioned here based on this study. This is not necessarily a shortcoming specific to our analysis; it is just the reality of using secondary data. In order to establish causality, controlled experiments should be conducted. However, we can confidently say that our results provide strong support for the idea that relative online reviews matter and their effects on product revenue are asymmetric.

Assessing the quality of experience goods before purchase is deemed to be a more challenging task than assessing the quality of search goods before purchase (Murray and Schlacter 1990; Zeithaml 1981). Zeithaml (1981) argues that consumers may rely to a greater extent on word-of-mouth prior to purchase of an experience good (e.g., hotel stays) because they may perceive a greater risk associated with selecting an alternative. For that reason, online WOM (either relative or absolute) may play a more significant role for experience goods in terms of shaping consumer purchase decisions. Future research could investigate the generalizability of our findings across search goods and other types of experience goods.

Recently, Minnema et al. (2016) and Sahoo, Dellarocas, and Srinivasan (2016) emphasize the importance of accounting for product returns in evaluating the financial impact of online reviews. Their research show that while online reviews increase product sales, they can also lead to increased product returns. Future research investigating the impact of online reviews on sales or revenue should be mindful of this consideration. In our context, the issue of product returns does not exist. However, studying the impact of online reviews on repeat purchase behavior or customer attrition can be a fruitful area for future research.

Table 3.1. Previous Empirical Research Related to Online Consumer Reviews

Research paper	Product category	Methodology	Competitor reviews considered	Key findings
Godes and Mayzlin (2004)	TV shows	Regression	NO	Number of blog posts and entropy of blog posts are associated with TV ratings.
Liu (2006)	Movies	Regression	NO	WOM has significant explanatory power for box office revenue. This explanatory power comes from its volume but not its valence.
Chevalier and Mayzlin (2006)	Books	Differences-in-differences	NO	Average review ratings and number of reviews improve book sales. The impact of a negative review is greater than that of a positive review.
Dellarocas, Zhang, and Awad (2007)	Movies	Diffusion model	NO	Both average review ratings and number of reviews increase accuracy of motion picture forecasting models.
Duan, Gu, and Whinston (2008)	Movies	Simultaneous equations	NO	Higher average ratings do not result in higher sales, but the number of reviews is significantly associated with movie sales.
Forman, Ghose, and Wiesenfeld (2008)	Books	Regression	NO	The prevalence of reviewer disclosure of identity-descriptive information is positively related to sales whereas the average review rating is not a significant predictor of sales.
Dhar and Chang (2009)	Music	Regression	NO	The volume of blog posts predicts future sales. The evidence on average review ratings and the volume of reviews is mixed, with volume being more significant in more cases.
Chintagunta, Gopinath, and Venkataraman (2010)	Movies	GMM with IV	NO	Average user rating is the main driver of box office performance whereas the volume of reviews is not.
Zhu and Zhang (2010)	Video games	Differences-in-differences	NO	Valence, variance and volume of online reviews are more important for less popular and online video games.
(Wendy W. Moe and Trusov 2011)	Bath and beauty	Hazard model	NO	Average review rating has (i) a direct effect on immediate sales and (ii) an indirect effect on futures sales by means of its impact on future ratings.
Sun (2011)	Books	Differences-in-differences	NO	The standard deviation of online ratings is positively associated with sales if and only if the average rating is low.
Archak, Ghose, and Ipeirotis (2011)	Cameras and camcorders	GMM with IV	NO	Consumer choices are influenced by the textual contents of the reviews. The average review rating, the volume of reviews, and the standard deviation of review ratings have a positive impact on sales.
Sonnier, McAlister, and Rutz (2011)	Undisclosed	LIV approach	NO	Sentiment has an effect on sales; positive and neutral comments increase, whereas negative comments decrease.
Ghose, Ipeirotis, and Li (2012)	Hotels	GMM	NO	The textual content and style of reviews are significantly associated with demand.
Jabr and Zheng (2014)	Books	GMM with IV	YES	Improvements in the reviews of a competing product decreases sales.
Luca (2016b)	Restaurants	Regression discontinuity	NO	A one-star increase in Yelp rating leads to a 5-9 percent increase in revenue. Ratings do not affect restaurants with chain affiliation.

Table 3.2. Focal hotel characteristics.

Brand	Hotel star ratings (Brand scale*)	Comparable to brands, such as	*valence*	*volume*
Focal Brand 1	4 stars or above (Luxury)	Lowes, Ritz-Carlton, Waldorf Astoria, etc.	4.19 (0.32)	1,781 (1,059)
Focal Brand 2	3.5 stars (Upscale)	Best Western Premier, DoubleTree, Crowne Plaza, etc.	3.83 (0.29)	585 (871)
Focal Brand 3	3.5 stars or 3 stars (Upscale)		4.11 (0.28)	476 (260)
Focal Brand 4	3 stars (Upscale)		4.37 (0.22)	175 (166)
Focal Brand 5	3 stars (Upper midscale)	Comfort Inn, Fairfield Inn, Holiday Inn, etc.	3.91 (0.37)	264 (274)
Focal Brand 6	2.5 stars (Upper midscale)		4.07 (0.34)	134 (129)
Focal Brand 7	2 stars (Midscale)	Quality Inn, Ramada, Wingate by Wyndham, etc.	4.07 (0.31)	89 (112)
Overall			4.06 (0.35)	179 (270)
Total number of hotels	1,992			

Notes. *Brand scale ratings are collected from STR, the primary global provider of competitive benchmarking to the hotel industry (www.str.com). Valence and Volume are calculated using reviews as of the end of the data collection period, i.e., August, 2015. Standard deviations are presented in parenthesis.

Table 3.3. Competitive landscape of focal hotels

Description of competitive landscape	Number of focal hotels	Percentage of focal hotels
At least one same quality competitor exists	1,848	93%
At least one lower quality competitor exists	1,218	61%
At least one higher quality competitor exists	1,039	52%

Table 3.4. Description of independent variables

Independent variables	Description
Online WOM characteristics of focal hotels	
$\boldsymbol{valence_{it}}$	Average rating of all reviews written for hotel *i* until the beginning of month *t*
$\boldsymbol{volume_{it}}$	Cumulative number of reviews written for hotel *i* until the beginning of month *t*
$\mathbf{ln}(\boldsymbol{volume_{it}})$	Natural logarithm of $volume_{it}$
Online WOM characteristics of competing hotels	
$\boldsymbol{valence_{same,it}}$	*Maximum valence* across *all* competing hotels within the same quality (star) level of hotel *i* at the beginning of month *t*
$\mathbf{ln}(\boldsymbol{volume_{same,it}})$	Natural logarithm of *maximum volume* across *all* competing hotels within the same quality (star) level of hotel *i* at the beginning of month *t*
Relative online WOM characteristics	
$\boldsymbol{Relative\_valence_{same,it}}$	$valence_{it} - valence_{same,it}$
$\boldsymbol{Relative\_volume_{same,it}}$	$\ln(volume_{it}) - \ln(volume_{same,it})$
$\boldsymbol{Relative\_valence_{same,pos,it}}$	$=valence_{it} - valence_{same,it}$ if $valence_{it} - valence_{same,it} > 0$ and at least one same quality competitor exists =0 otherwise
$\boldsymbol{Relative\_valence_{same,neg,it}}$	$=\lvert valence_{it} - valence_{same,it}\rvert$ if $valence_{it} - valence_{same,it} < 0$ and at least one same quality competitor exists =0 otherwise
$\boldsymbol{Relative\_volume_{same,pos,it}}$	$=\ln(volume_{it}) - \ln(volume_{same,it})$ if $\ln(volume_{it}) - \ln(volume_{same,it}) > 0$ and at least one same quality competitor exists =0 otherwise
$\boldsymbol{Relative\_volume_{same,neg,it}}$	$=\lvert\ln(volume_{it}) - \ln(volume_{same,it})\rvert$ if $\ln(volume_{it}) - \ln(volume_{same,it}) < 0$ and at least one same quality competitor exists =0 otherwise

Notes. Relative valence and volume with respect to lower and higher quality (star) level competitors are calculated in similar fashion.

Table 3.5. Summary descriptive statistics

Variable name	Mean	Std. dev.	Min	Max
$valence$	4.02	0.44	1	5
$\ln(volume)$	3.91	1.08	0	8.75
Relative online WOM characteristics with respect to same quality competitors				
$Relative\_valence_{same,pos}$	0.11	0.29	0	3.5
$Relative\_valence_{same,neg}$	0.33	0.39	0	3.2
$Relative\_volume_{same,pos}$	0.09	0.26	0	5.28
$Relative\_volume_{same,neg}$	0.58	0.57	0	3.94
Relative online WOM characteristics with respect to lower quality competitors				
$Relative\_valence_{lower,pos}$	0.19	0.41	0	4
$Relative\_valence_{lower,neg}$	0.15	0.31	0	2.46
$Relative\_volume_{lower,pos}$	0.15	0.36	0	4.93
$Relative\_volume_{lower,neg}$	0.22	0.41	0	4.43
Relative online WOM characteristics with respect to higher quality competitor				
$Relative\_valence_{higher,pos}$	0.08	0.22	0	2.83
$Relative\_valence_{higher,neg}$	0.15	0.30	0	3
$Relative\_volume_{higher,pos}$	0.04	0.17	0	4.28
$Relative\_volume_{higher,neg}$	0.38	0.59	0	4.57
Number of observations	107,523			
Number of focal hotels	1,992			

Table 3.6. Correlations among Independent Variables

	1	2	3	4	5	6	7	8	9	10	11
1.$Relative\_valence_{lower,pos}$	1										
2.$Relative\_valence_{lower,neg}$	-0.225	1									
3.$Relative\_valence_{same,pos}$	0.022	-0.142	1								
4.$Relative\_valence_{same,neg}$	-0.104	0.325	-0.330	1							
5.$Relative\_valence_{higher,pos}$	0.066	-0.126	0.100	-0.151	1						
6.$Relative\_valence_{higher,neg}$	-0.129	-0.031	-0.040	0.262	-0.171	1					
7.$Relative\_volume_{lower,pos}$	0.385	0.016	-0.024	-0.012	0.004	-0.066	1				
8.$Relative\_volume_{lower,neg}$	-0.037	0.304	-0.067	0.038	-0.054	-0.115	-0.227	1			
9.$Relative\_volume_{same,pos}$	0.025	0.010	0.250	-0.076	0.023	-0.007	0.161	-0.116	1		
10.$Relative\_volume_{same,neg}$	-0.018	-0.047	-0.142	0.242	-0.009	0.035	-0.175	0.222	-0.335	1	
11.$Relative\_volume_{higher,pos}$	0.018	-0.057	-0.002	-0.015	0.240	0.041	0.096	-0.081	0.116	-0.107	1
12.$Relative\_volume_{higher,neg}$	-0.105	-0.151	0.118	-0.059	0.057	0.449	-0.142	-0.092	-0.078	0.170	-0.140

Notes. Correlations among independent variables do not suggest any problems with multicollinearity.

Table 3.7. The impact of relative reviews with respect to same quality competitor

	DV: $\ln(RevPAR)$					
	No endogeneity correction			GMM Instrumental variables correction heteroskedasticity+ autocorrelation		
	(1)	(2)	(3)	(4)	(5)	(6)
$valence$	0.022**		-0.001	0.023***		0.004
	(0.010)		(0.013)	(0.007)		(0.010)
$\ln(volume)$	0.039***		0.019**	0.056***		0.026*
	(0.007)		(0.009)	(0.009)		(0.016)
$Relative\_valence_{same,pos}$		0.008	0.009		0.002	0.0002
		(0.010)	(0.011)		(0.009)	(0.010)
$Relative\_valence_{same,neg}$		-0.035***	-0.033***		-0.037***	-0.031***
		(0.009)	(0.011)		(0.008)	(0.011)
$Relative\_volume_{same,pos}$		0.012	0.008		0.027	0.027
		(0.008)	(0.008)		(0.021)	(0.021)
$Relative\_volume_{same,neg}$		-0.032***	-0.023***		-0.052***	-0.032*
		(0.006)	(0.008)		(0.011)	(0.019)
Hotel fixed effects	Included	Included	Included	Included	Included	Included
Time fixed effects	Included	Included	Included	Included	Included	Included
Number of focal hotels	1,992	1,992	1,992	1,992	1,992	1,992
Number of observations	107,523	107,523	107,523	107,523	107,523	107,523
R2	0.6856	0.6858	0.6858			

Notes. Coefficients of hotel and time fixed effects are not reported.
*$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8. Extended Model. Impact of relative valence and relative volume on financial performance

	DV: $\ln(RevPAR)$			
	No endogeneity correction		GMM Instrumental variables correction heteroskedasticity+ autocorrelation	
	(1)	(2)	(3)	(4)
Absolute Review Characteristics				
$valence$		-0.017		-0.006
		(0.015)		(0.012)
$\ln(volume)$		0.005		0.017
		(0.011)		(0.019)
Relative Valence				
$Relative\_valence_{lower,pos}$	0.007	0.009	-0.002	-0.001
	(0.011)	(0.011)	(0.010)	(0.010)
$Relative\_valence_{lower,neg}$	-0.011	-0.005	0.008	0.007
	(0.014)	(0.015)	(0.012)	(0.012)
$Relative\_valence_{same,pos}$	0.004	0.008	-0.002	-0.004
	(0.010)	(0.011)	(0.009)	(0.010)
$Relative\_valence_{same,neg}$	-0.026***	-0.034***	-0.030***	-0.032***
	(0.010)	(0.011)	(0.009)	(0.011)
$Relative\_valence_{higher,pos}$	0.011	0.015	0.005	0.006
	(0.018)	(0.018)	(0.015)	(0.015)
$Relative\_valence_{higher,neg}$	-0.019	-0.026*	-0.025**	-0.027**
	(0.013)	(0.015)	(0.011)	(0.013)
Relative Volume				
$Relative\_volume_{lower,pos}$	0.031***	0.030***	0.041**	0.039**
	(0.009)	(0.009)	(0.020)	(0.020)
$Relative\_volume_{lower,neg}$	-0.016*	-0.014	0.006	0.012
	(0.009)	(0.009)	(0.022)	(0.024)
$Relative\_volume_{same,pos}$	0.006	0.005	0.025	0.027
	(0.009)	(0.009)	(0.021)	(0.021)
$Relative\_volume_{same,neg}$	-0.024***	-0.023***	-0.046***	-0.036*
	(0.007)	(0.008)	(0.014)	(0.018)
$Relative\_volume_{higher,pos}$	-0.004	-0.005	0.007	0.007
	(0.011)	(0.011)	(0.021)	(0.021)
$Relative\_volume_{higher,neg}$	-0.010	-0.009	0.001	0.007
	(0.008)	(0.009)	(0.016)	(0.017)
Hotel fixed effects	Included	Included	Included	Included
Time fixed effects	Included	Included	Included	Included
Number of focal hotels	1,992	1,992	1,992	1,992
Number of observations	107,523	107,523	107,523	107,523
R2	0.6861	0.6862		

Notes. Coefficients of hotel and time fixed effects are not reported.
*$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

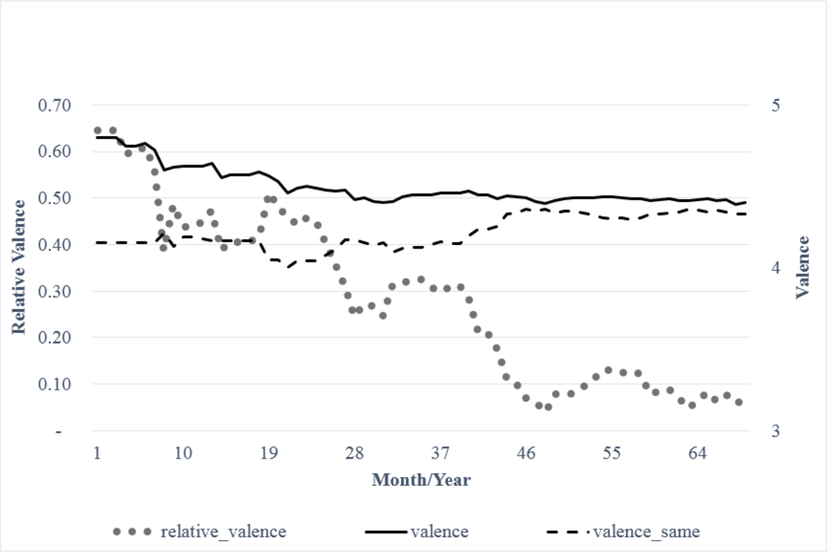
Table 3.9. Robustness Check: Using *average valence* and *volume* across *all* competing hotels within corresponding quality levels

	DV: $\ln(RevPAR)$			
	No endogeneity correction		GMM Instrumental variables correction heteroskedasticity+ autocorrelation	
	(1)	(2)	(3)	(4)
Relative Valence				
$Relative\_valence_{lower,pos}$		0.005		-0.003
		(0.010)		(0.009)
$Relative\_valence_{lower,neg}$		-0.014		-0.010
		(0.019)		(0.015)
$Relative\_valence_{same,pos}$	0.021**	0.017*	0.012	0.009
	(0.010)	(0.010)	(0.008)	(0.008)
$Relative\_valence_{same,neg}$	-0.036***	-0.022*	-0.034***	-0.019*
	(0.012)	(0.013)	(0.010)	(0.011)
$Relative\_valence_{higher,pos}$		0.008		0.0001
		(0.016)		(0.013)
$Relative\_valence_{higher,neg}$		-0.026		-0.038***
		(0.016)		(0.013)
Relative Volume				
$Relative\_volume_{lower,pos}$		0.031***		0.040**
		(0.009)		(0.016)
$Relative\_volume_{lower,neg}$		-0.022*		0.006
		(0.011)		(0.024)
$Relative\_volume_{same,pos}$	0.020**	0.013	0.043***	0.036**
	(0.008)	(0.008)	(0.014)	(0.014)
$Relative\_volume_{same,neg}$	-0.028***	-0.016**	-0.055***	-0.048***
	(0.007)	(0.008)	(0.012)	(0.015)
$Relative\_volume_{higher,pos}$		-0.002		0.018
		(0.010)		(0.019)
$Relative\_volume_{higher,neg}$		-0.012		-0.005
		(0.009)		(0.016)
Hotel fixed effects	Included	Included	Included	Included
Time fixed effects	Included	Included	Included	Included
Number of focal hotels	1,992	1,992	1,992	1,992
Number of observations	107,523	107,523	107,523	107,523
R2	0.6856	0.6860		

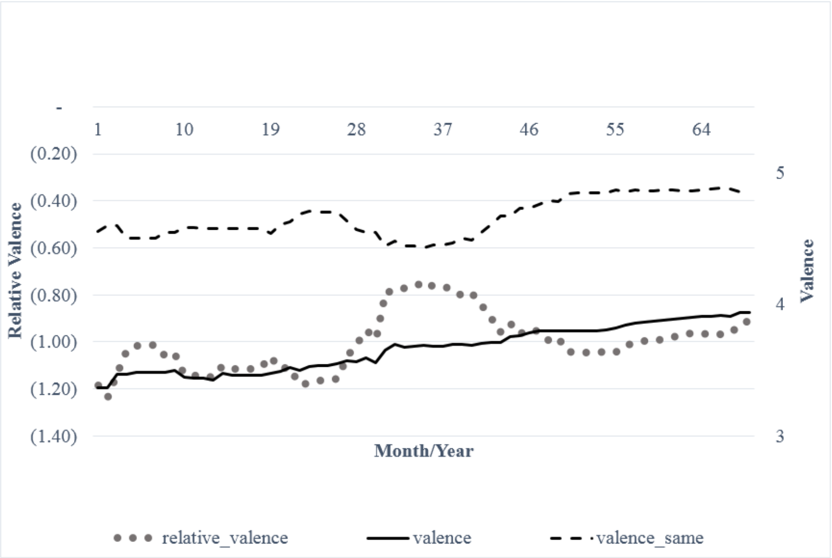
Notes. Coefficients of hotel and time fixed effects are not reported.
p* < 0.10, ** *p* < 0.05, * *p* < 0.01

Figure 3.1. Evolution of relative valence over time

Hotel A



Hotel B



CHAPTER IV: Future Work

The literature on online WOM has demonstrated that online reviews have a significant impact on a product's demand (Archak, Ghose, and Ipeirotis 2011; Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008; Forman, Ghose, and Wiesenfeld 2008; Ghose, Ipeirotis, and Li 2012). In response to a vast number of online reviews, managers are striving to find effective ways to respond to them as part of their reputation management strategy. Management response to online reviews can have multiple consequences: (1) it may have an impact on subsequent online review posting behavior of reviewers on the website, (2) it may have an impact on repeat purchase behavior of the individual who directly receives a response, (3) it may have an indirect impact on repeat purchase behavior of individuals who are exposed to managers' responses at the time of purchase.

Recently, researchers have started investigating the first outcome (Proserpio and Zervas forthcoming; Wang and Chaudhry forthcoming). However, the other two outcomes have been ignored by researchers. In this study, we intend to focus on the second outcome and study the impact of management response on the reviewer who is at the receiving end of the response. More specifically, we aim at identifying textual elements that managers can incorporate in their responses to induce repeat purchase from highly satisfied or dissatisfied customers. We will particularly focus on highly dissatisfied customers because they are the ones who are least likely to repeat purchase from the firm.

Service recovery literature has extensively studied how managers should respond to service failures. This literature shows that complaint management can significantly improve customer satisfaction, repurchase intent and profitability. However, the increasing number of online reviews forced managers to adopt a novel response strategy. Managers have started responding to online reviews by posting responses to them on online review websites. To the best of our knowledge, this paper will be the first to study the impact of such management response on retention of the customer who is the direct recipient of the response.

The most studied consequence of management response is how it impacts online reviewer ratings posted after management starts responding to online reviews. Two recent papers investigate the impact of management response on subsequent review posting behavior. Proserpio and Zervas (forthcoming) investigate the relationship between a firm's use of management responses and its online reputation. They show that when hotels start responding to online reviews they receive fewer but longer negative reviews. Subsequently, Wang and Chaudhry (forthcoming) show that managers' responses to negative reviews positively impact subsequent opinion whereas managers' responses to positive reviews result in smaller effect in the opposite direction.

In sum, the literature to date has been focusing on the impact of management response on subsequent reviewers who are indirect observers of these responses whereas in this study, we will focus on the impact of management response on the direct recipient of that response, especially focusing on the highly dissatisfied consumers. This future study aims at helping managers understand how they should respond to online reviews.

REFERENCES

Adomavicius, Gediminas, Jesse C. Bockstedt, Shawn P. Curley, and Jingjing Zhang (2013), "Do recommender systems manipulate consumer preferences? A study of anchoring effects," *Information Systems Research*, 24 (4), 956–75.

Amabile, Teresa M. (1983), "Brilliant but cruel: Perceptions of negative evaluators," *Journal of Experimental Social Psychology*, 19 (2), 146–56.

——— and Ann H. Glazebrook (1982), "A negativity bias in interpersonal evaluation," *Journal of Experimental Social Psychology*, 18 (1), 1–22.

Anderson, Eric T. and Duncan I. Simester (2014), "Reviews without a purchase: Low ratings, loyal customers, and deception," *Journal of Marketing Research*, 51 (3), 249–69.

Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis (2011), "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science*, 57 (8), 1485–1509.

Ariely, Dan and Jonathan Levav (2000), "Sequential choice in group settings: Taking the road less traveled and less enjoyed," *Journal of consumer Research*, 27 (3), 279–90.

Arndt, Johan (1967), "Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 4 (3), 291–95.

Asch, Solomon E. (1955), "Opinions and social pressure.," *Scientific American*, 193 (5), 31–35.

Babić Rosario, Ana, Francesca Sotgiu, Kristine de Valck, and Tammo H.A. Bijmolt (2016), "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research*, 53 (3), 297–318.

Banerjee, Abhijit V. (1992), "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107 (3), 797–817.

Baum, C. F., M. E. Schaffer, and S. Stillman (2007), "Enhanced routines for instrumental variables/generalized method of moments estimation and testing," *Stata Journal*, 7 (4), 465–506.

Bell, David E. (1982), "Regret in Decision Making under Uncertainty," *Operations Research*, 30 (5), 961–81.

Berger, Jonah and Chip Heath (2007), "Where Consumers Diverge from Others: Identity Signaling and Product Domains," *Journal of Consumer Research*, 34 (2), 121–34.

Bernheim, B. Douglas (1994), "A Theory of Conformity," *Journal of Political Economy*, 102 (5), 841–77.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1998), "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *The Journal of Economic Perspectives*, 12 (3), 151–70.

Blattberg, Robert C. and Kenneth J. Wisniewski (1989), "Price-induced patterns of competition," *Marketing science*, 8 (4), 291–309.

Bolton, Ruth N., P. KandBramlettMatthewD Kannan, and Matthew D. Bramlett (2000), "Implications of loyalty program membership and service experiences for customer retention and value," *Journal of the academy of marketing science*, 28 (1), 95–108.

Bone, Paula Fitzgerald (1995), "Word-of-mouth effects on short-term and long-term product judgments," *Interpersonal buyer behavior in marketing*, 32 (3), 213–23.

Borah, Abhishek and Gerard J. Tellis (2016), "Halo (spillover) effects in social media: do product recalls of one brand hurt or help rival brands?," *Journal of Marketing Research*, 53 (2), 143–60.

Bourne, Francis S. (1957), "Group influence in marketing and public relations," *Some applications of behavioral research*, 207–55.

Bowman, Douglas and Das Narayandas (2001), "Managing customer-initiated contacts with manufacturers: The impact on share of category requirements and word-of-mouth behavior," *Journal of Marketing Research*, 38 (3), 281–97.

Burnkrant, Robert E. and Alain Cousineau (1975), "Informational and Normative Social Influence in Buyer Behavior," *Journal of Consumer Research*, 2 (3), 206–15.

Cai, Hongbin, Yuyu Chen, and Hanming Fang (2009), "Observational Learning: Evidence from a Randomized Natural Field Experiment," *The American Economic Review*, 99 (3), 864–82.

Chen, Yubo, Qi Wang, and Jinhong Xie (2011), "Online social interactions: A natural experiment on word of mouth versus observational learning," *Journal of marketing research*, 48 (2), 238–54.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research (JMR)*, 43 (3), 345–54.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29 (5), 944–57.

Cialdini, Robert B. and Noah J. Goldstein (2004), "Social influence: Compliance and conformity," *Annu. Rev. Psychol.*, 55, 591–621.

Cukierman, Alex (1991), "Asymmetric Information and the Electoral Momentum of Public Opinion Polls," *Public Choice*, 70 (2), 181–213.

Dellarocas, Chrysanthos, Xiaoquan (Michael) Zhang, and Neveen F. Awad (2007), "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," *Journal of Interactive Marketing*, 21 (4), 23–45.

Deutsch, Morton and Harold B. Gerard (1955), "A study of normative and informational social influences upon individual judgment.," *The journal of abnormal and social psychology*, 51 (3), 629.

Dhar, Vasant and Elaine A. Chang (2009), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing*, 23 (4), 300–307.

Dickey, David A. and Wayne A. Fuller (1979), "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American statistical association*, 74 (366a), 427–31.

Dickson, Peter R. and Alan G. Sawyer (1990), "The price knowledge and search of supermarket shoppers," *The Journal of Marketing*, 42–53.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do online reviews matter? — An empirical investigation of panel data," *Information Technology and Systems in the Internet-Era*, 45 (4), 1007–16.

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008), "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, 19 (3), 291–313.

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

Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman (1996), "Crime and Social Interactions," *The Quarterly Journal of Economics*, 111 (2), 507–48.

Godes, David and Dina Mayzlin (2004), “Using Online Conversations to Study Word-of-Mouth Communication.,” *Marketing Science,* 23 (4), 545–60.
——— and José C. Silva (2012), “Sequential and Temporal Dynamics of Online Opinion,” *Marketing Science*, 31 (3), 448–73.
Gupta, Sunil and Valarie Zeithaml (2006), “Customer Metrics and Their Impact on Financial Performance,” *Marketing Science,* 25 (6), 718–39.
Hardie, Bruce GS, Eric J. Johnson, and Peter S. Fader (1993), “Modeling loss aversion and reference dependence effects on brand choice,” *Marketing science*, 12 (4), 378–94.
Helson, Harry (1964), “Adaptation-level theory.”
Hirschman, Albert O. (1970), *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*, Harvard university press.
Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein (2004), “Social Interaction and Stock-Market Participation,” *The Journal of Finance,* 59 (1), 137–63.
Hsee, Christopher K. (1996), “The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives.”
——— and Jiao Zhang (2010), “General Evaluability Theory,” *Perspectives on Psychological Science*, 5 (4), 343–55.
Huber, Joel, John W. Payne, and Christopher Puto (1982), “Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis,” *Journal of consumer research*, 90–98.
Iacus, Stefano M., Gary King, and Giuseppe Porro (2011), “Multivariate matching methods that are monotonic imbalance bounding,” *Journal of the American Statistical Association,* 106 (493), 345–61.
———, ———, and ——— (2012), “Causal Inference without Balance Checking: Coarsened Exact Matching,” *Political Analysis*, 20 (1), 1–24.
Inman, J. Jeffrey, James S. Dyer, and Jianmin Jia (1997), “A Generalized Utility Model of Disappointment and Regret Effects on Post-Choice Valuation,” *Marketing Science*, 16 (2), 97–111.
Ismail, Joseph A., Michael C. Dalbor, and Juline E. Mills (2002), “Using RevPAR to analyze lodging-segment variability,” *The Cornell Hotel and Restaurant Administration Quarterly*, 43 (6), 73–80.
Jabr, Wael and Z. Zheng (2014), “Know Yourself and Know Your Enemy: An Analysis of Firm Recommendations and Consumer Reviews in a Competitive Environment,” *Mis Quarterly*, 38 (3), 635–54.
Jonathan E. Alevy, Michael S. Haigh, and John A. List (2007), “Information Cascades: Evidence from a Field Experiment with Financial Market Professionals,” *The Journal of Finance*, 62 (1), 151–80.
Kahneman, Daniel and Amos Tversky (1979), “Prospect theory: An analysis of decision under risk,” *Econometrica: Journal of the Econometric Society*, 263–91.
Kalnins, Arturs (2017), “Pricing Variation Within Dual-Distribution Chains: The Different Implications of Externalities and Signaling for High- and Low-Quality Brands,” *Management Science*, 63 (1), 139–52.
Kalyanaram, Gurumurthy and Russell S. Winer (1995), “Empirical Generalizations from Reference Price Research,” *Marketing Science,* 14 (3), G161–69.

King, Robert Allen, Pradeep Racherla, and Victoria D. Bush (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), 167–83.

Lee, Young-Jin, Kartik Hosanagar, and Yong Tan (2015), "Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings," *Management Science*, 61 (9), 2241–58.

Lerner, Jennifer and Philip Tetlock (1999), "Accounting for the effects of accountability," *Psychological Bulletin*, 125, 255–75.

Li, Xinxin and Lorin M. Hitt (2008), "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), 456–74.

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

Luca, Michael (2016a), "User-Generated Content and Social Media," in *Handbook of Media Economics*, North-Holland Publishing Company, Chap. 12.

——— (2016b), "Reviews, Reputation, and Revenue: The Case of Yelp.com.," *Harvard Business School*, Working Paper No. 12-016.

——— and Georgios Zervas (forthcoming), "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud.," *Management Science*.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014), "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, 104 (8), 2421–55.

Mazumdar, Tridib, S. P. Raj, and Indrajit Sinha (2005), "Reference price research: Review and propositions," *Journal of marketing*, 69 (4), 84–102.

Minnema, Alec, Tammo H.A. Bijmolt, Sonja Gensler, and Thorsten Wiesel (2016), "To Keep or Not to Keep: Effects of Online Customer Reviews on Product Returns," *Journal of Retailing*, 92 (3), 253–67.

Moe, Wendy W. and David A. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution.," *Marketing Science*, 31 (3), 372–86.

——— and Michael Trusov (2011), "The value of social dynamics in online product ratings forums," *Journal of Marketing Research*, 48 (3), 444–56.

Moe, Wendy W and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums.," *Journal of Marketing Research (JMR)*, 48 (3), 444–56.

Muchnik, Lev, Sinan Aral, and Sean J. Taylor (2013), "Social Influence Bias: A Randomized Experiment," *Science*, 341 (6146), 647–51.

Murray, Keith B. and John L. Schlacter (1990), "The impact of services versus goods on consumers' assessment of perceived risk and variability," *Journal of the Academy of Marketing science*, 18 (1), 51–65.

Nagle, Frank and Christoph Riedl (2014), "Online word of mouth and product quality disagreement," in *Academy of Management Proceedings*, Academy of Management, 15681.

Newey, Whitney K. and Kenneth D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55 (3), 703–8.

Porac, Joseph F. and Howard Thomas (1990), "Taxonomic mental models in competitor definition," *Academy of management Review*, 15 (2), 224–40.

——— and ——— (1994), “Cognitive categorization and subjective rivalry among retailers in a small city.,” *Journal of Applied Psychology,* 79 (1), 54–66.
———, ———, Fiona Wilson, Douglas Paton, and Alaina Kanfer (1995), “Rivalry and the Industry Model of Scottish Knitwear Producers.,” *Administrative Science Quarterly,* 40 (2), 203–27.
Prelec, Drazen, Birger Wernerfelt, and Florian Zettelmeyer (1997), “The role of inference in context effects: Inferring what you want from what is available,” *Journal of Consumer research,* 24 (1), 118–25.
Proserpio, Davide and Georgios Zervas (forthcoming), “Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews,” *Marketing Science*.
Rossi, Peter E. (2014), “Even the rich can make themselves poor: A critical examination of IV methods in marketing applications,” *Marketing Science*, 33 (5), 655–72.
Sahoo, Nachiketa, Chrysanthos Dellarocas, and Shuba Srinivasan (2016), “The Impact of Online Product Reviews on Product Returns,” *Boston U. School of Management Research Paper,* No. 2491276.
Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts (2006), “Experimental study of inequality and unpredictability in an artificial cultural market,” *Science*, 311 (5762), 854–56.
——— and Duncan J. Watts (2008), “Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market,” *Social Psychology Quarterly*, 71 (4), 338–55.
Schlosser, Ann E. (2005), “Posting versus lurking: Communicating in a multiple audience context,” *Journal of Consumer Research*, 32 (2), 260–65.
Sherif, Muzafer, Daniel Taub, and Carl I. Hovland (1958), “Assimilation and contrast effects of anchoring stimuli on judgments.,” *Journal of Experimental Psychology,* 55 (2), 150.
Simonson, Itamar (1989), “Choice based on reasons: The case of attraction and compromise effects,” *Journal of consumer research*, 158–74.
——— (1992), “The Influence of Anticipating Regret and Responsibility on Purchase Decisions,” *Journal of Consumer Research*, 19 (1), 105–18.
——— and Amos Tversky (1992), “Choice in context: Tradeoff contrast and extremeness aversion,” *JMR, Journal of Marketing Research*, 29 (3), 281.
Sonnier, Garrett P., Leigh McAlister, and Oliver J. Rutz (2011), “A Dynamic Model of the Effect of Online Communications on Firm Sales,” *Marketing Science,* 30 (4), 702–16.
Sridhar, Shrihari and Raji Srinivasan (2012), “Social influence effects in online product ratings,” *Journal of Marketing*, 76 (5), 70–88.
Sun, Monic (2011), “How Does the Variance of Product Ratings Matter?,” *Management Science,* 58 (4), 696–707.
Tellis, Gerard J. and Joseph Johnson (2007), “The value of quality,” *Marketing Science*, 26 (6), 758–73.
Tomz, Michael, Jason Wittenberg, and Gary King (2003), “CLARIFY: Software for interpreting and presenting statistical results,” *Journal of Statistical Software*, 8 (1), 1–30.
Tucker, Catherine and Juanjuan Zhang (2011), “How Does Popularity Information Affect Choices? A Field Experiment,” *Management Science,* 57 (5), 828–42.
Venkatesan, M. (1966), “Experimental Study of Consumer Behavior Conformity and Independence,” *Journal of Marketing Research,* 3 (4), 384–87.

Wang, Y and A Chaudhry (forthcoming), "To See or not to See: Salience and the sentiment externalities of managers' responses to online reviews," *Journal of Marketing Research*.

Wangenheim, Florian and Tomás Bayón (2004), "Satisfaction, loyalty and word of mouth within the customer base of a utility provider: differences between stayers, switchers and referral switchers," *Journal of Consumer Behaviour*, 3 (3), 211–20.

Wu, Fang and Bernardo A. Huberman (2008), "How public opinion forms," in *Internet and Network Economics*, Springer, 334–41.

Yoon, Song-Oh and Itamar Simonson (2008), "Choice set configuration as a determinant of preference attribution and strength," *Journal of Consumer Research*, 35 (2), 324–36.

Zeithaml, Valarie A. (1981), "How Consumer Evaluation Processes Differ Between Goods and Services."

Zhu, Feng and Xiaoquan (Michael) Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74 (2), 133–48.

Yelp (2016), "An Introduction to Yelp Metrics," (accessed June 2, 2016), [available at http://www.yelp.com/factsheet]