CONSUMER RESPONSE TO USER GENERATED CONTENT

AND ONLINE ADVERTISING

by

Yiyi Li

APPROVED BY SUPERVISORY COMMITTEE:

Ying Xie, Chair

Dmitri Kuksov

Brian T. Ratchford

Zhiqiang (Eric) Zheng

In dedication to my husband Ting, our daughter Claire, and my parents.

CONSUMER RESPONSE TO USER GENERATED CONTENT AND ONLINE ADVERTISING

by

YIYI LI, BBA

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

MANAGEMENT SCIENCE

THE UNIVERSITY OF TEXAS AT DALLAS

August 2017

ACKNOWLEDGMENTS

First and foremost, I want to thank my advisor Dr. Ying Xie. She has encouraged me to pursue the research topics that interest me and has taught me how to make good ideas into great papers. She has also provided an excellent example as a successful female professor. I enjoyed every discussion with her because they are always productive and inspiring. I could not have imagined having a better advisor and mentor for my PhD study.

I would also like to thank the rest of my dissertation committee: Dr. Brian T. Ratchford, Dr. Eric Zheng and Dr. Dmitri Kuksov for their time and insightful comments to better my work. I thank them for their contribution and their good-natured support.

I am grateful to Dr. Xiongwen Lu, Dr. Qingyun Jiang and Dr. Yiping Song from Fudan University for enlightening me the first glance of research and encouraging me to pursue a PhD degree in the United States back to my undergraduate studies.

Lastly, I would like to thank my family for all their love and spiritual support. For my parents who made me an independent and rational person and gave me the strength to chase my dreams. For my parents-in-law who sacrificed their own time to help me with my life. For my loving and supportive husband Ting who encouraged me during the tough times in the PhD pursuit. And finally, for my beloved daughter Claire who is such an adorable girl always cheering me up. Thank you all!

June 2017

CONSUMER RESPONSE TO USER GENERATED CONTENT

AND ONLINE ADVERTISING

Yiyi Li, PhD
The University of Texas at Dallas, 2017

Supervising Professor: Ying Xie

The bursts and multiplicity of online user generated content and Internet advertising have posed new challenges to marketing managers. In order to design effective marketing campaigns and allocate advertising budgets wisely, we must understand the role of how consumers respond to user generated content and different advertising formats. I explore two specific questions in this domain in the present dissertation: imagery content in user generated content, and multichannel advertising in a competitive setting.

The first part of the dissertation examines three effects of imagery content in a social media post on users' liking and sharing behaviors. The results indicate a strong mere presence effect: the inclusion of an immediately viewable picture in a Tweet lifts the number of Likes and Retweets for the Tweet by 231% and 479% respectively. However, Tweets with linked pictures receive fewer Likes and Retweets than Tweets without any pictures. Moreover, we find that pictorial characteristics have disparate effects on liking and sharing: personal content increases liking while artistic and informational content increases sharing. Last, we find that picture-text congruency also matters: unexpected pictures increase both liking and sharing, while relevant pictures only increase

liking but not sharing. These results will shed light on how to improve social media content strategies and foster reader engagement.

The second part of this dissertation extends the first chapter by further examining how online reviews with photos are different from those without. Online review is an important information source for both consumers and marketers. However, the aggregated rating for a business or product may not fully reflect individual consumer's opinion due to several reasons. This study contributes to the literature by examining the association between star ratings on restaurants and the existence of photo in the review. Meanwhile, we also aim to understand why some consumers choose to upload photos to accompany text reviews. We use a data set containing reviews and photos on restaurants provided by Yelp.com and find that reviews with photos tend to have higher star ratings compared to those without photos. We also find that photo uploading decision is affected by the consumer's intention to show effort, the motivation to gain image-utility and the underlying topic and valence of the text review. We provide two explanations for our findings and discuss the implications for marketers and consumers.

The third part of this dissertation extends the multi-channel advertising attribution literature by developing an integrated individual-level choice model that considers consumers' online visit and purchase decisions across all competitors within one industry. We specifically analyze the effects of multi-channel advertising on: (1) consumer choice of entry site, (2) consumer search decisions concerning the remaining websites that compete in the same industry, and (3) subsequent purchase at one of the searched websites. We quantify the impact of different digital advertising channels on consumers' decisions at different purchase funnel stages based on an individual-level click stream data for the online air ticket booking industry. We find that information stock collected

through all advertising channels contributes significantly to consumers' visit and purchase decisions, among which search advertising is more effective in driving the choice of entry site while email advertising has a larger effect on visit decision concerning remaining websites and purchase decision. The own- and cross-marginal impacts of various ad channels on each firm vary widely across competitors, and this is true at all purchase funnel stages. We also show that neglecting competition may lead to underestimated advertising effects and worse predictions, by comparing the estimated advertising effectiveness and predictive performance of our proposed model with those of the common baseline model that only models consumers' binary purchase decisions on a focal firm.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

IMAGERY CONTENT AND SOCIAL MEDIA ENGAGEMENT

1.1 Introduction

The past few years have witnessed a swift shift in social media platforms from text-centric webpages to visual-oriented experience. The trend towards visual social media is partly driven by the shifting habits of social media users thanks to the popularization of smartphones and improved mobile Internet experience. As people engage with social media via apps on their smartphones, they quickly find out that taking a picture "on the go" using a high-resolution phone camera is much easier than typing a status update on a tiny keyboard. According to KPCB (2016), social network users were sharing an average of 3.2 billion digital images every day in 2015 on Snapchat, Facebook, Instagram and WhatsApp combined. It is no coincidence that Instagram and Pinterest, two image-centric social media apps, have quickly risen to become the second and fourth largest social networks in the United States in less than five years of time (eMarketer 2015). The old idiom "A picture is worth a thousand words" has become the new maxim among social media marketers.

Although anecdotal evidence is abundant that people love to post, view and share pictures online, we know little about whether social media posts with pictures are indeed more popular than those without, and furthermore, which pictorial characteristics induce more interaction and propagation. While many studies have examined the determinants that drive virality or popularity in the context of user-generated content, the vast majority of these studies primarily focus on the text or video content (e.g., Berger and Milkman 2012, Porter and Golan 2006, Tucker 2015), leaving the role of imagery content a largely unexplored question. Is it true that including pictures in a post always leads to more views and higher engagement as some industry studies seem to

suggest (Vavrek 2012, WebLiquid 2011)? Alternatively, should only pictures with certain characteristics be included to boost the popularity of a social media post while others be avoided? Answers to these questions will guide social media users to create more effective social media messages to engage readers. They will also help brand managers improve the efficiency of their brand-related social media listening efforts by mining imagery content for additional insights.

In this study, we aim to answer these questions by quantifying the impact of imagery content on users' social media engagement.[1] Drawing from the extant literature on visual marketing, we propose three effects through which imagery content can affect the appeal of social media post and thus lead to higher user engagement: (1) "Mere Presence Effect": The imagery content in a social media post helps the post stand out from the media clutter when the majority of online content is text-based. Attracting eyeballs is the very first step before any further engagement takes place. Furthermore, some social platforms allow users to share a hyperlink to a picture posted on other social platforms. We specifically distinguish these linked pictures from the immediately viewable pictures in our analysis to capture the benefit of exhibiting imagery content directly in a post. (2) "Pictorial Characteristics Effect": The imagery content in a social media post may provide informational or aesthetic value independent of the text content and therefore may increase the total appeal of the post. On the contrary, an image with uninteresting content or low quality may backfire and lead to lower user engagement. (3) "Picture-Text Congruency Effect": Congruency involves both *expectancy* and *relevancy*. A long list of studies in visual advertising have shown that while unexpected pictures in an advertisement can enhance ad recall and lead to more favorable attitude, irrelevant pictures may have the opposite effect because they create extra

[1] We use "imagery content", "image", and "picture" interchangeably in this chapter.

difficulty for the readers to comprehend the main message (Heckler and Childers 1992, Lee and Mason 1999). We contemplate that the congruency between the text content and the imagery content in a social media post is likely to have similar effects on user engagement with the posts.

We use a large dataset of brand related posts collected from Twitter.com to quantify the three proposed effects of imagery content. The data include the text and imagery content of 16,897 original Tweets posted in September 2015 that mention one of the following five airline companies: JetBlue Airways, American Airlines, Southwest Airlines, Virgin America, and Frontier Airlines. We follow the standard industry practice to measure engagement level of each Tweet using the number of Likes and Retweets. We collected the data six months after the original Tweets were posted to ensure that the observation period is long enough to capture all user engagement with the posts. As Twitter users' motivations to like or to retweet a post may differ from each other, the pictorial characteristics that induce liking may also differ from those that encourage retweeting. By treating Likes and Retweets as correlated yet distinct engagement measures, we examine how imagery content in a Tweet may affect readers' liking and retweeting behaviors differently.

One of the biggest challenges in analyzing online user-generated content, as well documented in previous literature (Netzer et al. 2012), is to convert the overwhelming quantity of unstructured data into usable information. Thanks to the advancements in automated text mining techniques, several recent studies have made use of the information contained in text data beyond simple volume to derive useful insights from user-generated content (see Purnawirawan et al. 2015, Pang and Lee 2008 for two reviews of relevant studies). Following this literature, we adopt the Naïve Bayes classifier, a commonly used text mining algorithm, to analyze the text content in

the Tweet posts. However, the emergence of the visual web has presented us with new challenges in mining imagery data. In this study, we employ Google Cloud Vision API (Application Program Interface) that encapsulates powerful machine learning models, together with manual coding to extract information from the imagery data.

We develop a Zero-Inflated Bivariate Negative Binomial model to examine how imagery content affects the number of Likes and Retweets received by original Tweet posts. The presence of excess zeros (zero-inflation) in the number of Likes and Retweets in our data is primarily driven by the fact that many Tweets failed to attract attention from readers. This case is often termed as structural zeros in count data analysis. We use a binary logit model to account for the probability that a Tweet receives sufficient attention from the audience to warrant their further interactions with the post through either liking or retweeting. Once this minimal level of attention is attained, the Bivariate Negative Binomial component in the model is used to estimate the impact of imagery content on the expected number of Likes and Retweets, while controlling for other post and account characteristics. The inclusion of imagery content in a Tweet post can be an endogenous decision, which may depend on factors that also influence user engagement with the post. In an additional analysis, we use the propensity score subclassification approach to address the endogeneity concern and test the robustness of the mere presence effect.

Our findings lent support for all three proposed effects of imagery content on social media engagement. First, we find a significant positive mere presence effect: the inclusion of an immediately viewable image in a Tweet lifts the number of Likes for the Tweet by 231% and the number of Retweets by 479%. However, the “linked” pictures have the adverse effect that Tweet posts with picture hyperlinks receive fewer Likes and Retweets than Tweets without any pictures.

Second, we find that pictorial characteristics of the imagery content matter: while characteristics such as the colorfulness increase both liking and retweeting, other characteristics are found to have varying effects on liking and retweeting. For example, Tweets with pictures featuring human faces receive more Likes but fewer Retweets while artistic pictures in Tweet posts only increase the number of Retweets but not the number of Likes. Third, we find that an unexpected picture increases both liking and sharing, but a relevant picture only affects liking but not sharing.

Our research contributes to the rapidly growing literature on social media content effectiveness. Previous studies in this area have primarily focused on the text or video content, paying much less attention to the role of imagery content. Given the rapid rise of visual social media, it is timely that we expand the current knowledge on social media content effectiveness by exploring how imagery content affects social media engagement. Our results offer important managerial implications to social media content providers aiming to enhance their content effectiveness as well as companies seeking to improve the efficiency of their brand related social media listening effort. Such insights are especially useful for service industries where many customers make service complaints through the brand official social media pages or their own personal social media accounts. Better identification of potentially influential posts gives these companies an opportunity to take actions before the harmful social buzz becomes viral.

Second, our research also adds to the literature on visual marketing. Extant research primarily focuses on the impact of visual elements on advertising effectiveness. With the wide spread of smartphones and ubiquitous access to Internet, imagery content in social media is increasingly created by ordinary users rather than professional marketing and advertising agencies. Content effectiveness metrics in social media are broadened to include many pre-purchase engagement

measures in addition to brand and product sales. Our research extends the literature on visual marketing from the traditional "paid media" channels to the emerging "earned media" area. Our findings regarding the mere presence effect and the pictorial characteristics effect of the imagery content are generally in line with the previous literature. Noticeably, our finding that picture-text relevancy increases liking but does not affect retweeting provides evidence for the Elaboration Likelihood Model (ELH) in "earned media" (Petty and Cacioppo 1986). Since liking is more personal than retweeting and happens more frequently between familiar users, readers' elaboration level is likely to be higher as well. Therefore, our result is consistent with the ELM's prediction that the influence of picture-text relevancy should grow as involvement increases.

1.2 Background

In this section, we first introduce the two social media engagement metrics we will use in the current study and describe how these two metrics may differ from each other. We then provide a brief literature review on imagery content in print and banner advertising and discuss how imagery content may affect the two social media engagement metrics based on the extant literature.

1.2.1 User Engagement on Social Media

Kaplan and Haenlein (2010) describe social media as a group of Internet-based applications allowing the creation and exchange of user generated content. People share not only their daily life but also opinions on companies and brands on social media, making it possible for marketers to gain powerful insights into their customers through social listening. At the meanwhile, marketers also use social media brand pages as an online communication portal to broadcast brand-related updates and interact with their customers. As a result, firms have continued to increase their spending on social media marketing (Moorman 2016).

Despite the fact that social media has become a major marketing tool, demonstrating its contribution remains a major challenge for many marketers (Moorman 2016). Among many other metrics, engagement has been suggested by many marketing practitioners as the start point to measure the influence of social media posts. Social media engagement measures can be grouped into two broad types. The first type includes direct responses to the original post including likes, comments, and favorites. In this research, we focus on liking, a commonly adopted metric, which allows readers to show enjoyment, appreciation or endorsement of the content without leaving a comment. For example, Twitter and Facebook have "Like", Google+ has "+1", and many blogs have "vote up" or similar measures. The second type concerns sharing or propagation of original posts, which allows the audience to recommend the content to their followers. Examples include Retweet for Twitter, Shares on Facebook and Google+, and Repins on Pinterest.

Although both metrics reflect a deeper level of engagement than simple views, we argue that social media users' motivation to like or to share a post may differ due to the varying visibility caused by these two actions. Take Twitter for example - once a user sends a Retweet, the retweeted content will be updated on "Home" page of all her followers, effectively expanding the content exposure beyond the original poster's own follower base. In contrast, the action of liking is more private in the sense that it gives affirmation to the poster but does not propagate the content. As a result, what people share are usually different from what they like. Toubia and Stephen (2013) find that social media users derive both *intrinsic utility* (the inherent satisfaction gained from doing of an activity) and *image-related utility* (perceptions of others, status seeking or prestige) through posting contents. When choosing which content to share, a social media user will consider how her followers might perceive her actions due to its high social visibility; therefore, image utility

may play a much bigger role in the sharing decision than in the liking decision. Past research has shown that people are more likely to share positive, high-arousal (Berger and Milkman 2012) and humorous (Golan and Zaidner 2008, Brown et al. 2010) content because such content reflects positively on the sharer. They are also more likely to share useful information to appear knowledgeable (Wojnicki and Godes 2008). When it comes to decide whether to "like" a post on social media, Seiter (2015) suggests a social media user is driven by one of the following motivations: 1) to show agreement or support; 2) to affirm something about herself; 3) to express virtual empathy; 4) to get benefit (such as a coupon) in return. These four motivations are mostly associated with intrinsic utility, which we believe is the primary driver for the liking behavior. Due to the difference in the underlying motivations, we expect that imagery content in a Tweet may affect readers' liking and sharing behaviors differently.

1.2.2 Imagery Content and Social Media Engagement

Social media users are subject to information overload. A post needs to stand out from many others to gain awareness and attention from readers before any engagement takes place. This is very similar to the challenge faced by banner or print ads because they also need to compete for consumers' attention before any action further down the purchase funnel occurs. In this sub-section, we briefly review the literature on visual advertising research based on which we discuss how imagery content may affect social media engagement.

However, before relating to the visual advertising literature, we must note that social media posts differ from advertisements in at least two important aspects. First, advertisements are designed carefully by professionals who are skilled and experienced, while the majority of social media posts are created by ordinary people for the purpose of recording and sharing their daily

life. Even though pictorial characteristics may still work the same way in these two settings, we expect the picture quality to be worse on average and to vary more from post to post in social media. Second, in the visual advertising literature, the outcome measure of consumer interest is more or less related to purchase intention. However, in social media, the content itself is the "product" for consumption and is often actively sought by social media users. The more relevant and immediate measures of user interest in this particular setting would be those reflecting the popularity of the posts such as the number of views, likes, comments, or shares. Given these differences, it remains an unanswered empirical question whether the imagery characteristics that affect visual advertising effectiveness would also affect the popularity of social media posts.

According to the extant visual advertising literature, imagery components in advertising design can affect cognitive outcomes (such as attention, attitude and preference) and behavioral outcomes (such as clicks, purchase intention or sales) in three ways. First, the mere presence of imagery content matters. Using eye-tracking methodology, Pieters and Wedel (2004) find that the pictorial component in print ads can capture superior baseline attention of magazine readers regardless of its size. Various studies on banner ads also show that obtrusive ads are more effective because they grab a viewer's attention in the first place (Goldfarb and Tucker 2011, Bruce et al. 2016). Based on these findings, we expect that the mere presence of imagery content in a social media post helps the post stand out from the vast majority of text only posts and as a result attract more attention. Attention is the pre-requisite for any further interactions with the post. With hundreds of millions of posts generated on Twitter every day, it is very likely that a Tweet would be washed away by new Tweets before it attracts any attention. Beyond attention, imagery content may also affect engagement directly by enhancing the perceived quality of the Tweet. As most of

the Tweets consist of text content only, readers may appreciate Tweets with pictures more because of the extra effort put in by the posters when composing such Tweets. As a result, they may be more likely to engage with these posts.

Second, extant literature has shown that imagery content affects attention, attitude, affect, or purchase intention through various pictorial characteristics beyond the mere presence effect. Finn (1988) reports that color is a powerful pictorial characteristic that consistently enhances a viewer's attention for advertisements based on findings from more than a dozen research studies. More recently, Wedel and Pieters (2015) find that color helps ads quickly communicate their meaning during brief and blurred exposures. Social media posts are subject to similar brief and blurred exposures because users often scroll down pages rapidly when surfing on social network apps. Therefore, we expect that colorful pictures help increase user engagement with social media content by attracting attention and enhancing information processing.

Objects featured in an image and the design complexity of the pictorials are also found to affect the effectiveness of ads and website design (e.g., Mitchell and Olson 1981, Petty et al. 1983, Palmer 1999, Pieters et al. 2010). Cyr et al. (2009) find that human images enhance website aesthetic and emotional appeal for users. Xiao and Ding (2014) show that hedonic facial features affect advertising effectiveness, but the effect size varies across people and product category. Pieters et al. (2010) demonstrate that design complexity such as the quantity of objects, asymmetry and irregularity of object arrangement increases both attention and improves attitude towards the ad. Following these studies, we also consider objects featured in the image, and the asymmetry of object arrangement as pictorial characteristics that may affect social media engagement in our analysis.

Lastly, since advertisements are usually composed of both visual and verbal elements, the relationship between these two components affects advertising effectiveness (e.g., Houston et al. 1987, Heckler and Childers 1992, Lee and Mason 1999). Following Heckler and Childers (1992), we focus on two dimensions of congruency, *expectancy* and *relevancy*, which are hypothesized to affect information processing in social cognition literature (Goodman 1980). *Expectancy* refers to whether the imagery content falls into some predetermined pattern or structure evoked by the verbal information. Heckler and Childers (1992) find that unexpected information lead to superior ad recall because greater processing effort is required to encode the information and the successful encoding of the ambiguity in the unexpected visual stimulus is rewarding and pleasurable for the reader (Berlyne 1971). Lee and Mason (1999) report similar findings that ads with unexpected information elicited more favorable attitudes than did ads with expected information. Based on findings from these studies, we expect that pictures with unexpected information will elicit greater cognitive elaboration and thus increase engagement.

Relevancy reflects whether information contained in the picture contributes to or detracts from the clear identification of the verbal information. Heckler and Childers (1992) show irrelevant information negatively affects ad recall because it is less easily recognized. This is because the irrelevant visual stimulus creates extra difficulty for the reader to make sense of the ad information and therefore may lead to frustration rather than resolution. Lee and Mason (1999) also find that ads with irrelevant information yielded less favorable attitudes than did ads with relevant information. However, the effect of *relevancy* can be mediated by reader's involvement level with the product. Petty and Cacioppo's (1986) elaboration likelihood model (ELM) provides a good explanation for how elaboration (i.e., issue-relevant thinking) moderates the relationship

between the *relevancy* of a picture (to the verbal message in the ad or to the product) and advertising persuasiveness. According to ELM, the influence of relevant pictures grows when an individual gets more involved in issue-relevant thinking. Miniard et al. (1991) test the implications of this theory using two experiments and show that higher product involvement reduces the persuasive impact of peripheral pictures, but enhances the influence of product-relevant pictures. Following these studies, we are interested in finding out how the *relevancy* between the imagery content and the text content in a social media post affects user engagement. On one hand, relevant pictures can help identify the primary message conveyed by the text content and therefore may enhance a user's comprehension of the message. On the other hand, social media users are usually in a low involvement mental state due to multi-tasking while browsing their social media pages. Under the ELM framework, we should expect picture *relevancy* to have a rather small or a null effect on user engagement with social media posts.

1.2.3 Modeling Framework

Besides the imagery component, users' engagement with a social media post is also influenced by the text component in the post, the account characteristics, and the timing of the post. We consider all these factors in our model. We present our modeling framework for the determinants of social media post engagement in Figure 1.1.

Peters et al. (2013) suggest that (text) content should be a major element of social media metrics. They further propose that content characteristics (e.g., interactivity, vividness, usefulness), content domain (e.g., topic, fact vs. opinion), and content valence (i.e., positive, negative, neutral) are important content factors to be considered. Following their proposal, we consider the following characteristics of the text component in our modeling framework:

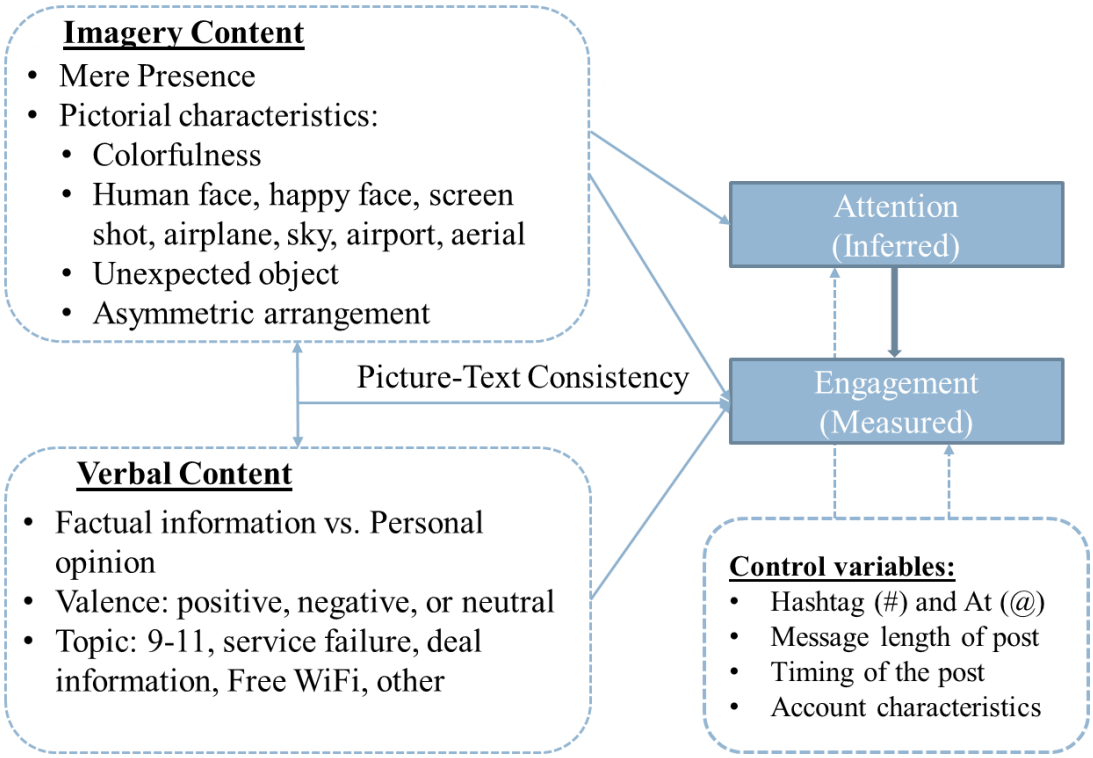


Figure 1.1. Modeling Framework

1) Factual information versus personal opinion: whether the text content contains objectively verifiable information about the firm, or whether it conveys a personal opinion;

2) Valence: whether the text content is positive, negative or neutral;

3) Topic: the theme of the post, such as complaints, deal information, etc.

1.3 DATA DESCRIPTION

For our empirical investigation, we obtained a dataset of brand related Twitter posts from a social media consultancy. This dataset includes the text and imagery content (if any) of every Tweet posted in September 2015 if at least one of the following keywords is mentioned: JetBlue,

TrueBlue, American Airlines, Frontier Airlines, Southwest Airlines, and Virgin America. As one of the largest sectors in the service industry, airline industry is an ideal setting for this study because many customers use social media websites as their primary platforms to share experiences, seek information and make complaints to companies directly. All five airline companies in our study have official Twitter accounts, and they actively manage their Twitter accounts by regularly posting brand related information and interacting with customers on their official brand pages.

We eliminated all Tweets that include hyperlinks to other webpages because the engagements received by these Tweets can be confounded with the content quality of the original webpages. We also eliminated all Retweeted Tweets for the same reason. Our final sample contains 16,896 original Tweets. We then collected the number of Likes and Retweets received by each Tweet in our data sample in March 2016, i.e., six months after the Tweets were posted. Industry reports have shown that the life span of a normal Tweet is less than half an hour (Rey 2014). Therefore, we believe a six-month window is long enough to capture any potential engagement. The data exhibits a left skewed distribution for both Likes and Retweets: 84.97% of the Tweets in our sample did not receive any Likes or Retweets. Among those Tweets which received at least one Like or Retweet, the average number of Likes per Tweet is 12.07 (SD = 40.76) and the average number of Retweets per Tweet is 19.48 (SD = 54.38).

We present the summary statistics of our estimation sample in Table 1.1. On average, the Tweets in our data are 16 words long, which is quite short due to the 140-character limit imposed by Twitter, a micro-blogging application. Regarding timing of the post, 47.8% of Tweets in our data are posted during 8 a.m. to 4 p.m., and 75.64% of Tweets are posted during workdays. This is consistent with industry reports about the peak time of Twitter activity (Sysomos 2009). We

have information on three account characteristics: number of followers, the Klout score and account verification status. On average, an account has 17,834.5 followers. However, the distribution of number of followers is strongly left-skewed. The median number of follower is 389, much smaller than the mean. The Klout score is an index developed by Klout, Inc. to indicate the influence power of an online account (Rao et al. 2015). It is a score between 1 and 100 with a higher score indicating a larger influence. The average Klout score is 39.4, with a standard deviation of 15.9. The verification status indicates the credibility of the account: 3.64% of Tweets in our data are posted by verified accounts.

Since the text and imagery content contained in a Tweet post are unstructured information, we have to transform them into meaningful numerical measures that can be readily used in quantitative analysis. To process such a large amount of unstructured information efficiently, we use machine learning techniques to label text content and image processing tools to classify image content. Measurement error is inevitable during the process of text and image categorization. Our purpose is to extract accurate-enough information from the raw text and image data to be used in the empirical analysis. As long as the measurement errors are of zero mean and are uncorrelated with other independent variables, the estimates of the model will be unbiased and consistent (Greene 2003).

Prior to data coding, we develop the operational definition for all content related variables we will use in the analysis based on previous literature and industry characteristics of our data. Table 1.2 summarizes the definitions for these variables.

Table 1.1. Summary Statistics

Categorical Variables	Levels	Count	Percent	Continuous Variables	Mean	Std Dev
Image	No	12493	73.94%	N(Retweets)	1.19	14.24
	Direct	2953	17.48%	N(Likes)	1.62	15.47
	Link	1450	8.58%	Pixel Fraction Sum	0.34	0.22
Relevant	Yes	3990	90.62%	Log(Followers)	5.97	2.33
Unexpected	Yes	310	7.04%	Klout Score	38.32	17.03
Objects	Face	983	22.33%	Word Count	16.02	5.72
	Happy	513	52.19%			
	Not_happy	470	47.81%			
	Airplane	1218	27.66%			
	Sky	267	6.06%			
	Airport	53	1.20%			
	Aerial	99	2.25%			
	ScreenShot	802	18.21%			
Symmetric		2485	56.44%			
Emoji		1258	7.45%			
Topic Sentiment	Positive Opinion	2639	15.62%			
	Negative Opinion	3217	19.04%			
	Neutral Opinion	6041	35.75%			
	Positive Fact	589	3.49%			
	Negative Fact	1631	9.65%			
	Neutral Fact	2779	16.45%			
Topic	9-11	1054	6.24%			
	Service Failure	1750	10.36%			
	Deal	462	2.73%			
	Entertainment	485	2.87%			
	Others	13145	77.80%			
Brands	AA	5539	32.78%			
	JetBlue	8773	51.92%			
	Frontier	497	2.94%			
	Southwest	1438	8.51%			
	Virgin	665	3.94%			
At		7635	45.19%			
Hashtag		5428	32.13%			
Verified		615	3.64%			

Table 1.2. Coding Categories

Variables	Definition
Imagery Content	
Color Variation	Sum of pixel fraction of top 3 colors. A smaller sum of fraction indicates a larger color variation.
Human Face Presence	Content is coded “1” if at least one human face is detected; otherwise, it is coded “0”.
Happy Face	Content is coded “1” if at least one happy human face is detected; otherwise, it is coded “0”.
Other objects	
Airplane	Content is coded “1” if airplane is detected; otherwise, it is coded “0”.
Sky	Content is coded “1” if sky is detected; otherwise, it is coded “0”.
Airport	Content is coded “1” if the picture is detected to be shot in an airport; otherwise, it is coded “0”.
Aerial	Content is coded “1” if the picture is an aerial photo; otherwise, it is coded “0”.
Screenshot	Content is coded “1” if the picture is a screenshot of a webpage; otherwise, it is coded “0”.
Symmetric Arrangement	Content is coded “1” if the spatial arrangement of objects is symmetric; otherwise, it is coded “0”.
Unexpected Object	Content is coded “1” if the picture contains an unexpected object; otherwise, it is coded “0”.
Relevant Picture	Content is coded “1” if the picture is related to the topic of text component; otherwise, it is coded “0”.
Text Content	
Opinion vs. fact	Content is coded as “opinion” if it contains subjective judgement; it is coded as “fact” if it only contains objective brand information, such as a price or route.
Valence	Content is coded as “positive”, “negative”, or “neutral” based on the over-riding sentiment of the post.
Topics	
9-11	Content is coded as “1” if its content concerns the Sep 11 attacks.
Failure	Content is coded as “1” if its content concerns airline’s service failures such as flight cancellations and delays.
Deal	Content is coded as “1” if its content concerns deals or discounts.
Wifi	Content is coded as “1” if its content concerns on-board Wifi.

1.3.1 Imagery Content Coding

4,403 Tweets (26.06%) in our sample contain imagery content. We combine Google's Cloud Vision API (available at https://cloud.google.com/vision/) with manual coding to categorize our imagery content. Google Cloud Vision API is an easy-to-use API that enables developers to perform a number of tasks on images. According to industry reviews, this API has fast response speed and a reasonably high accuracy rate (e.g., Casalboni 2016, Loeb 2016). We used three functions offered by Google Cloud Vision API. First, we used the "label detection" service, which returns up to 10 objects detected in the image along with confidence scores. As shown in Table 1.2, we control for four airline-industry related object categories that appear most frequently in our data: airplane, sky, airport, and aerial photography. We then utilize the detected labels to determine whether an image contains any objects in these categories.[2] We also utilize the label detection result to determine whether the image is a screenshot or not.[3] Second, we use the "face detection" service to determine whether at least one human face is present in the image and the emotional state if a human face is present. Lastly, the "image attributes" function provided by the API helps us detect the dominant colors of an image in an RGB format along with the fraction of each color in the image. We sum the pixel fraction of the top three colors as our measure of the amount of color variation in an image.

The remaining three image-related variables, i.e., symmetric arrangement, unexpected object, and relevant picture, have to be coded manually. We hired two research assistants who

[2] Specific labels in each object category include 1) Airplane: airplane, aircraft; 2) Sky: sky, cloud, sun; 3) Airport: airport, boarding; 4) Aerial photography: aerial photography, cityscape, bird's eye.
[3] The image is a screenshot if at least one of the following labels is detected: text, font, document, web page, banner, and advertising.

have basic knowledge about photography. They were given a detailed explanation of the variables and were shown several examples to make sure that they fully understood the coding procedures. They then independently coded the imagery content in the 4,403 Tweet posts, and their coding results are consistent in more than 95% of the cases. We hired a third research assistant to examine and adjudicate any discrepancies in the manual coding.

The frequency distribution of all categorical variables including presence of imagery content, objects featured in the picture, symmetric arrangement, unexpected object, relevant picture are reported in the left column of Table 1.1. We report the mean and standard deviation of the top three-color fraction sum in the right column of Table 1.1 (Row 3).

1.3.2 Text Content Coding

Our text content coding task involves dividing the text content of each Tweet into either fact or opinion and determining its valence and topic. A Tweet either states the opinion of the poster or shares a fact about the brand. Similarly, it can only fall into one of three categories in terms of valence - positive, neutral, or negative. Regarding the topic, with Twitter's 140-character limit, it is difficult to introduce multiple topics in a single Tweet. To fulfill this single-category coding task, we utilize the Naïve Bayes classifier (NBC) in the Nature Language Toolkit (NLTK), a suite of libraries for the Python programming language. NBC is a commonly used method to document text classification because it is fast, easy to implement and relatively effective (Lewis 1998). It has also been recently adopted by marketing researchers for text mining purposes (e.g., Tirunillai and Tellis 2012, Kumar et al. 2016). We provide a detailed description of the text mining procedure in the Appendix A.1.

We tested the prediction accuracy of our NBC algorithm using another set of 1,000 hand-labeled Tweets that are randomly selected and the result is presented in Figure A1 in the Appendix A.1.[4] As a whole, the machine learning algorithm achieves a high accuracy rate, especially when used to determine the post's topic and whether it states an opinion or a fact (all being close to 90%). The accuracy rate for valence is relatively lower (76.7%), but they are within acceptable ranges considering the fact that people agree on valence only around 80% of the time (Wilson et al. 2005). The frequency distribution of resulting categorical variables including opinion versus fact, valence, and topic can be found in Table 1.1.

1.4 THE MODEL

Our main interest is to understand how imagery content affects the two types of engagement with Tweets: 1) liking, which is measured by the number of Likes, and 2) sharing, which is measured by the number of Retweets. Since these measures are discrete and non-negative, a count regression (e.g., Poisson Regression, Negative Binomial Regression) is more appropriate than a linear regression. However, we have to extend the standard univariate count regression model to account for two patterns that are present in our data.

First, we have to account for the presence of excess zeros (i.e., zero-inflation) because of the typical low level of responses to Tweets posted by ordinary social media users. Zero responses to a Tweet may be due to the lack of views from followers, or due to the lack of interest to respond after followers see the post. The former case is often termed structural zero in count data analysis. A user must read a post before any further engagement can take place. Therefore, if a Tweet gets

[4] We use only a subset of the testing sample to ensure balanced class distributions when calculating accuracy rate. For example, only 248 Tweets are hand-labeled as "fact" out of the 1,000 Tweets in the testing sample. Therefore, we randomly selected 248 Tweets that are hand-labeled as "opinion" to match and balance the testing sample.

zero views, it will surely receive zero number of Likes and Retweets. However, we do not observe the number of views for Tweets in our data. Instead, we rely on post attributes and account characteristics that may affect the amount of attention received by each Tweet to predict the success of the Tweet in getting at least one view. The attention that a Tweet attracts is first affected by the obtrusiveness of the post. We argue that the obtrusiveness of a Tweet is influenced by the presence of pictures and emojis as well as the length of the Tweet. Emojis contribute to obtrusiveness because they are visually different from text in terms of shape and color, while the number of words in a Tweet determines how much space it occupies on the screen. Pictures have both merits. Therefore, we expect the presence of a picture in a Tweet to increase its obtrusiveness. We also include the heuristic features of the picture, such as its colorfulness, objects featured in the picture and symmetric arrangement in the attention model because these features can be easily captured at first sight and may induce readers to investigate more into the post. While *expectancy* and *relevancy* of the picture may also affect engagement, they are less likely to affect attention because these judgements can only be made conditional on attention. For a similar reason, we do not include variables related to the specific message conveyed by the post, such as whether the message is a fact or an opinion, valence of the message and its topic, because users' knowledge about these variables is also dependent on their attention.

The amount of attention received by a Tweet also depends on the specific audience it targets, which can be captured by whether the At (@) and/or the Hashtag (#) function are used. These two unique features on Twitter allow the poster to address to a specific user or a specific topic in the post. As a result, the addressed user or anyone who follows the topic may pay special attention to the post. Moreover, the phrase following the At (@) or the Hashtag (#) sign are highlighted in a

different color than plain text, thus they also contribute to the post's obtrusiveness. Finally, timing of the post and the account characteristics may affect the amount of attention a Tweet can attract.

More formally, we model the latent attention level received by a Tweet i, Att_i, as follows.

$$\begin{aligned} Att_i = {} & \gamma_0 + \gamma_1 Img\_direct_i + \gamma_2 Img\_link_i + \gamma_3 Color_i + \gamma_4 Face_i + \gamma_5 FaceHp_i \\ & + \gamma_6 Obj_i + \gamma_7 Scr_i + \gamma_8 Sym_i + \gamma_9 Emoji_i + \gamma_{10} Length_i + \gamma_{11} At_i + \gamma_{12} Hashtag_i \\ & + \gamma_{13} Time_i + \gamma_{14} Weekend_i + \gamma_{15} Follower_i + \gamma_{16} Klout_i + \gamma_{17} Verified + \varepsilon_i \end{aligned} \tag{1.1}$$

where Img_direct_i is a dummy variable indicating that Tweet i has imagery content that is directly viewable on the twitter home timeline, while Img_link_i is a dummy variable indicating that Tweet i has imagery content which requires clicking on a link to be viewed. $Color_i$, $Face_i$, $FaceHp_i$, Obj_i, Scr_i, Sym_i are pictorial characteristics defined in Table 1.2. $Emoji_i$, At_i, and $Hashtag_i$ are dummy variables indicating whether Tweet i includes at least one emoji, @ sign, and # sign respectively. $Time_i$ and $Weekend_i$ are dummy variables indicating Tweet i is posted during one of the six 4-hour time period (baseline is 12am to 4am PDT) and weekend. $Follower_i$ is the natural logarithm of the number of followers of the poster. $Klout_i$ is 0.1 × Klout score, a numerical value between 1 and 100 that represents an account's online social influence. $Verified_i$ is a dummy variable indicating whether Tweet i is posted by a verified account, which is used to establish authenticity of identities of key individuals and brands on twitter. The error term ε_i follows iid Gumbel Distribution.

The minimal level of attention that needs to be obtained to warrant at least one view on the Tweet is represented by $Att_{i0} = \alpha_0 + \varepsilon_{i0}$, where α_0 is the deterministic level of attention and ε_{i0} is the iid stochastic error term that follows the Extreme Value Type I distribution. We assume a Tweet gets at least one view if its attention level exceeds the minimal attention level defined above. The probability that Tweet i receives a positive number of views can be expressed as follows:

$$\Pr\left(view_i > 0\right) = \Pr\left(Att_i - Att_0 > 0\right) = \frac{\exp\left(\overline{Att}_i\right)}{1+\exp\left(\overline{Att}_i\right)} \tag{1.2}$$

After a Tweet attracts at least one view, we model the number of Likes and the number of Retweets received by the Tweet using a Bivariate Negative Binomial model. We choose the Bivariate Negative Binomial model for two reasons. First, the two engagement measures jointly reflect the popularity of a Tweet and thus are likely to be correlated. Accounting for the correlation is important for the efficiency of the estimator and the computation of correct standard errors. Second, the variances of both dependent variables in our estimation sample are much larger than the means even after removing observations with zero counts. The presence of excess zeros in count data can lead to over-dispersion, where the variance of the count distribution exceeds its mean (Greene 1994). We choose the Negative Binomial model over the Poisson model due to its flexibility in handling over-dispersion.

There are several ways to introduce correlation among count variables (see Winkelmann (2008) for two alternative methods other than the one used in this paper). For computational convenience, we adopt the Multivariate Poisson-Gamma Mixture Model proposed by Hausman et al. (1984). Let y_{ij} ~ Poisson($\lambda_{ij}u_i$) and u_i follow a gamma distribution with E(u_i) = 1 and Var(u_i) = α. It can be shown that the joint distribution function of $y_i = y_{i1} + y_{i2}$, which denotes the sum of the numbers of Likes and Retweets for Tweet *i*, is of a negative binomial form with the following distribution function

$$f\left(y_i \mid x_i\right) = \frac{\Gamma\left(y_{i1}+y_{i2}+\alpha^{-1}\right)}{y_{i1}!\,y_{i2}!\,\Gamma\left(\alpha^{-1}\right)}\left(\frac{\lambda_{i1}}{\lambda_{i1}+\lambda_{i2}+\alpha^{-1}}\right)^{y_{i1}}\left(\frac{\lambda_{i2}}{\lambda_{i1}+\lambda_{i2}+\alpha^{-1}}\right)^{y_{i2}}\left(\frac{\alpha^{-1}}{\lambda_{i1}+\lambda_{i2}+\alpha^{-1}}\right)^{\alpha^{-1}} \tag{1.3}$$

The marginal distributions of the Bivariate Poisson-Gamma model are Univariate Negative Binomial with $E(y_{ij}) = \lambda_{ij}$ and $Var(y_{ij}) = \lambda_{ij} + \alpha\lambda_{ij}^2$ where λ_{ij}, for $j = 1, 2$, are specified as follows:

$$\begin{aligned}\log\left(\lambda_{ij}\right) = {} & \beta_0{}^j + \beta_{i1}{}^j Img\_direct_i + \beta_{i2}{}^j Img\_link_i + \beta_{i3}{}^j Unexp_i + \beta_{i4}{}^j Relev_i + \beta_{i5}{}^j Face_i \\ & + \beta_{i6}{}^j FaceHp_i + \beta_{i7}{}^j Obj_i + \beta_{i8}{}^j Scr_i + \beta_{i9}{}^j Color_i + \beta_{i10}{}^j Sym_i + \beta_{i11}{}^j Emoji_i \\ & + \beta_{i12}{}^j Opinion\_valence_i + \beta_{i13}{}^j Topic_i + \beta_{i14}{}^j Brand_i + \beta_{i15}{}^j At_i + \beta_{i16}{}^j Hashtag_i \\ & + \beta_{i17}{}^j Length_i + \beta_{i18}{}^j Follower_i + \beta_{i19}{}^j Klout_i + \beta_{i20}{}^j Verified_i\end{aligned} \tag{1.4}$$

where Img_direct_i, Img_link_i, $Face_i$, $FaceHp_i$, Obj_i, Scr_i, $Color_i$, Sym_i, $Emoji_i$, $Length_i$, At_i, $Hashtag_i$, $Follower_i$, $Klout_i$, $Verified_i$ are defined in the same way as in equation 1.2. $Unexp_i$ is a dummy variable indicating the presence of unexpected object(s) in the picture. $Relev_i$ is a dummy variable indicating whether the picture is relevant to the text content. $Opinion\_valence_i$ is a vector of dummy variables indicating whether the text content of the post falls into one of the following six categories: positive opinion, negative opinion, neutral opinion, positive fact, negative fact, and neutral fact, where the last category serves as the baseline. $Topic_i$ is a vector of dummy variables indicating whether the topic of the text content is "9-11", "complaint", "deal", or "in-flight entertainment" (baseline category is "other"). $Brand_i$ is a vector of dummy variables indicating whether a particular airline's name was mentioned in Tweet *i* (baseline category is "American Airlines").

Combining the Logistic model and the Bivariate Negative Binomial model, we get the resulting Bivariate Zero-Inflated Negative Binomial model (BVZINB) is as follows:

$$
\begin{aligned}
& p\left(y_i \mid x_i\right) \\
& = \begin{cases} \Pr\left(view_i = 0\right) + \Pr\left(view_i > 0\right) \times \Pr_{NB}\left(y_{i1} = 0, y_{i2} = 0\right) & if \quad y_{i1} = y_{i2} = 0 \\ \Pr\left(view_i > 0\right) \times \Pr_{NB}\left(y_{i1} = k_{i1}, y_{i2} = k_{i2}\right) & otherwise \end{cases} \\
& = \begin{cases} \dfrac{1}{1+\exp\left(\overline{Att}_i\right)} + \dfrac{\exp\left(\overline{Att}_i\right)}{1+\exp\left(\overline{Att}_i\right)} \left(\dfrac{\alpha^{-1}}{\lambda_{i1} + \lambda_{i2} + \alpha^{-1}} \right)^{\alpha^{-1}} & if \quad y_{i1} = y_{i2} = 0 \\ \dfrac{\exp\left(\overline{Att}_i\right)}{1+\exp\left(\overline{Att}_i\right)} \dfrac{\Gamma\left(k_{i1} + k_{i2} + \alpha^{-1}\right)}{k_{i1}!k_{i2}!\Gamma\left(\alpha^{-1}\right)} \left(\dfrac{\alpha^{-1}}{\lambda_{i1} + \lambda_{i2} + \alpha^{-1}} \right)^{\alpha^{-1}} \prod_{k=1,2} \left(\dfrac{\lambda_{ik}}{\lambda_{i1} + \lambda_{i2} + \alpha^{-1}} \right)^{y_{ik}} & otherwise. \end{cases}
\end{aligned}
\tag{1.5}
$$

We estimate the model described in Equations 1.1-1.5 using a maximum likelihood estimator. The resulting set of parameters θ = {γ, β, α} maximizes the following log likelihood function:

$$
\begin{aligned}
\ell = & \sum_{\{i: y_{i1} = y_{i2} = 0\}} \ln \left[\exp\left(z_i' \gamma\right) + \left(1 + \alpha \sum_{j=1,2} \exp\left(x_i' \beta_j\right)\right)^{-\alpha^{-1}} \right] \\
& + \sum_{\{i: y_{i1} + y_{i2} > 0\}} \left[\sum_{s_0=0}^{y_{i1}+y_{i2}-1} \ln\left(s_0 + \alpha^{-1}\right) - \sum_{s_1=1}^{y_{i1}} \ln s_1 - \sum_{s_2=1}^{y_{i2}} \ln s_2 \right] \\
& + \sum_{\{i: y_{i1} + y_{i2} > 0\}} \left[-\left(y_{i1} + y_{i2} + \alpha^{-1}\right) \ln\left(1 + \alpha \sum_{j=1,2} \exp\left(x_i' \beta_j\right)\right) + \left(y_{i1} + y_{i2}\right) \ln \alpha + \sum_{j=1,2} y_{ij} x_i' \beta_j \right] \\
& - \sum_{i=1}^{N} \ln\left[1 + \exp\left(z_i' \gamma\right)\right].
\end{aligned}
\tag{1.6}
$$

1.5 RESULTS

In this section, we present the model estimates and discuss how we can extract meaningful and managerially relevant inferences from social media data based on the model estimates. Before discussing the details of our results, we first compare the in-sample fit statistics of three successive models to demonstrate the importance of accounting for the imagery content when predicting the number of Likes and Retweets. All three models incorporate the effects of text content variables

and all other control variables, but they differ in the extent to which they take into account the effect(s) of imagery content. Model 1 is our proposed model, which accounts for all three proposed effects of imagery content, i.e., the mere presence effect, the pictorial characteristics effect, and the text-image congruency effect. Model 2 only considers the mere presence effect but does not account for the other two imagery content effects. Model 3 does not consider any effects related to the imagery content. Table 1.3 reports the -2LogLikelihood, AIC, BIC, and in-sample root mean square error of these three models. Our proposed model (Model 1) outperforms the two nested models (Model 2 and Model 3) on all four statistics, suggesting the inclusion of the effects of imagery content significantly improves the model fit.

Table 1.3. Model Comparison Based on Marginal Likelihood and In-Sample Goodness-of-Fit Measures Across Models with Different Components

	-2LL	AIC	BIC	MSE	
				Retweets	Likes
Model 1					
Proposed model	21,112.06	21,301.12	22,027.13	1.84	2.55
Model 2					
Proposed Model without pictorial characteristics and congruency variables	21,651.55	21,778.03	22,264.84	1.87	2.56
Model 3					
Proposed model without any image-related variables	22,049.96	22,164.35	22,604.84	1.86	2.62

1.5.1 Imagery Content

How imagery content affects social media post engagement is the primary interest of our research. In this subsection, we discuss our estimation results regarding three aspects of imagery content: the mere presence effect, the pictorial characteristics effect, and the picture-text congruency effect.

● *The mere presence effect:* We find that the mere presence of directly viewable imagery content significantly increases attention and both types of engagement regardless of its characteristics and relationship to the text content of the post (Table 1.4, Row 2). However, social media users seem to rarely bother to read and engage with a Tweet that contains a hyper-linked image as indicated by the three significantly negative coefficients in Row 3 of Table 1.4. Tweets with hyper-linked images are often created using other photo-focused applications such as Instagram. Although it may be more time efficient for a user to post a hyper-link instead of uploading a picture twice, our results suggest that this extra time is not wasted. A user who posts a Tweet with directly viewable image content is rewarded by more views, Likes, and Retweets.

● *Pictorial Characteristics Effect*: The relationships between specific pictorial characteristics and user engagement with the content suggest that the role of imagery content in the popularity of social media posts is more complex than its mere presence alone. We examine three types of pictorial characteristics, namely colorfulness, arrangement, and objects. We find that pictures with less color variation lead to higher attention level, but yield fewer Likes and Retweets conditional on attention (Table 1.4, Row 4). This finding suggests that although simple monotonic pictures may help to attract attention, people tend to engage with more colorful pictures. Our results also show that the symmetric arrangement of objects in a picture decreases attention, but it does not have any significant impact on the number of Likes and Retweets beyond that (Table 1.4, Row 12). This suggests that while design complexity makes the picture more eye-catching, it does not affect engagement directly probably because it does not change the information value of the post.

Table 1.4. Estimation Results for Tweet Engagement

Variables	Levels	Attention	Likes	Retweets
Intercept		**-3.10**	**-2.45**	**-5.02**
Image Presence	Direct Image	**1.04**	**0.47**	**1.57**
	Linked Image	**-0.58**	**-0.73**	**-0.92**
Colorfulness	Top 3 Color %	**0.64**	**-0.36**	**-1.15**
Objects	Human Face	**0.73**	**0.69**	-0.06
	Happy Face	**0.86**	**-0.32**	**-0.74**
	Airplane	**-1.41**	**-0.16**	**-0.29**
	Sky	-0.03	**0.38**	**0.96**
	Airport	**0.69**	**0.64**	**0.97**
	Aerial Photos	-0.63	0.23	**0.43**
	Screenshots	**-0.66**	**-0.37**	**-0.42**
Symmetric Arrangement		**-0.36**	-0.08	-0.04
Unexpected Object		-----	**0.42**	**0.33**
Relevant Picture		-----	**0.17**	-0.02
Valence	Positive Opinion	-----	**0.37**	-0.01
	Negative Opinion	-----	**-0.16**	**-0.58**
	Neutral Opinion	-----	**-0.07**	**-0.42**
	Positive Fact	-----	-0.04	**0.19**
	Negative Fact	-----	**-0.66**	0.01
Topics	Sept-11 Event	-----	**0.07**	0.03
	Service	-----	**-0.72**	**-0.08**
	Deal	-----	**-1.22**	**-0.81**
	Free Wi-Fi	-----	**-0.15**	**-1.06**
Brands	JetBlue	-----	**-0.61**	**-0.88**
	Frontier	-----	**-0.65**	**-0.78**
	Southwest	-----	**-0.69**	**-0.60**
	Virgin	-----	**-0.48**	**-1.35**
At (@)		**0.10**	**-0.11**	0.00
Hashtag		**0.24**	-0.07	0.07
Emoji		**0.32**	**0.11**	**0.45**
Log of Word Count		-0.05	**0.34**	**-0.32**
Log of Number of Followers		**0.31**	**0.31**	**0.35**
Klout Score		-0.01	**0.18**	**0.45**
Verified Account		**2.31**	-0.03	**-1.06**
Time-of-day fixed effects		✓		
Weekend fixed effects		✓		

*Bold figures: *p*-value < 0.05

The findings regarding the effect of specific objects featured in the image are more nuanced (Table 1.4, Row 5 to 11). We find that pictures featuring sky and airport attract more attention and both types of engagement, but those featuring airplanes or a screen shot lead to lower attention level and discourage both types of engagement. More interestingly, we find that the presence of certain objects can have disparate effects on users' liking and sharing behaviors. For example, while the presence of human faces helps to increase attention and the number of Likes, it does not affect the number of Retweets for the post. A plausible explanation is that the majority of human images in our sample of Tweet posts are personal photos. Although they help to induce intimacy so that the followers may respond by sending a Like, these personal photos may not contain useful information for others and are too private to be shared beyond the original poster's own follower base. Moreover, we find that pictures with happy faces tend to induce significantly fewer Likes and Retweets than pictures containing other facial expressions. This finding may seem to be counterintuitive at the first glance. However, after a closer examination, we discover that the facial expressions classified as "other" expressions in our sample are not necessarily unhappy faces (such as sad or angry faces) but cool and funny faces. Aerial scenes are another type of imagery objects that has varying effects on users' liking and sharing behaviors. Unlike personal photos, these pictures tend to boost sharing, but not liking. Aerial photos tend to have higher aesthetic values than pictures featuring other objects in our setting. We suspect that readers are more likely to share such pictures with others to tout their superior tastes.

- *Congruency Effect*: Our result on congruency effect is consistent with the extant literature (Table 1.4, Row 13 and 14). We find that people prefer pictures with unexpected objects. This is because readers gain extra satisfaction by resolving the tension created by unexpected objects. We

find that image relevancy with the topic only affects the number of Likes, but not the number of Retweets. This result is in line with the prediction of the ELM model because the elaboration level can be different when a reader chooses to like or to share a post. Approval is more personal and often happens between friends to show support or applause. We expect a reader to pay more attention to her friends' posts than posts by others, thus exhibiting a higher elaboration level. However, sharing is mostly driven by the usefulness of the content rather than the relationship between the reader and the poster. Therefore, we expect a relatively lower elaboration level given the large number of Tweets a reader may be exposed to every day.

1.5.2 Text Content

Although not the focus of this research, we also examine how valence and topic(s) of a Tweet post affect user engagement (Table 1.4, Row 19 to 23). In general, we find social media users to rather retweet facts than opinions. However, at the same time, they tend to like opinions more than facts. This is again in line with varying motivations underlying the sharing and liking behavior. Opinions are more personal so they induce personal interactions such as comments and likes. Facts tend to be more informational and useful and thus are more likely to be shared by readers. In terms of valence, our results indicate that positive content increases engagement and negative content decreases engagement compared to neutral content, although the effect also depends on whether the text content is an opinion or a fact.

We control for the topic of the text content. Among the various topics discussed, we find Tweets mentioning the 9-11 event receive more Likes than Tweets of other topics related to the airline industry (Table 1.4, Row 24). We also find that Tweets mentioning air travel deals receive significantly fewer Likes and Retweets than other Tweets in our sample (Table 1.4, Row 26). This

result suggests Twitter may not be the ideal platform for airline companies to advertise flight deals. Tweets about service failures, such as technique issues, flight cancelations and delays, receive fewer Likes than other Tweets because such events are definitely not something to cheer for (Table 1.4, Row 25). Many passengers tweet in flight when there is free Wifi on board (a special complimentary service offered by JetBlue Airways). We find that such Tweets tend to receive fewer Retweets, mainly because they are about personal experiences and therefore uninformative to others (Table 1.4, Row 27). We also control for brand-specific impacts. Our results indicate that Tweets about American Airlines (AA) are most popular compared to other brands in our dataset which is consistent with AA's higher market share and brand equity.

1.5.3 Account Characteristics

We examine three account characteristics of the poster, namely, the number of followers, the Klout score and the account verification status (Table 1.4, Row 32 to 34). As expected, more followers lead to higher attention levels and more Likes and Retweets, simply because the potential audience size is larger. Regarding the effect of influence power of the account, our results show that a higher Klout score increases the number of Likes and Retweets, but not the attention level. Although the exact formula to calculate the Klout score is not revealed by the company, it considers both the account's follower size and the content quality of past posts. Since we already control follower size, the impact of the Klout score is not significant at the attention level. However, a higher Klout score does lead to significantly more Likes and Retweets conditional on attention because it indicates a higher expected quality of content posted by the account. Regarding the account verification status, we find that after controlling for follower size and Klout score, verified accounts still enjoy a higher attention level, but tend to receive fewer Retweets. Since

verified accounts are marked with a blue tick, they are visually different from other account and thus may draw readers' attention. However, liking or sharing decisions are mainly based on the quality of the content, thus the status being a verified account does not necessarily bring extra value to the content.

1.5.4 Quantifying the Imagery Effects

In the previous four subsections, we have reported our findings on the role of imagery content, text content and other control variables on attention, and the two conditional engagement measures separately. In the current subsection, we discuss the net effect of imagery content on the number of Likes and Retweets. We first compute the expected number of Likes and Retweets of a typical Tweet post without any imagery content as the baseline for our comparison.[5] We then compute the percentage change in the expected number of Likes and Retweets due to the inclusion of imagery content to quantify the net effect size of mere presence (for both directly viewable and linked picture), pictorial characteristics and congruency effects (See Figure 1.2).

Our results suggest that having a directly viewable picture, regardless of its characteristics, increase the expected number of Likes and Retweets by 231% and 479% respectively. However, a linked picture decreases the expected number of Likes and Retweets by 71% and 78% respectively. After combining its impact on attention and engagement, we find that the presence of human face always leads to higher number of Likes, but happy faces can decrease the expected number of Retweets. Our results also indicate that among all the objects related to air-traveling,

[5] A typical Tweet without picture is defined as a post with text content and other control variables set to take the mode value (for categorical variables) or the median value (for continuous variables) in our estimation sample. All imagery content variables are set to be zero.

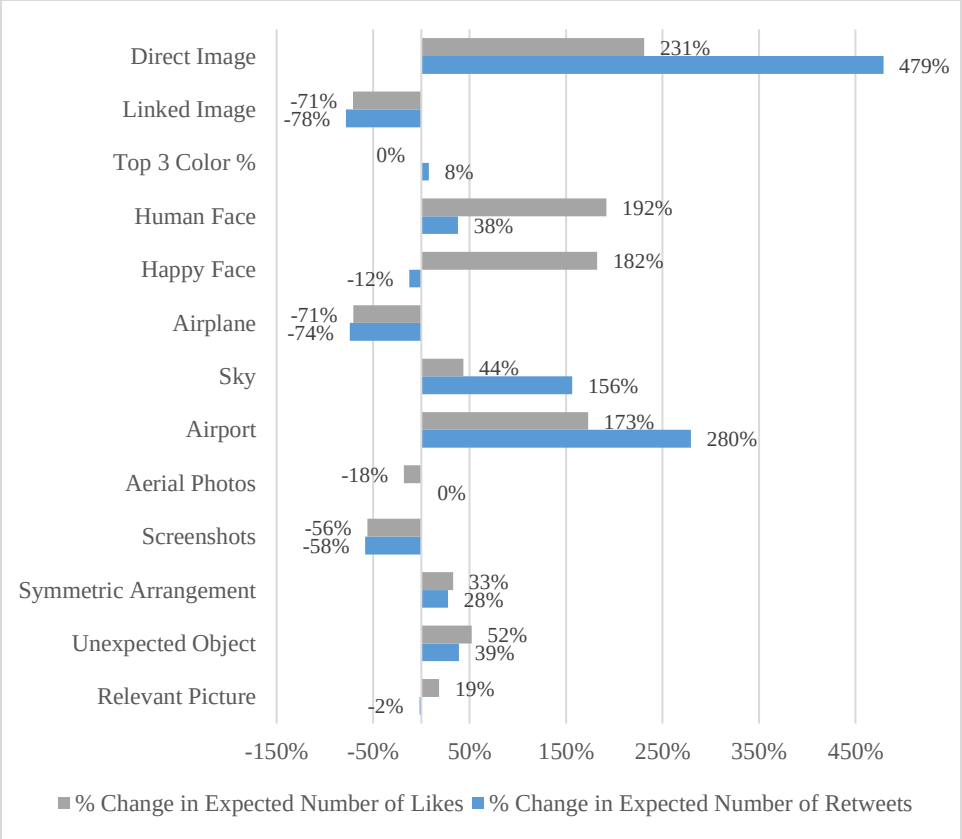


Figure 1.2. Percentage Change in Expected Number of Likes and Retweets

airport pictures are the most effective in boosting the expected number of Likes and Retweets. Regarding congruency, our results suggest that expectancy has a larger impact than relevancy on both types of engagement.

1.6 Endogeneity and Propensity Score Subclassification

The inclusion of imagery content in a Tweet post can be an endogenous decision, which correlates with other factors that also influence user engagement with the post. If endogeneity is present but not accounted for, model estimates will be biased. There are several potential sources for endogeneity in our case. First, it might be easier to find a relevant picture for certain topics

compared to others. For example, one may post a photo featuring a JetBlue airplane next to the gate with a Tweet which says "Waiting to depart for New York @Jetblue." However, it will be much more difficult to find a picture to go with "@Jetblue your call center never picks up the phone." We account for this source of endogeneity in our main model by controlling for text-specific features such as opinion/fact, valence, and topic. Second, it is possible that posters know that imagery content will add to the popularity of a post. Therefore, they may be more likely to attach a picture to a message that has the potential to become popular than other ordinary messages. This is commonly referred to as the self-selection issue. If the decision to add a picture is not independent of the inherent quality of the message, our estimates can be biased. Lastly, the existence of imagery content may be due to unobservable post or account features that drive both the popularity of the Tweets and the propensity to have imagery content in those Tweets. In this section, we describe how we use a propensity score subclassification approach to alleviate the endogeneity concern.

1.6.1 Propensity Score Subclassification

Ideally, we would solve the endogeneity problem using a randomized experiment by randomly assigning the followers of each Tweet into one of two groups: followers in the treatment group are exposed to both text and imagery content, while those in the control group are exposed to only the text content of the same Tweet. We would then be able to draw conclusions on the treatment effect of imagery content on followers' engagement by calculating the differences in the number of Likes and Retweets received by the same Tweet from these two groups. However, with observational data, all followers are either in the treatment group (if the Tweet contains both text and imagery content) or in the control group (if the Tweet only contains text content). As a result, followers in

the treatment group may be systematically different from followers in the control group, thus making the inference about the treatment effect biased.

Propensity score matching is an effective way to adjust for the differences in the treatment and control group, which may bias the inferences about the treatment effect (Rosenaum and Rubin 1983). Propensity score is the predicted probability that an observational unit (a Tweet) receives a treatment (the existence of imagery content) conditional on all observable covariates. Therefore, for observations with propensity scores that are close enough to each other, we can consider the covariates to be independent of the treatment. Thus, the biases in the comparisons between treated and control units are eliminated. The propensity score method is more suitable for settings with binary treatments (Imbens and Wooldridge 2009). Therefore, we will focus on the effect of mere presence of imagery content for the propensity score analysis. We do not differentiate directly viewable pictures from linked pictures in the propensity score analysis because it is difficult to predict whether the poster will post the picture directly or share a picture link based on observable post and poster characteristics. Neither do we control for pictorial characteristics or picture-text congruency because the resulting strata would be too small to yield credible inferences. We argue that controlling for the presence of imagery content is adequate because it is hard to imagine a poster is not strategic when deciding whether to attach a picture or not in the first place but is strategic when deciding which picture to attach.

There are multiple methods to adjust for propensity scores, such as matching, weighting and subclassification, and these adjustment methods can be further combined with regressions to control for the effects of other observable covariates (see Imbens and Wooldridge 2009 for a comprehensive review and discussion about the pros and cons for each method). We adopt a

subclassification approach and combine it with a regression, mainly because matching will significantly reduce our sample size and weighting is not suitable for nonlinear regression models. The idea of subclassification is that we can partition the sample into strata using discretized values of the propensity score. Within each stratum, due to the very small differences in propensity scores, we can view the data as coming from a completely randomized experiment. Then, we can estimate stratum-specific treatment effects by regressing the outcome variable on a constant, an indicator for the treatment, and other covariates within each stratum. Finally, we can calculate the overall treatment effect by averaging the stratum-specific effects weighted by the relative stratum size. Cochran (1968) shows five equal-sized blocks can remove over 95% of the bias in the difference between treatment and control groups, so we classify our sample into five strata.

As discussed earlier, both observable and unobservable factors may lead to endogeneity. Therefore, we include all the text content related covariates that may directly affect engagement when constructing the propensity score. To account for the unobservable factors, we use three instrumental variables: 1) whether the Twitter account is linked to the holder's Instagram account, 2) whether the account holder mentions anything related to photo-enthusiasm in the account self-description, 3) the percentage of Tweets with images in the account history. We argue that these instruments are only related to an account holder's propensity to post a picture in general, but are uncorrelated to the popularity of a specific Tweet post. We use the logit link function to estimate the probability of the presence of a picture as a function of these covariates, and then calculate the propensity score for every Tweet in our sample based on the parameter estimates.

1.6.2 Validation of Propensity Score Subclassification Approach

The model fit statistics show that our model explains a significant proportion of variances in the presence of imagery content (Adjusted R-Square = 0.246). We then assign each observation to one of the five equal-sized strata based on the percentage of its propensity score. We use two methods to verify that the data is balanced within each subclass and treatment is independent of covariates after subclassification. First, we compare the subclass membership across strata. The result shows that strata associated with higher propensity score are composed of more Tweets coming from the treatment group than those from the control group, which verifies the predictive power of our propensity score model. Second, following Guo and Fraser (2009), we use each covariate as an independent variable and the treatment as the dependent variable and run the regression within each stratum to check whether the covariates can still predict treatment. The results show that after subclassification, all covariates are insignificant predictors of the treatment. Therefore, we believe that our method successfully removes the differences in treatment and control groups and makes the data balanced.

1.6.3 Effect of Mere Presence after Propensity Score Subclassification

To get the stratum-specific estimate for the mere presence effect of a picture, we estimate equation (5) using the presence of imagery content, characteristics of the text content, timing of the post, and account characteristics as explanatory variables. The results are presented in Table 1.5. We find that the presence of an image significantly increases the attention level and the two types of engagement in almost all strata. The weighted average estimates for attention, the number of Retweets and the number of Likes are 0.534, 1.228 and 0.560, and all of them are significantly different from zero. These results are consistent with the findings from our proposed model. In

particular, we observe the same pattern that the mere presence effect of imagery content has a relatively stronger influence on the number of Retweets than on the number of Likes. Based on this analysis, we conclude that self-selection of posting a picture is not a significant concern in our sample and our findings regarding the mere presence effect of imagery content are robust.

Table 1.5. Estimates for Mere Presence Effect Using Propensity Score Subclassification

Strata	Attention	Retweets	Likes
1	**0.961**	**1.587**	**0.836**
2	0.162	**1.276**	**0.786**
3	**0.472**	**1.412**	**0.064**
4	**0.493**	**1.289**	0.225
5	**0.582**	**0.575**	**0.890**
Weighted Average	**0.534**	**1.228**	**0.560**

*Bold figures: p-value < 0.05

1.7 Conclusions, Limitations and Future Research

Social media has enabled average Internet users to publish their experiences and opinions online and to let their voice be heard by many others. In extreme cases, a local event can become a viral topic on social media and reach millions of people in just a few hours. Social media users who desire to broaden their reach are interested in learning what factors drive the popularity of social media posts. So far, our knowledge about popularity or virality of user-generated content (UGC) is mainly on the text content. However, with the swift tide shift towards visual social media, it is crucial to expand our understanding how images affect the popularity of UGC in social media.

In the current study, we use observational data to examine the effect of imagery content on readers' engagement with social media posts. We propose three ways through which imagery content exerts influence on engagement – mere presence, pictorial characteristics, and picture-text congruency – and test for the existence of these effects using a dataset containing user

conversations about five airline companies on Twitter.com. Following the industry customs, we define two types of engagement – liking and sharing – and measure them using the number of Likes and Retweets, respectively.

Our findings enrich the current knowledge about UGC popularity. While common wisdom suggests that images help a post to become more popular, we offer a more detailed explanation on how it happens. Our findings confirm that the mere presence of a directly viewable image always helps a Tweet post to receive more Likes and Retweets. However, a linked picture backfires, in that it significantly reduces the number of Likes and Retweets because readers do not bother to click on external links. Therefore, when posting on social media platforms, the poster should always re-upload the pictures in order to facilitate readers' engagement with the content. Moreover, the relationship between imagery content and engagement is more complex than the mere presence effect alone. Pictorial characteristics and picture-text congruency also affect user engagement. Interestingly, we find that pictorial characteristics can have disparate effects on users' liking and sharing. These findings are consistent with previous literature on the differences in users' underlying motivations to like versus to share content: while personal content boosts liking, informational and useful content increases sharing. We also find that relevancy between the text and imagery content in a Tweet affects liking but not sharing, which is in line with the explanation based on Elaboration Likelihood Model.

Although not a focus of our analysis, our study also generates additional insights regarding the impact of text content on UGC popularity in micro-blogging websites. It is an ongoing debate whether positive or negative sentiments drive the popularity of UGC. Our findings suggest that the effect of sentiment may be more nuanced: it depends on the text category and the type of

engagement. Positive opinions increase liking but not sharing, while positive facts increase sharing but not liking. We also find that negative facts decrease liking, but do not affect sharing. These findings can be explained by differences in users' motivation to like versus to share content.

Our study offers important managerial implications for social media practitioners. First, our findings can guide social media marketers in identifying important Tweets when conducting brand related social listening. Many Twitter users utilize the platform to make service complaints. The better identification of potentially influential posts gives these companies an opportunity to take actions before the harmful social buzz becomes viral. Most of the current social listening tools mainly track the total number of brand mentions and the sentiment composition while leaving imagery content aside. Our research suggests that companies should include the presence of imagery content into their social listening metrics. They may also consider using online image processing tools such as Google Cloud Vision API to further analyze the content of the picture and pay special attention to images with certain characteristics, e.g., colorful pictures. Second, our findings also shed light for social media users on how to create posts that are more influential. The ubiquity of mobile phone cameras has enabled social media users to take pictures on the go. They should take advantage of this powerful tool and attach a picture to their posts to boost reach and engagement. Meanwhile, they should be more selective when choosing which picture to attach. Our findings give many clues on what types of pictures increase the liking or sharing of a post.

To the best of our knowledge, this research is the first study that explores the impact of imagery content on UGC popularity in the context of social media engagement. This is an important yet very new research topic that offers plenty of opportunities for future research. First, our empirical study is based on Tweet posts related to airline companies. It will be interesting to

test whether our findings are generalizable to other industries, especially industries that heavily rely on user-generated photos to demonstrate product benefits such as apparel, cosmetics, and food industries. Second, the focal social media platform in our study is Twitter, a general micro-blogging service. Given the quick rise of photo-sharing apps such as Instagram, Pinterest and Flickr, it is important to examine whether the pictorial characteristics and picture-text congruency effect work the same way on these platforms. Third, an in-depth investigation on what motivates social media users to post pictures online would be a nice complement to the current study. Last but not least, a natural extension of this research is to further explore the economic effects of UGC with imagery content, for example, how would it affect readers' purchase intentions or purchase behavior.

CHAPTER 2

IMAGERY CONTENT AND STAR RATING IN ONLINE REVIEWS

2.1. Introduction

Online customer reviews are playing an increasingly important role for both consumers and marketers. According to the Local Consumer Review Survey (BrightLocal 2016), 91% of consumers regularly or occasionally read online reviews, and 84% of people trust online reviews as much as a personal recommendation, even though most of these reviews are posted by total strangers. As a result, the volume and valence of online reviews have been shown to significantly affect business' performance. For example, it has been shown that a one-star increase in Yelp rating can increase the revenue for local restaurants by 5 to 9 percent (Luca 2011).

As smart phones become popular, taking a picture of the business or the product and attach it to the review become much easier than before. Previous research has shown that pictures increase the attention and user engagement on social media (Li and Xie 2017) and are associated with longer business page views and even sales (Yelp Inc. 2012), fewer studies have examined how reviews with photos might be different from those without. If the inclusion of imagery content leads to a higher star rating, marketers should have more incentive to encourage their customers to upload photos and make their business more photo-friendly. On the other hand, review readers may want to discount the rating of those reviews with photos because they might be biased. Therefore, our objective in this research is to quantify the effect of imagery content inclusion on review ratings while controlling for other factors that affect rating. Because the decision of whether to upload photos to accompany text content can be related to star rating, we jointly model both star rating and photo-uploading decision as two separate but related processes.

We estimate our model using data provided by Yelp Data Challenge. The estimation sample contains more than 60,000 reviews on 474 restaurants in Pittsburgh, PA, from May 2005 to January 2017. We consider three factors that may influence individual's decision of uploading photos. First, users may derive image-related utility by posting photos because this makes them look more professional and knowledgeable. Evidence suggests that image-related utility is related to a user's number of followers. Therefore, we include the number of followers a user has and the user's elite status. Second, a user may upload photo to show her effort, thus making the review more informational and persuasive. Another way to show effort is through writing a long review. Therefore, we include the number of words in the text review. Third, the decision of uploading photos can also be affected by the review's topic and underlying sentiment, which we extract using unsupervised latent Dirichlet allocation (LDA; Blei et al. 2003). In the star rating model, we control for topic, and restaurant- and reviewer-specific characteristics and evaluate the impact of the inclusion of photo(s) in the review. The identification of the model is made possible due to a natural experiment conducted by Yelp. On June 13, 2013, Yelp started to match photos and reviews contributed by the same user for the same business and display them together on that business' Yelp page. Before this event, the reviews are text-only. The users must go to the photo section to view the photos. Co-displaying photos and reviews encourage users to post photos, but is unrelated to the rating they have in mind, thus making it a perfect exogenous shock that help us to identify the impact of the inclusion of photo(s) in the review on review valence.

Our results show significant difference in average rating across reviews with photos versus reviews without photos after controlling for the self-selection of uploading photo, underlying sentiment and topics, restaurant attributes and reviewer characteristics. This association might be

attributed to two reasons, both pointing to an upward bias in star rating of reviews with photos. First, it can be explained by the inherent difference in users who love to take photos and share their works on social networks from those who don't. These people enjoy life more and are more optimistic than others. Therefore, their average opinions on things tend to be more positive. Second, consumers have limited brain resources to remember all the details about their experience at the restaurant. Having a photo on hand helps them to retrieve the memory more easily and derive extra utility from memory (Elster and Loewenstein 1992), thus give a higher star rating when writing the review.

In the next section, we discuss how this research contributes to the current literature on online word-of-mouth (WOM). We then describe our data in section 2.3 and detail our proposed bivariate ordered probit model in section 2.4. In section 2.5, we present the model results and highlight the importance of accounting for the variation across reviews with or without photo when deriving insights from business' star ratings. We conclude by discussing the implications of our research for marketers and consumers and suggest future research in this area.

2.2 Literature Review

Research on online word-of-mouth (WOM) has grown rapidly in the past decade. Numerous studies have shown that online WOM can affect purchase decision, product sales and even firm's stock value (e.g., Chevalier and Mayzlin 2006, Luca 2011, Tirunillai and Tellis 2012). In recent years, we have observed a trend that more studies have focused on the text of online content provided (e.g., Archak et al. 2011, Tirunillai and Tellis 2014). However, few studies have examined the imagery content despite evidence of its effects on consumer decision-making. By

modeling consumer's photo uploading decision and the impact of photo on star rating, this study contributes to the online WOM literature in two ways.

First, our research helps to further understand what drives people to engage in WOM and contribute content online. Prior research has identified that a helpful personality, intrinsic satisfaction, approval on past decision, and a desire for social status, among many others, are drivers of WOM (e.g., Sundaram et al. 1998, Gatignon and Robertson 1986). Cheema and Kaikati (2010) study the effect of need for uniqueness on WOM and find that high-uniqueness consumers, compared to low-uniqueness consumers, are less likely to promote publicly consumed products through WOM because this may decrease the uniqueness of their possession. But high uniqueness does not decrease the willingness to generate WOM for privately consumed products. Berger and Schwatz (2011) examine the psychological drivers of WOM and find that products that are publicly visible or cued more receive more WOM both immediately and over time compared to interesting products because they are top of mind.

On the e-WOM side, Moe and Schweidel (2012) find that an individual's posting decision is affected by the ratings environment. Positive ratings environment encourages posting while negative ratings environments discourage posting. Using a field experiment, Toubia and Stephen (2013) found that users are motivated by both intrinsic utility and image-related utility to contribute content to Twitter, but the later has a larger effect on most users. Anderson and Simester (2014) try to consider a strange phenomenon that loyal customers write negative reviews on products that they did not purchase and suggest that these customers are acting as self-appointed brand managers and use online review as a method to give feedback to the firm. The primary focus in prior research is on consumer's decision to engage in WOM. Our focus in this study is on

consumer's decision to upload photo, an option to provide more information given the consumer has already engaged in WOM.

Our study also contributes to the ongoing discussion about review representativeness. Online reviews have a huge effect on consumer decision and can be used as an effective marketing research tool. However, recent research suggests that online reviews can be systematically biased and easily manipulated. Venue format, social influence, consumer innovativeness, and expertise in online review are among the factors that can affect an individual's decision on what to post. Schweidel and Moe (2014) show that people express different sentiment in reviews posted in different venues, emphasizing the importance of differentiating multiple venue formats when conducting social listening. Several studies identify a downward trend in average rating and aim to provide explanation on that. Li and Hitt (2008) posit that the decrease in rating is due to the difference in taste across early adopters who write the initial ratings and the later customers. Godes and Silva (2009) show evidence to support their explanation that the decrease in average rating is due to the difficulty in making accurate purchases as the total number of rating grows. Using experiment, Schlosser (2005) shows that high initial ratings tend to encourage the subsequent posting of negative ratings because posters strive to differentiate their reviews to be perceived as more intelligent (Amabile 1983). Moe and Trusov (2011) verify this finding empirically using a product rating sample from a national retailer's website. However, Muchnik et al (2013) find the opposite. Using a randomized experiment, they find that later customers are more likely to engage in positive herding while negative social influence inspired users to correct manipulated ratings. Moe and Schweidel's (2012) study may explain the disagreement in previous studies. They find that less frequent posters are more prone to positive herding while more active customers tend to

differentiate from existing reviews. Therefore, the evolution of ratings is shaped by the underlying customer base. Initial reviews not only affect the valence of subsequent reviews but also the content discussed in later reviews. Hamilton et al. (2016) find that the content of subsequent discussions are affected by early postings more than the initial query in an online discussion forum. Our study contributes to this stream of literature by examining how existence of photos relates to star rating. As the penetration of smartphone and 4G network increases, taking a photo and uploading it to social networks become easier for consumers. Therefore, it is vital to understand the role of photos in social media, especially online reviews sites.

2.3 Data Description

We obtained our data set from Yelp Dataset Challenge.[6] This data set contains reviews, tips and photos related to businesses in 11 cities around the world. In choosing the sample, we only include restaurants that meet the following two criteria. First, they must have started business before June 13, 2013, the day when Yelp started to show photos and reviews posted by the same user side-by-side. We need this criterion for identification purpose. Second, all these restaurants have more than 50 reviews. This is to ensure enough observations for each restaurant for reliable estimation results. The resulting data sample we used to estimate the model contains 62,310 reviews posted by 24,735 Yelp users on 473 restaurants in Pittsburgh, PA from May 2005 to January 2017. We matched the photos uploaded for each restaurant to the review posted by the same user and identified 8,184 reviews that are accompanied by at least one photo. Table 2.1 shows the distribution of star ratings of our sample and that of all businesses on Yelp.com. The average star rating in our sample is 3.75,

[6] http://www.yelp.com/dataset_challenge

which is very close to the average star rating (3.77) across all businesses including other categories reported by Yelp. This suggests that our data sample is representative.

Table 2.1. Star Rating Distribution

Stars	Freq.	Percent
1	4,913	7.88
2	5,970	9.58
3	9,918	15.92
4	20,344	32.65
5	21,165	33.97

2.3.1 Summary Statistics

The restaurants in our sample varies in terms of price range, type of food served, and attributes such as availability of table service and food delivery, friendliness to kids, etc. The reviewers in our data sample ranges in terms of date to start yelping, elite status and number of fans. Table 2.2 presents the summary statistics of restaurant and user characteristics.

2.3.2 Topic Modeling

We posit that the decision of uploading photo and star rating are jointly determined by the underlying sentiment and topic that the user has in mind, which are reflected the text content of the review. Therefore, we need to make sense of the text and create quantitative measures to include them in the data analysis. We adopt the latent Dirichlet allocation (LDA; Blei et al. 2003) to simultaneously extract the latent topic and valence mentioned in the review text. The LDA technique has been widely adopted in recent marketing literature and has been proved to be flexible, efficient and accurate. It does not require the researches to know the latent dimensions in advance. In addition, it allows for computation of the importance of the extracted topic by the intensity of the reviews on each dimension. We use the Machine Learning for Language Toolket

Table 2.2. Summary Statistics of Restaurant and User Characteristics

Attributes	Levels	Freq.	Percent
Price Range[7]	$	103	21.78
	$$	332	70.19
	$$$	33	6.98
	$$$$	5	1.06
Type of Food	Domestic	236	49.89
	Mideastern	32	6.77
	South American	40	8.46
	Asian	91	19.24
	European	66	13.95
	Others	8	1.69
Good for Kids		351	74.21
Delivery		82	17.34
Table Service		398	84.14
Elite Member		2,896	11.71
Variable		Mean	Std Dev
Yelp_age		1.79	1.39
Number of Fans		3.39	28.87

(Mallet), a Java-based package for statistical nature language processing to model the latent dimension (McCallum 2002). Before ruing the LDA, we preprocess the text by lower-casing each word and eliminating non-English words and stop words because they do not typically have any informational value about the topic or the sentiment we are interested in. One task for researchers when conducting LDA is to determine the number of topics. We have run the model using 5, 10 and 20 topics and find 20 topics to capture the latent topics and valence the best. Table 2.3 reports the percent of each topic, topic keywords and a label that we create for each topic based on the keywords. It is interesting to note that the topics varies in terms of specificity and valence. For

[7] Price range is the approximate cost per person for a meal including one drink, tax and tip. Yelp defines it as follows:
$= under $10
$$= $11-$30
$$$= $31-$60
$$$$= above $61

Table 2.3. Topics Yield by LDA

No.	Percent	Label	Keywords
1	7.77	Sandwich	Sandwich fries sandwiches cheese hot good bread dog slaw beef dogs steak french meat coleslaw order lunch primanti turkey chicken
2	5.26	Mexican	Tacos mexican salsa chips taco burrito chicken good mad mex beans margaritas rice fish food margarita sauce guacamole spicy authentic
3	41.72	Great-overall	Food great service good friendly love staff delicious time back amazing atmosphere recommend nice excellent awesome wait favorite restaurant fantastic
4	30.61	Great-feeling	Time make love eat years find review people back home feel thing places food things menu city eating made makes
5	22.25	Long-wait	Food service minutes table time order asked wait back waitress server told restaurant people waiting manager experience drinks waited hour
6	8.48	Delicious-breakfast	Breakfast brunch coffee eggs pancakes toast french sunday delicious waffles diner bacon waffle wait cream sweet bloody morning potatoes chocolate
7	17.81	Nice-atmosphere	Bar nice seating tables area people room music night dining inside drinks great sit patio atmosphere table loud street space
8	9.30	Seafood and steak	Crab fish shrimp good cake dessert lobster seafood chocolate salmon tuna cream dinner cooked steak ordered cheesecake great ice meal
9	5.76	Italian	Pasta sauce italian bread wine mussels dish restaurant delicious salad good meal gnocchi ravioli cheese byob dessert dinner fresh appetizer
10	27.59	Good-service	Ordered time back good menu decided dinner friend night table meal pretty order nice wanted made enjoyed seated thought waitress
11	7.64	Meat-lover	Meat steak pork potatoes good cooked chicken sauce cheese sandwich beef delicious ribs bbq tender pulled fried perfectly bread flavor
12	37.87	Good-value	Good food menu service pretty price lunch great prices restaurant bit quality nice options small times decent portions lot worth
13	5.28	Burgers	Burger fries burgers cheese good bacon bun burgatory cooked medium onion meat toppings beef brgr great potato shake chips delicious
14	19.30	Bad-taste	Food ordered good bad tasted back bland service taste time dry pretty flavor chicken cold stars give eat meal quality
15	15.52	Beer	Beer bar good selection great beers wings food happy night hour drink game tap menu pretty drinks draft church fun

16	12.32	Wine-and-cocktails	Menu wine dinner restaurant great delicious experience bar excellent drinks night service meal cocktails made food dining special dessert list
17	13.17	Thai	Thai sushi chicken spicy noodles soup rice rolls curry chinese roll tea good sauce pho dishes pad dish fried pork
18	4.37	Pizza	Pizza cheese crust sauce pizzas good pie toppings pepperoni oven slice thin gelato fresh dough slices italian order wings delivery
19	11.41	Salad	Salad chicken cheese soup fresh sauce ordered dressing delicious sweet side bread served grilled lunch red flavor tomato tasty hummus
20	3.25	Asian	Buffet indian food chicken vegan korean vegetarian dishes spicy masala naan rice dish good restaurant curry options tikka delicious india

example, Topic 5 and 14 are negative, whereas topic 3, 4, 6, 7, 10, 12 are positive. Some topics are general, such as topic 3 which is about the restaurant being great overall, whereas some are more specific, such as topic 6 which is about delicious breakfast.

2.3.3 Model-Free Evidence

Before moving to model develop, we show model-free evidence that motivates our research in this section.

Table 2.4 shows the star rating distribution by whether a review is accompanied by at least one photo.

We also observe a significant jump in the number of reviews with photos before and after June 13, 2013, the day that Yelp started to co-display photos and reviews. Figure 2.1 displays how the percent of reviews with at least one photo in each quarter change over time. The event happened in quarter 25.

Table 2.4. Star Rating Frequency (With Photo vs. Without Photo)

Stars	With Photo		Without Photo	
	Freq.	Percent	Freq.	Percent
1	4,700	8.68	213	2.60
2	5,603	10.35	367	4.48
3	8,641	15.96	1,277	15.60
4	17,249	31.87	3,095	37.82
5	17,933	33.13	3,232	39.49
Average Rating	3.70 (1.26)		4.07 (0.98)	

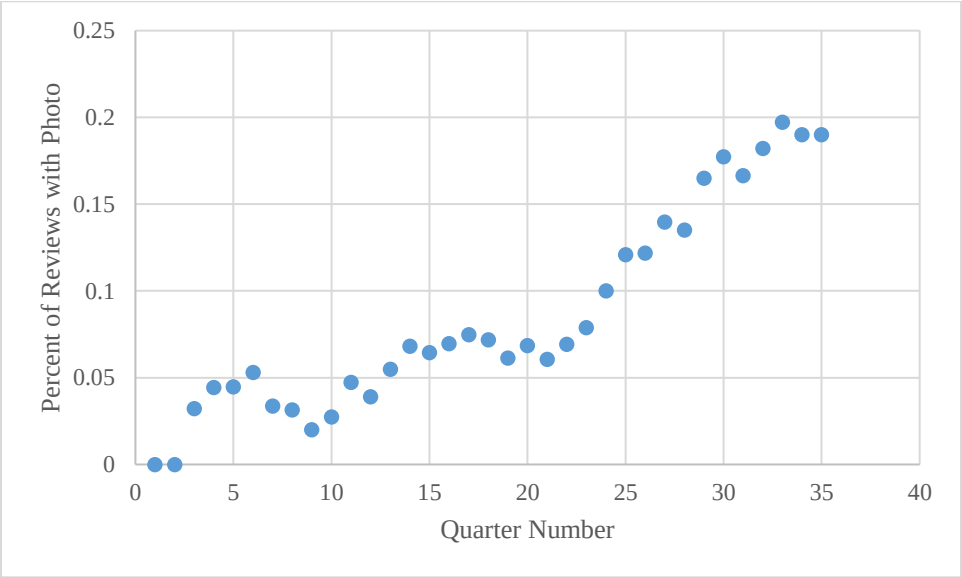


Figure 2.1. Percent of Reviews with Photo over Time

2.4 Model Development

Our modeling objective is to examine how existence of photos relates to star rating. Because the decision of whether to upload photos to accompany text content can be related to star rating, we jointly model both star rating and photo-uploading decision as two separate but related processes.

We are not trying to claim that there is a causal relationship of having photos on the star rating. Instead, we model the two decisions as two interrelated processes arising from the same underlying topic and sentiment that a consumer has in mind.

We adopt the bivariate ordered probit model to allow for correlated error terms of the photo-uploading model and the star rating model. The two models are specified as follows:

The latent attractiveness of posting photo in review *j* is given by:

$$\begin{aligned} PhtAttr_j &= \alpha_1 \cdot \tau_j + \alpha_2 \cdot topic_j + \alpha_3 \cdot length + \alpha_4 \cdot p\_rng_j + \alpha_5 \cdot kid_j + \alpha_6 \cdot table_j \\ &+ \alpha_7 \cdot deliver_j + \alpha_8 \cdot food_j + \alpha_9 \cdot elite_j + \alpha_{10} \cdot fans_j + \alpha_{11} \cdot y\_age_j + \alpha_{12} \cdot qtr_j + \xi_j \end{aligned} \tag{2.1}$$

And the latent opinion driving the star rating in review *j* is given by:

$$\begin{aligned} U_j &= \gamma \cdot PhtAttr_j + \beta_1 \cdot topic_j + \beta_2 \cdot length_j + \beta_3 \cdot p\_rng_j + \beta_4 \cdot kid_j + \beta_5 \cdot table_j \\ &+ \beta_6 \cdot deliver_j + \beta_7 \cdot food_j + \beta_8 \cdot elite_j + \beta 9 \cdot fans_j + \beta_{10} \cdot y\_age_j + \varepsilon_j \end{aligned} \tag{2.2}$$

where ξ_j and ε_j are the error terms that follow bivariate Normal distribution specified in Equation 2.3:

$$\begin{bmatrix} \xi_j \\ \varepsilon_j \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{2.3}$$

Topic$_j$ is a vector that contains the review's loading on the 20 topics resulting from the LDA model with topic 20 serving as the baseline. The details about the 20 topics can be found in Table 2.3. *Length*$_j$ is the logarithm of number of words contained in review *j*. *P_rng*$_j$ is a vector of binary variables that indicate the restaurant's price range with $$$$ serving as the baseline. *Kid*$_j$, *table*$_j$, *deliver*$_j$ are binary variables indicating whether the restaurant is good for parties with kids, offers table service and delivery service. *Food*$_j$ is a vector of binary variables that indicate the type of food served in the restaurant. *Elite*$_j$, *fans*$_j$ and *y_age*$_j$ are user-specific variables, representing

whether the user has been chosen as Yelp elite member in the past, logarithm of number of followers on Yelp and how long s/he has been a member on Yelp when writing the review.

τ_j and qtr_j are two variables exclusive to the photo-uploading model. τ_j is a binary variable indicating whether review j is posted after June 13, 2013. qtr_j is an integer indicating which quarter the review was posted in. It is used to control for the increase in penetration of smartphones. May to July 2005 is defined as quarter 1 and the remaining quarters are defined accordingly.

These two variables are key point to model identification. For the system of bivariate ordered probit model to be identified, the exclusion restriction that at least one explanatory variable in equation 2.1 should not be present in equation 2.2. τ_j serves as a perfect instrumental variable because it is an exogenous shock that should only affect user's photo-uploading decision, but does not have an effect on the star rating. Before June 13, 2013, photos and reviews uploaded by the same user on the same restaurant are displayed in two different sections on the webpage, so the readers are not able to associate the photo and the review. However, starting from June 13, 2013, Yelp matches the photos and review posted by the same user in the review section. We believe this change should significantly increase the user's motivation to upload photos because the co-display of photos may lead to more attention on their review and thus more helpful votes. However, we don't think co-displaying review and photo should change the star rating, which is more likely to be driven by the user's underlying opinion on the restaurant.

The observed pair of outcomes are governed by equation 2.1 and 2.2. More specifically,

$$pht_j = \begin{cases} 0 & if \quad c_{10} < PhtAttr_j \le c_{11} \\ 1 & if \quad c_{11} < PhtAttr_j \le c_{12} \end{cases} \text{ and } star_j = \begin{cases} 1 & if \quad c_{20} < U_j \le c_{21} \\ 2 & if \quad c_{21} < U_j \le c_{22} \\ 3 & if \quad c_{22} < U_j \le c_{23} \\ 4 & if \quad c_{23} < U_j \le c_{24} \\ 5 & if \quad c_{24} < U_j \le c_{25} \end{cases} \tag{2.4}$$

The unknown cutoffs satisfy the condition that $c_{10} < c_{11} < c_{12}$ and $c_{20} < c_{21} < c_{22} < c_{23} < c_{24} < c_{25}$. We define $c_{10} = c_{20} = -\infty$ and $c_{12} = c_{25} = +\infty$ in order to avoid handling the boundary cases separately.

The probability that $pht_j = j$ and $star_j = k$ is:

$$\begin{aligned} \Pr\left(pht_j = i, star_j = k\right) &= \Pr\left(c_i < PhtAttr_j \le c_{1,i+1}, c_{2,k-1} < U_j \le c_{2k}\right) \\ &= \Pr\left(PhtAttr_j \le c_{1,i+1}, U_j \le c_{2k}\right) \\ &\quad - \Pr\left(PhtAttr_j \le c_{1i}, U_j \le c_{2k}\right) \\ &\quad - \Pr\left(PhtAttr_j \le c_{1,i+1}, U_j \le c_{2,k-1}\right) \\ &\quad + \Pr\left(PhtAttr_j \le c_{1i}, U_j \le c_{2,k-1}\right) \end{aligned} \tag{2.5}$$

We estimate the model using bioprobit, a package in Stata which adopts Monte Carlo experiment. The log likelihood function is as follows:

$$\text{Ln}\, L = \sum_{j=1}^{N} \sum_{i=0}^{1} \sum_{k=1}^{5} I\left(pht_j = i, star_j = k\right) \ln \Pr\left(pht_j = i, star_j = k\right) \tag{2.6}$$

2.5 Results

In this section, we present the estimated parameters resulting from our proposed model. Table 2.5 presents the posterior means, standard deviations across iterations of the MCMC sampler and 95% confidence interval for the parameter of interest.

Table 2.5. Photo-Uploading and Star Rating Parameter Estimate

	Photo-uploading				Star Rating			
	Coef.	Std. Err	95% Conf. Int.		Coef.	Std. Err	95% Conf. Int.	
Tao	0.24	0.03	[0.19,	0.30]	----			
Qtr	0.06	0.00	[0.05,	0.06]	----			
Tpc1	0.45	0.14	[0.17,	0.73]	-0.74	0.10	[-0.93,	-0.55]
Tpc2	0.20	0.15	[-0.09,	0.50]	-0.62	0.10	[-0.82,	-0.43]
Tpc3	-0.17	0.13	[-0.42,	0.09]	1.49	0.09	[1.32,	1.66]
Tpc4	-0.46	0.14	[-0.74,	-0.19]	0.05	0.09	[-0.13,	0.23]
Tpc5	-1.07	0.13	[-1.33,	-0.81]	-4.29	0.09	[-4.46,	-4.13]
Tpc6	0.62	0.14	[0.35,	0.89]	0.44	0.09	[0.26,	0.63]
Tpc7	-0.10	0.14	[-0.38,	0.18]	-1.09	0.09	[-1.27,	-0.91]
Tpc8	0.53	0.15	[0.25,	0.82]	-0.15	0.10	[-0.34,	0.05]
Tpc9	-0.04	0.16	[-0.35,	0.27]	0.33	0.11	[0.13,	0.54]
Tpc10	-0.07	0.14	[-0.33,	0.20]	-1.18	0.09	[-1.36,	-1.01]
Tpc11	0.53	0.15	[0.23,	0.82]	0.41	0.10	[0.21,	0.61]
Tpc12	-0.78	0.13	[-1.04,	-0.51]	-1.63	0.09	[-1.80,	-1.46]
Tpc13	0.35	0.15	[0.05,	0.65]	-0.11	0.10	[-0.31,	0.09]
Tpc14	-1.19	0.14	[-1.48,	-0.91]	-6.22	0.09	[-6.41,	-6.04]
Tpc15	-0.39	0.13	[-0.66,	-0.13]	-0.43	0.09	[-0.61,	-0.26]
Tpc16	-0.07	0.15	[-0.36,	0.21]	1.01	0.10	[0.82,	1.20]
Tpc17	0.33	0.13	[0.08,	0.58]	-0.15	0.08	[-0.31,	0.02]
Tpc18	0.12	0.15	[-0.18,	0.42]	-0.33	0.10	[-0.53,	-0.13]
Tpc19	0.34	0.15	[0.06,	0.63]	-0.50	0.10	[-0.69,	-0.30]
Ln_words	0.14	0.01	[0.11,	0.16]	-0.02	0.01	[-0.04,	-0.01]
$	0.14	0.08	[-0.01,	0.30]	-0.44	0.05	[-0.54,	-0.34]
$$	0.16	0.08	[0.01,	0.31]	-0.40	0.05	[-0.50,	-0.30]
$$$	0.10	0.08	[-0.05,	0.26]	-0.33	0.05	[-0.43,	-0.23]
Kid	0.02	0.02	[-0.01,	0.06]	0.06	0.01	[0.03,	0.08]
Deliver	-0.08	0.02	[-0.13,	-0.04]	-0.07	0.01	[-0.10,	-0.04]
Table	0.03	0.02	[-0.02,	0.08]	-0.37	0.02	[-0.40,	-0.34]
Domestic	0.02	0.02	[-0.02,	0.05]	0.04	0.01	[0.02,	0.06]
Mideastern	-0.05	0.03	[-0.12,	0.01]	0.06	0.02	[0.02,	0.10]
SouthAmer	0.01	0.03	[-0.04,	0.07]	0.13	0.02	[0.10,	0.17]
Asian	-0.03	0.02	[-0.08,	0.01]	0.03	0.01	[0.00,	0.05]
European	-0.04	0.02	[-0.08,	0.01]	0.00	0.01	[-0.03,	0.03]
Yelp_age	-0.06	0.00	[-0.07,	-0.05]	0.01	0.00	[0.00,	0.01]
Elite	0.01	0.02	[-0.03,	0.06]	-0.02	0.02	[-0.05,	0.01]
Ln_fans	0.29	0.01	[0.27,	0.30]	-0.05	0.01	[-0.06,	-0.04]
c_{11}	3.72	0.16	[3.40,	4.04]				
c_{21}					-3.55	0.10	[-3.75,	-3.34]
c_{22}					-2.77	0.10	[-2.97,	-2.57]

c_{23}	-1.96	0.10	[-2.17,	-1.76]
c_{24}	-0.81	0.10	[-1.01,	-0.61]
γ	0.09	0.01	[0.07,	0.11]

2.5.1 Photo Uploading Model Results

Consistent with the exploratory analysis presented in Figure 2.1, the likelihood of uploading photos increases over time. However, on top of the time trend, there is an extra boost due to the co-display of photos and reviews. The posterior mean of *ln_words* indicates that the longer the review is, the more likely that the user will also upload a photo. This result is consistent with our expectation that both uploading photo and writing a long review show the review's effort in being helpful and knowledgeable. The posterior mean of *elite* and *ln_fans* indicate that being an elite member and having more followers on the platform will increase the likelihood of uploading photo. This result is consistent with the extant literature that image-related utility motivates noncommercial social media users to contribute content (Toubia and Stephen 2013). Image-related utility assumes that users are motivated by the perceptions of others and is related to status seeking. Therefore, the user is more incentivized to upload photos if she is recognized as an elite member or has more followers on the platform. Topics also affect the likelihood of uploading photo. Our estimates suggest that specific topics regarding food, such as topic 1 – sandwich, topic 6 – delicious breakfast and topic 8 – seafood and steak, are positively associated with photo uploading behavior. In contrast, general topics talking about the feeling about the experience, such as topic 3 – great overall and topic 4 – great feeling, good value (topic 12) and atmosphere (topic 7) of the restaurant are associated with low photo uploading likelihood. This is because it is easier to take a relevant picture of the food, but it is much more difficult to use a photo to express the subjective feeling about the experience. Our estimates also suggest that positive topics are more likely to be associated with photo

uploading behavior compared to negative topics (such as topic 5 – long wait, and topic 14 – bad taste). This indicates that people are more likely to upload photos when they are happy.

The estimates of the restaurant-specific attributes are also consistent with our expectation. Our results suggest that people are less likely to take photos in extremely expensive restaurant because taking a photo in a high-end restaurant may be perceived as rude and improper. Some fancy restaurants even have clear policies that prohibit photos and cell phone use. In contrast, our result suggests that people are more likely to take a photo in kids-friendly restaurants due to the relaxing atmosphere. We also find that people are more likely to take photos in restaurants with table service, but less likely to take photos in restaurants that offer delivery service. In terms of food type, we find that people are more likely to upload photos for restaurants serving domestic or Mexican food compared to Mideastern, Asian or European food.

2.5.2 Ratings Model Results

Our main interest of this study is to find the relationship between uploading photo and star rating. The result shows that after controlling for underlying topics, restaurant attributes and user characteristics, having a photo is positively associated with star rating. We have computed the marginal effect of having photo on each star level by calculating the percentage change in choice probability for each star level from photo = 0 to photo =1 while keeping every remaining variable unchanged. The average marginal effects across all reviews are presented in Table 2.6. Compared to reviews with no photos, those with photos are more likely to have a high star rating (4 stars or 5 stars). We believe the positive association is due to two reasons. First, people to love taking photos and sharing their pictures with others are different from those who do not. Lee et al. (2014) find that extroverts upload photos and update their status more often than introverts. They tend to

be more optimistic and thus view things more positively. Second, many reviews are written a few days after the experience (Chen and Lurie 2013). Since consumers have limited brain resources, photos serve as a vehicle to help them recall what has happened at the restaurant, just like a take-home brochure provided by Ikea to remind the customers about what they have seen in store. In addition, Elster and Loewenstein (1992) prove that consumers may even derive extra utility by retrieving good memories. As a result, they tend to give a higher star rating when writing the review with photos.

Table 2.6. Marginal Effect of Having Photo

	Avg. Marginal Effect	Std. Dev.
1 star	-15.90%	0.083
2 stars	-10.10%	0.081
3 stars	-4.82%	0.078
4 stars	2.05%	0.078
5 stars	12.04%	0.090

Other coefficients in the star rating model are also consistent with our expectation. In terms of topics, we find that negative topics (topic 5 – long wait, and topic 14 – bad taste) are associated with low star ratings, while positive topics such as topic 3 – great overall, topic 4 – great feeling and topic 6 – delicious breakfast are associated with high star ratings. Longer reviews are associated with lower ratings, probably because they are more critical and contain mixing opinions. We also find that high-priced restaurants and kids-friendly restaurants tend to receive higher ratings, and restaurants offering European food tend to receive lower star ratings. In terms of user characteristics, we find that old Yelp users are more generous when rating the restaurants, but elite members and users with large follower base tend to give lower ratings.

2.6 Discussion

Online reviews are an important information source for both consumers and marketers. However, the overall rating could be biased due to various reasons. This study aims to examine the association between uploading photo and star rating on restaurants and understand why users choose to upload photos to accompany text reviews. We find that reviews with photos tend to have higher star ratings compared to those without and photo uploading decision is affected by user's intention to show effort, motivation to gain image-utility and underlying topic and valence of text review.

Our result has important implications for marketers who are present on the online review sites. Traditional wisdom tells us that having photos may raise viewer's attention and provide more information to them. Our result shows that in addition to these benefits, photos may even lift the star rating for a business. Since a high average star rating has been shown to have a positive effect on a business' sales (Luca 2011), encouraging customers to upload photos becomes even more important. There are a number of things that restaurants can do to increase their customer's intention to take and upload photos. For example, they can improve the furnishing, decoration and lighting to make themselves more photo-friendly. They may also consider using beautiful dishes to display the food and arrange and decorate the food nicely. Moreover, they can train their staff to offer more help when their customers want to take photos.

Although we have managed to control for underlying topic and several characteristics specific to restaurants and users, there are still a lot of unobserved factors that may lead to changes in star rating, such as change in management team or user's personality. Moreover, we are only able to propose the possible reasons behind the effects of photo on rating, but not able to test these

mechanisms. Further research could examine other factors that motivate users to upload photos and explore the mechanism behind the effects of photo on rating using proper data or experiment. Further research may also examine how photo and rating contribute to the perceived helpfulness of the review and their economic effect.

CHAPTER 3

MODELING MULTI-CHANNEL ADVERTISING ATTRIBUTION ACROSS COMPETITORS

3.1 Introduction

The advancement of digital technology has enabled firms to reach consumers through a variety of advertising channels such as search engines, email, referral sites, and display ads. The United States digital advertising market is estimated to reach US $72.09 billion by the end of 2016, surpassing the same year TV ad spending for the first time.[8] Consumers have also become more shopping savvy than ever before as many of them visit multiple stores repeatedly while seeking the best deal before making a purchase (Park and Fader 2004). During this process, it is highly likely that consumers are exposed to many ads posted by different advertisers through multiple digital channels at different points in their purchase journeys. It is therefore vital for firms to have a sound understanding of ad attribution and analytics regarding the effects of various types of online advertisement across multi-channels and multi-touch-points (MCMT hereafter).

Determining the value of digital advertising channels has attracted a growing interest recently among researchers in Information Systems (e.g. Ghose and Todri 2015, Xu et al. 2014) and Marketing (e.g. Li and Kannan 2014). The effectiveness of cross-channel advertising was initially assessed with simple heuristic attribution rules. The "last-touch" model, the linear model, the time decay model, and the position-based model are all examples of ad attribution models using heuristic rules (Tornquist and Tradewell 2012). However, these models often do not fully consider

[8] https://www.emarketer.com/Article/US-Digital-Ad-Spending-Surpass-TV-this-Year/1014469, last accessed on 12/13/2016.

consumers' entire purchase funnel (Li and Kannan 2014). To address this issue, researchers have proposed a number of algorithmic attribution models in recent years as richer multi-channel data become available (e.g., Li and Kannan 2014, Abhishek et al. 2014, Xu et al. 2014). Industry leaders such as Visual IQ, Convertro and Google also provide their clients with advanced algorithmic-based models for understanding marketing performance across multiple channels. These data-driven models attempt to go beyond advertising's direct effect on conversion by considering a consumer's entire purchase journey. However, these models predominantly focus on consumer interactions with a single focal firm while leaving out the impact of competitive actions implemented by competing firms, which we view as a critical determinant of consumer conversion.

Our goal is to develop a new cross-channel attribution model that incorporates consumers' both searching and purchasing behaviors across multiple competing online stores within a competitive online shopping environment. The most important distinction between our model and other extant attribution models is that we consider consumers' online store choice decisions across multiple sellers, thus expanding the literature's single-seller scope (e.g. Li and Kannan 2014, Xu et al. 2014) to a multi-seller scope. Consumers are likely to be exposed to various types of digital ads from various competing sellers, and failing to account for consumers' interactions with these sellers may yield biased estimates of ad effectiveness. Advertising can exert two separate effects on a consumer's final purchase decision: converting the consumer from not buying into buying, and influencing the consumer in choosing where to buy. With a single-seller model, we are only able to capture the second effect, thus can underestimate the overall effect of advertising. Moreover, the problem (of not accounting for competitive advertising) exacerbates in a purchase funnel model because competitive advertising not only influences the final conversion stage, but

also the earlier product information search and alternative evaluation stages. It is natural to expect that the marginal impact of different advertising channels on consumers' conversion probabilities varies across competitive firms and across different stages in a purchase funnel. Firms can benefit from having competitive intelligence by considering consumers' interactions with all other firms competing within the same industry.

Toward this goal, we propose an integrated two-stage choice model in which a consumer considers all available websites that offer the relevant product and then decide which ones to visit and which one to make the purchase from. We follow previous literature (e.g. Li and Kannan 2014, Abhishek et al. 2014) and adopt a funnel view of consumer online store choice decisions for two reasons. First, this allows us to capture the heterogeneity in consumers' consideration set which directly impacts their final purchase decisions. Ignoring this source of heterogeneity will lead to biased advertising effect estimates on purchase decisions (Goeree 2008). Second, this model structure differentiates between the impacts of various advertising channels on different purchase funnel stages, allowing us to separate the two advertising effects - the effect on a brand's inclusion within the consideration set and its effect on consumer purchase utility. The resulting inferences therefore provide guidelines for more efficient ad allocation along the purchase decision process.

We estimate this model using a unique individual-level panel dataset that records consumers' interactions in their purchase funnel with all competing websites within the online air booking industry through various online advertising channels including search engines, email, display ads, referral engines and direct channel. Our result shows that the effects of information stock gained through different channels are significant at both the visit stage and purchase stage. However, the size of these effects varies across channels and stages. Specifically, we find that information stock

gained through direct channel is the most effective in driving both visit and purchase decisions, followed by search advertising and email advertising. Display ads and referral engines are less effective on either decision compared to the other channels. With our competitive model, we are able to compute the own- and cross-marginal impacts of various ad channels on each website. Our marginal impact analysis shows that the effectiveness of advertising channels varies widely across competitors, and this is true for all purchase stages considered in our model. The disparate effectiveness of the same advertising channel across competitors highlights the importance of ad content and message strategies in a competitive environment. A firm needs to determine its standing among the competitive pack and revise its strategy accordingly by benchmarking its own ad effectiveness against competitors. Finally, we highlight the importance and necessity of considering competition by comparing the estimated advertising effectiveness and predictive performance derived from the two models: our proposed model and a baseline model where we assume each firm only knows consumers' interactions within its own website. It turns out that the baseline model significantly underestimates the advertising effectiveness and performs worse than our proposed model in predicting purchase of a holdout sample.

To the best of our knowledge, this paper represents the first attempt in modeling multi-channel attribution across competitors. Our primary contribution is that we develop an integrated model that captures individual consumers' interactions with all competing firms at different stages along their online purchase journey. Our competition-centric analysis represents a significant enrichment over extant ad attribution models that typically take a single firm perspective, while neglecting competitive interactions. As companies are attaching more importance to deriving competitive intelligence in ad attribution, there is an increasing availability in cross-competitor ad

clicks and browsing data provided by many marketing analytics companies. For example, both ACNielsen and ComScore constantly track representative consumers' web browsing and purchases through cookies installed on their browsers[9]. Companies can also form strategic alliance and share customer data with each other, e.g. the strategic data sharing agreement between Amazon and Salesforce in 2016[10]. In addition, the advance of the new digital technologies has also made it possible for companies to infer competitive intelligence through social media and online surveys[11]. In light of the increasing availability of data that tracks consumers' MCMT interactions with all competing firms, our study represents a timely precursor to what is to come in the imminent competitive analytics field. Our model will become more and more feasible as the information technology advances. The inferences from our model will provide companies critical competitive intelligence that helps guide their advertising allocation decisions in an online shopping environment that is characterized by intense competition.

The rest of this paper is organized as follows. In §3.2 we review the related research and discuss how our study contributes to the literature. In §3.3 we set up the proposed integrated two-stage model. We then provide an overview of the data used for this study and show model-free evidence of the online shoppers' searching and purchasing patterns in §3.4. In §3.5 we present the estimation results and evaluate the effectiveness of different ad formats implied by these model estimates. In §6, we demonstrate the importance and necessity of modeling competition between

[9] These firms are even able to link consumers' behavior across multiple devices. According to their websites (e.g., http://www.comscore.com/Industries), many leading firms (including those in our analysis) have subscribed to these services.

[10] http://www.nasdaq.com/article/amazon-and-salesforce-extend-strategic-alliance-to-deliver-service-integrations-20161202-00256/amp, last accessed on 12/13/2016.

[11] Tracking customers' social media activities can also help firms to identify their customers' interactions with their competitors. In fact, SAS's social media analytics tool provides such analysis (https://www.sas.com/content/dam/SAS/pl_pl/doc/factsheet/sas-social-media-analytics.pdf) .

websites by comparing our model with a baseline model which only considers consumers' interactions with a focal firm. Finally, we conclude with discussions and directions for future research in §3.7.

3.2 Literature Review

Information Systems and Marketing researchers have long been interested in measuring the effectiveness of different online advertising formats such as search advertising (Agarwal et al. 2011 & 2015, Ghose and Yang 2009) and display ads (e.g., Lewis and Reiley 2014, Rutz and Bucklin 2012, Goldfarb and Tucker 2011). Recently as rich multi-channel data becomes available, multi-channel ad attribution has gained special attention. Table 3.1 provides an overview of the extant research on multi-channel attribution, where we compare the studies along dimensions including the level of data aggregation, product category, ad channels investigated, the consideration of pre-purchase decisions, the consideration of competition and the modeling approach. Some notable recent papers in this stream of literature include Ghose and Todri (2015), Li and Kannan (2014), Xu et al. (2014), Abhishek et al. (2014), and Zantedeschi et al. (2016). Ghose and Todri (2015) study how exposure to various types of display ads (e.g., retargeting, affiliate targeting advertising, etc.) and the duration of exposures affect consumers' active and passive search behaviors and conversion probability using a difference-in-difference (DID) matching estimator. Li and Kannan (2014) model a consumer's consideration of advertising channels, the visit through a given channel and purchase decision jointly; and estimate the carryover and spillover effects of advertising channels. They find that information stock gained through firm-initiated channels significantly increases a consumer's conversion probability, but the incremental impact of paid search channels is smaller than previously estimated. Abhishek et

al. (2014) apply a Hidden Markov Model to uncover the impact of ads at different latent purchase stages. They show that display ads exert a positive impact on early stages but do not increase the conversion probability directly, while search ad effects show up throughout all stages. Xu et al. (2014) report similar findings using a Mutually Exciting Point Process Model in which ad clicks and purchases are modeled as different types of random points in continuous time and earlier points can affect the types of later points. Zantedeschi et al. (2016) conduct randomized field experiments to estimate the short- and long-term effects of multiple online and offline advertising formats on consumer purchases from a single firm.

Most of the above studies suggest that the effect size of ad channels may vary across different stages along the conversion funnel, emphasizing the importance of modeling multiple stages in consumer's purchase funnel. We accordingly adopt the same funnel view and model a consumer's website visit decisions in addition to her final purchase choice. However, the key distinction of our model from these previous studies lies in our explicit consideration of cross-competitor interactions. Table 3.1 shows that all extant studies rely on data from a single focal firm despite the fact that consumers' website visit and purchase decisions are clearly influenced by their interactions with competing firms. Disregarding these interactions will lead to inefficient and biased estimates of advertising channels' effectiveness and therefore suboptimal ad budget allocation of firms. This oversight in prior research motivates our study.

The richness of our model comes not only from the more complete data (which include consumers' interactions with all firms) we have access to, but also from the more accurate identification of the advertising effects in the presence of competition. This is a non-trivial task. By the same token, although we also develop a multi-stage model, our "funnel" view differs from

Li and Kannan (2014) and Ghose and Todri (2015) in that they modeled consumers' visit and purchase decisions concerning a single focal firm through various channels, whereas we are modeling consumers' visit and purchase decisions at different funnel stages across competing firms. Our model framework is therefore closer to the traditional AIDA (Awareness, Interest, Decision and Action) model where consumers consider and evaluate multiple alternatives before a purchase decision is made (Kerin et al. 2014). Abhishek et al. (2014) modeled three latent states and allowed advertising to affect transitions between latent states and conversion probability within a latent state, whereas we are modeling the observed visit and purchase decisions. Although their model has many merits, it is not quite feasible computationally to model the latent states in a competitive setting where consumers' states with each alternative can vary, leading to exponential growth in terms of computation requirement where more than ten (competing) alternatives and multiple states within each alternative need to be modeled.

3.3 Empirical Setting

We obtained a click-stream data set collected by a leading media measurement and analytics US company in 2010. This company manages a large-scale consumer panel that is representative of the U.S. population. The company recruits panelists to install a software meter so that all their PC-based Internet activities across all web entities will be recorded non-intrusively. The click stream data is organized into user sessions which are defined as visits to a sequence of specific websites consecutively without a break of more than 30 minutes. Each session tracks a panelist's ID, domain name, date and time of visit, referral domain name (from which advertising channel can be inferred), browsing activity measured by the number of pages viewed and duration in minutes, as

Table 3.1. Summary of Research on Multi-Channel Attribution

	Individual /aggregate	Data: product category	Ad channels	Funnel View	Dynamic Effect	Competitive Effect	Methodology
Ghose and Todri 2015	Individual	Concealed	1,2,4	Yes	No	No	DID Matching
Shao and Li 2011	Individual	Software	3,5,6,7	No	No	No	Bagged Logit
Li and Kannan 2014	Individual	Hospitality	1,2,4,5,6	Yes	No	No	Hierarchical Bayes model
Xu et al. 2014	Individual	Electronics	3,6,7	Yes	No	No	Mutually exciting point process model
Anderl et al. 2013	Individual	Fashion, luggage, travel	1,2,4,5,6,7	No	No	No	Graph-Based data mining
Anderl 2013	Individual	Fashion	1,2,4,5,6,7	No	No	No	Proportional hazard
Abhishek et al. 2014	Individual	Automobile	3,6	Yes	No	No	Dynamic Hidden Markov model
Zantedeschi et al. 2016	Individual	Specialty retailer	5,8	No	Yes	No	Bayesian Tobit model
Wiesel et al. 2011	Aggregate	Furniture	1,5,7	Yes	No	No	VAR
Haan et al. 2014	Aggregate	5 categories	3,4,5,6,7,8	Yes	No	No	Structural VAR
Demirci et al. 2014	Aggregate	4 categories	1,2,5,6,7,8	No	No	No	Bayesian VAR
Raman et al. 2012	Analytical	NR	2 channels	No	No	No	Time-varying Nerlove-Arrow
Our paper	Individual	Air travel	3,4,5,6	Yes	Yes	Yes	Two-Stage Choice Model

Note: 1) Ad channels: 1- search: paid, 2- search: organic, 3- search: general, 4- referral, 5- email, 6- display, 7- other online, 8- offline
2) NR = Not Relevant

well as the purchase activity including product category, product description, quantity and price paid if a transaction occurs.

The analysis is confined to data that pertain to online flight booking service. We choose this category among more than 60 product categories for the following reasons. First, travel is the largest eCommerce category, comprising of 36% of all B2C eCommerce sales in 2012[12], so we will have a sufficiently large sample to ensure reliable estimation. Second, air tickets tend to be non-differentiated search goods which reduce the possible confounding effect resulting from product-level heterogeneity. The price differentiation across online agents for a given trip at a given time point is also minimal; price is therefore unlikely to play a major role in a consumer's decision regarding which website to purchase from. Third, air tickets are relatively expensive, averaging US $450 per ticket (Li et al. 2014). Since the prices of air tickets are extremely dynamic, savvy consumers act strategically with extensive searches and patience to wait for discounted tickets (Li et al. 2014). Advertising plays an important role in informing and reminding consumers in such an environment, making it an appealing test-bed for our model. These features ensure that consumers will be highly subject to ad exposures and engage in substantial searches for air travel, making this category an ideal setting for our research.

We identify the set of relevant websites from the panelists' purchase records. A website is included in the analysis if at least one purchase in the air travel category occurred on the website during our entire observational period[13]. These websites can be grouped into two types: 1) major

[12] https://www.emarketer.com/Article/Slow-Steady-Continued-Gains-US-Digital-Travel-Sales/1009909, last accessed on 12/13/2016.

[13] There might be other websites that also sell air tickets excluded from our analysis. However, since no air-travel ticket purchase took place on those websites during the entire sample period these websites are likely to be very small players in this category. The exclusion of these websites should therefore not significantly alter the results.

airlines' official websites, including AA.com, Airtran.com, Alaskaair.com, Continental.com, Delta.com, Jetblueairways.com, Nwa.com, Southwest.com, Spiritair.com, United.com, and Usairways.com; 2) major online travel agencies (OTA hereafter), including Expedia.com, Priceline.com, Orbitz.com, Travelocity.com, Cheaptickets.com, Hotwire.com, Americanexpress-travel.com, Sabresonicweb.com, and Wwte1.com. We exclude Southwest.com from our analysis because their tickets are sold only via the company's direct website but not through OTAs. Consumers who purchase from Southwest.com may have completely different consideration set (only consists of Southwest.com) than others (consists of all available websites). Due to computational considerations (to alleviate the curse of dimensionality), we opted to consolidate some smaller airlines and OTAs when constructing the full choice set. For airlines, we separate the top three airlines, Delta Airlines, American Airlines and United Airlines from the rest because they were the top three largest major airlines in the US in year 2010 and were visited more frequently in our data sample than any single remaining airline website as shown in Table 3.2[14]. We accordingly combine the remaining airline websites into one "other airlines" option. By the same token, we group Hotwire.com, Americanexpress-travel.com, Sabresonicweb.com, and Wwte1.com into one "other OTAs" option because they represent a much smaller market share and are visited and purchased from less frequently compared to the top five OTAs, also shown in Table 3.2.

[14] In 2010, Delta , United and American Airlines enplaned 162.6, 145.6 and 104.5 billion passengers respectively, ranking 1, 2 and 4 in the US. The 5th largest airline was US Airways, enplaned only 59.8 million passengers. The 3rd largest airline was Southwest Airlines.

Table 3.2. List of Websites in the Air Travel Category

j	Domain Name	# of Occasions Visited	# of Occasions as Entry Site	# of Purchases
1	Delta.com	2,206	881	234
2	AA.com	2,190	875	190
3	United.com	1,047	336	88
4	Other Airlines	4,530	2,365	745
5	Expedia.com	4,379	1,748	431
6	Priceline.com	3,017	1,038	223
7	Orbitz.com	3,012	857	246
8	Travelocity.com	2,813	857	216
9	Cheaptickets.com	1,658	623	167
10	Other OTAs	1,483	386	50
Sum		26,335*	9,966	2,590

*This is the number of purchase occasions during which the website was visited at least once.

Toward a purchase, a consumer may need to search several websites and visit each website in her choice set multiple times before a purchase decision is made. These visits can occur across different sessions. Therefore, one important yet challenging task when processing data is to identify and group related sessions pertaining to a purchase occasion. There is no universal rule for grouping such sessions. We use a seven-day window policy which is also adopted by De los Santos et al. (2012) because one week is considered sufficient to capture all relevant visits for a purchase on Internet.[15] Another problem we are facing is that all the OTAs also sell other products such as hotel rooms, car rentals, vacation packages and cruises in addition to air tickets. Yet we do

[15] The validity of the 7-day window policy is verified though our personal communication with a Sabre executive who is in charge of the data analytics division. According to him, majority of customers start searching for a flight ticket three days before making the purchase. Saber is the leading online flight reservation platform in the US that powers OTAs and airlines' ticket booking. We have also conducted robustness checks by varying the window length to 15 days and 30 days and re-estimated the model. We find that our results are qualitatively unchanged. These results are available from the authors upon request.

not directly observe what product(s) the consumer is searching information for unless a purchase is made; for non-purchasing sessions, we do not have data on the specific contents of a visit such as what the consumer saw at a website.

Since we include non-purchase sessions in our analysis, we need to preprocess the data with caution to determine which search sessions are relevant to flight booking. We take several steps to ensure a clean data preprocessing. We first assume that if a flight ticket is purchased at the end of a seven-day window, then all preceding site-sessions are relevant for searching flight-related information. Second, if no purchase is made in the end, we eliminate irrelevant sessions and purchase occasions as follows: 1) if the session is generated within seven days *after* a fight ticket is booked then the session is considered irrelevant and thus deleted. This is because very few consumers would purchase different tickets within a short time frame and so these visits occurred after the purchase are more likely for after-purchase reassurance and confirmation purposes, rather than looking for information about a new purchase; 2) if less than 10 pages are browsed in total across all websites combined during the entire 7-day window, we assume it is too short for any serious product-related search to take place and thus such sessions are dropped; 3) if any website outside of the set of relevant websites as we have defined is visited and at the same time no airline website (which only sold flight tickets in 2010) is visited during a seven-day window, we deem such a session to be irrelevant because the consumer is more likely to be searching for other products, not flight tickets. The above steps yield a data sample containing 18,936 sessions, out of which 4,897 occasions ended with a flight ticket purchase. We then use the first three months (January to March 2010) to construct the historical variables for each individual user, and the next six months (April to September 2010) as the estimation sample to calibrate the model, and the last

three months (October to December 2010) as the out-of-sample to conduct predictive analysis. Our final estimation sample consists of 9,966 purchase occasions with 2,590 of them ended up with a purchase. Table 3.2 presents the complete list of websites included in our analysis and their respective frequencies of being searched, being chosen as the entry website as well as being the website to make the purchase in the estimation sample.

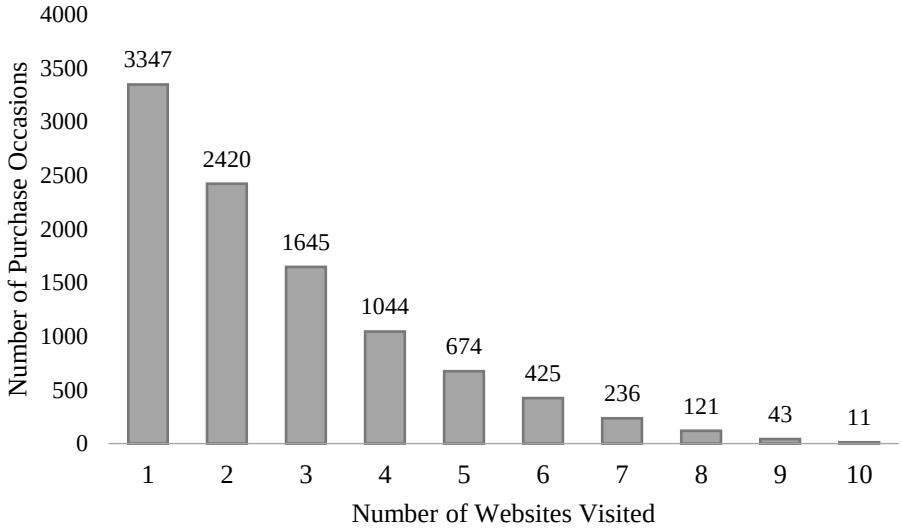


Figure 3.1. Frequency Distribution of the Number of Websites Visited

We observe that consumers generally only search a small number of websites before making purchases. Figure 3.1 plots the histogram for the number of websites visited during a purchase occasion, with 33.58% of consumers visiting only one website. We also observe that 1,528 transactions are made on the first website visited during the seven-day window, representing 58.99% of all transactions taking place in our estimation sample. These statistics lend support to the importance of modeling entry site choice as a distinct decision in the purchase funnel for the online air booking industry.

Table 3.3. Top Redirecting Websites in Each Marketing Channel

Rank	Search Engine	Display/Referral Engine	Email
1	google.com	kayak.com	live.com
2	yahoo.com	tripadvisor.com	comcast.net
3	aol.com	bookingbuddy.com	conduit.com
4	bing.com	lowfares.com	verizon.net
5	mywebsearch.com	doubleclick.net	juno.com

We identified three types of marketing channels: 1) search engines in which the consumer enters a keyword and redirects to the website by clicking a search result[16]; 2) display ads or referral where a third-party website runs banner (or video) ads or provide referral links (for example, referral engines such as TripAdvisor.com and Kayak.com)[17]; 3) email engines from which the consumer is directed to the website by clicking a link embedded in the email. In case that the referral domain name is absent when a consumer types the URL directly into the browser or locate the site through bookmarks, we label the referring channel for this session as "direct". The top five referral domains in each channel are presented in Table 3.3. We observe website visits from all referral domains at the individual level, and for every user we construct ad stock by weighting ad clicks by the number of pages viewed. We present summary statistics regarding the number of visits to each website through a specific channel in a purchase occasion in Table 3.4. Most consumers visit the websites directly. Among advertising channels, search engine is the most popular one, followed by display/referral ads. We also notice significant differences across websites in terms of different ad channel usage, suggesting the presence of large variations in ad budgets and allocations across competitors in this industry.

[16] We cannot differentiate between organic versus sponsored searches due to data limitation though.
[17] We do not further differentiate display ads versus referral engines because many referral engines also run display ads.

Table 3.4. Channel Usage (Conditional on Visit)

Alternatives	# Obs	Variable	Mean	Std Dev	Min	Max
Delta.com	2206	Search	0.65	1.25	0	17
		Display/Referral	0.08	0.32	0	4
		Email	0.06	0.36	0	9
		Direct	1.41	2.48	0	37
AA.com	2190	Search	0.52	0.90	0	14
		Display/Referral	0.10	0.33	0	3
		Email	0.06	0.30	0	4
		Direct	1.31	1.89	0	25
United.com	1047	Search	0.50	0.80	0	8
		Display/Referral	0.15	0.47	0	4
		Email	0.03	0.18	0	3
		Direct	1.24	2.33	0	28
Other Airlines	4530	Search	0.81	1.50	0	21
		Display/Referral	0.15	0.50	0	8
		Email	0.07	0.32	0	4
		Direct	1.65	2.72	0	46
Expedia.com	4379	Search	0.44	0.93	0	17
		Display/Referral	0.24	0.63	0	9
		Email	0.05	0.31	0	6
		Direct	1.52	2.73	0	45
Priceline.com	3017	Search	0.38	0.77	0	9
		Display/Referral	0.24	0.55	0	6
		Email	0.05	0.28	0	5
		Direct	1.23	1.47	0	17
Orbitz.com	3012	Search	0.32	0.79	0	20
		Display/Referral	0.35	0.78	0	13
		Email	0.02	0.17	0	4
		Direct	1.19	1.57	0	20
Travelocity.com	2813	Search	0.36	0.79	0	13
		Display/Referral	0.13	0.38	0	3
		Email	0.03	0.20	0	4
		Direct	1.31	1.43	0	18
Cheaptickets.com	1658	Search	0.44	0.89	0	15
		Display/Referral	0.12	0.38	0	3
		Email	0.03	0.23	0	5
		Direct	1.17	1.31	0	12
Other OTAs	1483	Search	0.31	0.71	0	9
		Display/Referral	0.28	0.56	0	4
		Email	0.04	0.27	0	4
		Direct	1.03	1.21	0	10

3.4 The Model

Our model adopts the purchase funnel view in the context of online shopping of high-involvement products within a competitive environment. The funnel view is grounded in the information processing theory describing the path through which consumers make their purchase decisions, from perceiving a need to post-purchase behaviors (Bettman et al., 1998). A consumer's decision process is metaphorically referred to as a funnel because the number of alternatives considered diminishes as she moves along the purchase journey.

The purchase funnel view can be applied to both product (or brand) choices and store (or website) choices. We apply it to the second scenario and study how online multi-channel advertising affects consumers' decision to visit and purchase from available websites. We specifically model a two-stage choice process in which consumers first select a subset of relevant websites to search for product information and then choose a single website from the subset to make the purchase or choose the outside option of no purchase.

We allow the effect of the information stock (accumulated through different advertising channels) of a consumer to vary across different funnel stages. Previous literature on multichannel marketing supports this assumption. For example, Abhishek et al. (2014) show that display ads tend to be more effective during the early stage in a purchase funnel; while the impact of email, search, and referral is more pronounced in later stages. Li and Kannan (2014) find that consumer-initiated channels (paid search, referral and direct) are more effective in reducing consumer search cost in early stages, while firm-initiated channels (email and display) contribute more at later stages. However, in contrast to most extant ad attribution models that focus on choices within a focal website only (for example, in-store product choices in Abhishek et al. 2014 and ad channel

choices in Li and Kannan 2014), our primary interest is to model consumer choices across competing websites at different stages of a purchase funnel. Next, we describe the details of each stage.

3.4.1 Stage 1: The Website Visit Decision

The objective of the Stage 1 model is to capture and represent consumer search process in terms of: 1) at which website to start the information search and 2) what other websites (if any) to visit after the initial search. A consumer's decision to visit a specific website depends on her perceived benefits derived from the visit relative to the costs incurred due to the visit. The perceived benefits depend on her intrinsic preference for the website and her goodwill towards the website built through various online and offline advertising channels. The costs are associated with the effort required to find the necessary product information and make purchase on the website (Shugan 1980). Hence, consumer *i*'s perceived utility of visiting website *j* can be expressed as:

$$U_{ijt} = \overline{U}_{ijt} + \varepsilon_{ijt} = \alpha_{ij} + G_{ijt} - S_{ijt} + \varepsilon_{ijt} \tag{3.1}$$

$$\text{where } G_{ijt} = \sum_{q=1}^{Q} \alpha_q \sum_{h=1}^{m_{ijt}} d_{ijqh} \times \left(1-\lambda_1\right)^{\left(m_{ijt}-h\right)} \tag{3.2}$$

$$S_{ijt} = \alpha_v v_{ij,t-1} + \alpha_s s_{ijt} + \sum_{c=1}^{C} \alpha_c E_{ijct} \tag{3.3}$$

$$E_{ijct} = \sum_{h=1}^{m_{ijt}} g_{ijch} \times \left(1-\lambda_2\right)^{\left(m_{ijt}-h\right)} \tag{3.4}$$

In the above equations, α_{ij} captures consumer *i*'s intrinsic net preference for visiting website *j* and can vary over people but stay constant over choice situations for each person (Erdem and Keane 1996), and ε_{ijt} follows a generalized extreme value distribution. The term G_{ijt} detailed in

Equation 3.2 represents the consumer's goodwill or ad-stock accumulated towards the website through exposures to various firm-initiated advertising formats (Nerlove and Arrow 1962). Ad exposures to channel q in each period h are approximated using the website's national and local ad spending in that period, d_{ijqh}. We will provide details on the ad expenditure data in Section 3.4.3. The informational effect of pervious exposures decays at a monthly discount rate λ_1, according to the elapsed number of months (m_{ijt}-h). The information channel-specific goodwill can take many possible functional forms; we adopt a parsimonious one that assumes a linear functional form, computed as a linear function of the decayed effects of a website's investment in various advertising channels. However, we must note that which functional form to use is nonessential to our model and our model can be readily adapted to incorporate other specifications for the latent goodwill accumulation process such as a nonlinear decay model.

The term S_{ijt} detailed in Equation 3.3 represents consumer i's cost of visiting website j. The cost is non-zero, so an intercept is needed in the cost function. However, we would not be able to separately identify the cost intercept and the website-specific preference. So, the term α_{ij} in Equation 1 denotes the composite intercept of consumer i with respect to site j, and represents i's "net preference" on j. Search cost usually reduces as a consumer gains familiarity with the website through previous visits and purchases (Li and Kannan 2014), exhibited as reduced time in finding relevant information and a faster checkout process (Shugan 1980). The dummy term $v_{ij,t\text{-}1}$ indexes lagged visit, indicating whether or not the consumer visited the same website in the preceding purchase occasion. The term s_{ijt} represents the logarithm of the consumer's cumulative spending on the website. The term E_{ijct} captures a channel-specific information stock accumulated through navigating the website in the previous purchase occasions, which is calculated according to

Equation 3.4. The term g_{ijch} denotes the logarithm of the number of pages viewed on website j coming through channel c in purchase occasion h. We then compute the channel-specific information stock as the sum of decayed page-views at a monthly discount rate of λ_2, according to the elapsed number of months (m_{ijt}-h). When coefficients α_v, α_s and α_c are negative, their associated variables reduce the cost of visiting the website.

The decision regarding which website to visit first is critical. The search theory asserts that a consumer begins her search process with the highest-utility alternative (Weitzman 1979). The conditional probability of observing website F_{it} being chosen as the entry site by consumer i at time t given the individual-specific preference is therefore expressed as:

$$P\left(F_{it}\middle|\alpha_i\right)=\prod_{j=1}^{J}\left[\frac{e^{\alpha_{ij}+G_{ijt}-S_{ijt}}}{\sum_{k=1}^{J}e^{\alpha_{ik}+G_{ikt}-S_{ikt}}}\right]^{f_{ijt}}\left[\frac{1}{\sum_{k=1}^{J}e^{\alpha_{ij}+G_{ikt}-S_{ikt}}}\right]^{1-f_{ijt}} \tag{3.5}$$

where f_{ijt} =1 if website j is chosen as the entry site and f_{ijt} = 0 otherwise.

Ideally, we would like to model all the subsequent website visits of the consumer one by one sequentially. However, the decision of which website to visit subsequently may depend on the consumer's interactions with the prior website(s). Such dependency can be complicated and arbitrary that can render the model intractable. Moreover Figure 3.2 shows that the number of observations for later visits (3rd, 4th, 5th, etc.) is relatively small in our estimation sample, prohibiting reliable estimates even if one would like to factor in such dependencies. In compromise, we simplify the search decision for the remaining websites within all available websites as a number of independent binary choices. A consumer will only visit a website if the perceived value of that visit is positive. Given the entry site F_{it}, and her individual net preferences,

the conditional probability of observing a consumer's choice set V_{it}, a combination of any number of websites out of J websites, is given as:

$$P\left(V_{it}\left|F_{it},\beta_i\right.\right)=\prod_{\substack{j=1,\\ j\neq F_{it}}}^{J}\left[\frac{e^{\beta_{ij}+G_{ijt}-S_{ijt}}}{1+e^{\beta_{ij}+G_{ijt}-S_{ijt}}}\right]^{v_{ijt}}\left[\frac{1}{1+e^{\beta_{ij}+G_{ijt}-S_{ijt}}}\right]^{1-v_{ijt}} \tag{3.6}$$

where v_{ijt} = 1 if website j is visited in the current purchase occasion and v_{ijt} = 0 otherwise.

3.4.2 Stage 2: The Purchase Decision

The purchase stage can be conceptualized as an option to purchase from a set of websites a consumer has already visited, V_{it}. Since we directly observe the entire search process made by each and every consumer, we are able to capture their respective choice set without any uncertainty. The consumer can also exercise the option of not buying if none of her searched alternatives yields a utility exceeding her purchasing threshold (which is normalized to zero). The utility of purchasing from website $j\in V_{it}$ can be written as:

$$W_{ijt}=\overline{W}_{ijt}+\zeta_{ijt}=\gamma_{ij}+\sum_{c=1}^{C}\gamma_c g_{ijct}+z_{ijt}\gamma+\zeta_{ijt} \tag{3.7}$$

$$\begin{aligned} z_{ijt}\gamma=\;&\gamma_{nn}\times FLN_{jt}+\gamma_{lc}\times FLL_{ijt}+\gamma_f\times FST_{ijt}+\gamma_l\times LST_{ijt}+\gamma_{np}\times LNP_{it}+\gamma_p\times LP_{ijt}\\ &+\gamma_v\times LV_{ijt}+\gamma_{cs}\times CS_{ijt}+\gamma_{cp}\times CP_{ijt}\end{aligned} \tag{3.8}$$

where γ_{ij} captures consumer i's intrinsic net preference for purchasing from website j and can vary over people but being constant over choice situations for each person (Erdem and Keane 1996), and ζ_{ijt} follows a generalized extreme value distribution. The term g_{ijct} is the cumulative informational stock derived from channel c. It is operationalized as the logarithm of the number of pages viewed through channel c in the current purchase occasion. These channel-specific

information stock terms are likely to be correlated with the error term ζ_{ijt} because the unobserved factors that drive purchase decision may also lead to more page views. We will discuss how we handle such endogeneity issues next in Section 3.4.3. The vector z_{ijt} includes other covariates that may affect consumer *i*'s purchase utility, which we specify in Equation 8[18]. The coefficients γ_{nn} and γ_{lc} capture the effect of website *j*'s offline advertising spending at the national and local level (where consumer *i* resided) at purchase occasion *t*, denoted by FLN_{lt} and FLL_{ilt}, respectively. The coefficients γ_f and γ_l capture the effect of being the first and last website visited during the current occasion, represented by two dummy variables FST_{ilt} and LST_{ilt}, respectively. The coefficients γ_{np} ,γ_p and γ_v measure the effect of visit and purchase decisions in the last purchase occasion. LNP_{it} is a dummy variable that indicates whether the last purchase occasion ended with a purchase, LP_{ijt} is a dummy variable that indexes whether the purchase made in the last purchase occasion was on website *j*, and LV_{ijt} is a dummy variable indicating whether website *j* was visited in the last purchase occasion. Finally, the coefficients γ_{cs} and γ_{cp} capture the effects of past spending and browsing history with website *j*, represented by two continuous stock variables CS_{ijt}, the cumulative amount of money spent on website *j*, and CP_{ijt}, the cumulative number of pages browsed on websites *j*.

[18] Some other factors may also affect consumer purchase decision, including but not limited to the product availability (not every carrier/time/stop combination is available at all websites), price information (for non-purchase sessions as well as non-chosen alternative for purchase sessions), and how far in advance the ticket is being purchase. We acknowledge these as a potential limitation of our study. These concerns can be addressed if screen-shot data for webpages visited is available. However, these variables and advertising are not likely to be correlated with each other, as advertising mainly serves as reminder in the airline industry. Omitting these variables is thus unlikely to affect the estimates for multi-channel ad stock.

The consumer also has the option of not buying and the utility of this outside option is normalized to 0. Taken together, the probability of observing website B_{it} being chosen as the conversion site by consumer i at time t conditional on entry site, choice set and individual-specific preferences is as follows:

$$P\left(B_{it}\left|F_{it},V_{it},\gamma_i\right.\right)=\prod_{j\in V_{it}}\left[\frac{e^{\gamma_{ij}+\sum_{c=1}^{C}\gamma_c g_{ijct}+z_{ijt}\gamma}}{1+\sum_{k\in V_{it}}e^{\gamma_{ik}+\sum_{c=1}^{C}\gamma_c g_{ikct}+z_{ikt}\gamma}}\right]^{b_{ijt}}\left[\frac{1}{1+\sum_{k\in V_{it}}e^{\gamma_{ik}+\sum_{c=1}^{C}\gamma_c g_{ikct}+z_{ikt}\gamma}}\right]^{1-b_{ijt}} \tag{3.9}$$

where b_{ijt} = 1 if the consumer converts on website j and b_{ijt} = 0 otherwise.

3.4.3 Endogeneity and the Instruments

Endogeneity in the context of measuring advertising effectiveness is a well-known concern for studies using observational data (Rossi 2014). In our case, the channel-specific informational stocks are likely to be correlated with the error term in the purchase stage and thus pose challenge to identification. Field experiments are ultimate arbitrator for causal inferences (Johnson et al. 2015). However, in our situation where many competing firms are involved, conducting a large-scale experiment is close to impossible because of the difficulty in controlling all competing firms' advertising actions. Alternatively, we adopt the control function approach in discrete choice models popularized by Petrin and Train (2010). The idea behind this approach is to partition the variation in the endogenous variable into two parts: one that is exogenous and the other that is potentially correlated with the error term. Then by including some extra instrument variables that contain the second part in the utility function, we can condition out the variation in the unobserved factor that is not independent of the endogenous variable. This IV approach has been proven to be effective in dealing with endogeneity when randomized experiment is not available, with the

premise that the instruments need to be valid and strong (Rossi 2014)[19] as weak instruments may cause more problems than they solve. We elaborate below how our proposed instruments help mitigate the endogeneity concerns coming from multiple potential sources.

The first potential source of endogeneity is *customer selection*, which happens when these consumers who are exposed to and interact with advertising are systematically different than the others who are not (Manchanda et al. 2004). This can occur for any marketing channels in our context, even for the direct channel. For example, consumers who are more loyal to a website are more likely to accumulate information by visiting the website directly more frequently than other people and to make purchases on this website eventually. This could also happen for firm-initiated marketing channels such as search advertising, email campaigns and display ads if the companies select customers to target for a specific campaign based on certain information such as past browsing or purchase history, responsiveness to ad campaigns, or demographics. If this is the case, the estimates for advertising effectiveness will be biased upward because these selected customers have a higher propensity to convert inherently. The second potential endogeneity source can arise when the unobserved ad impressions simultaneously drive advertising click and purchase intention. The third potential endogeneity source can come from the temporal simultaneity between advertising and unobserved demand shocks (Zantedeschi et al. 2016). It can happen at both the market level and the individual level. At the market level, firms may increase advertising spending when there is an external demand shock unobservable to researchers. As a result, consumers are exposed to more ads leading to simultaneous increases both in terms of ad clicks

[19] Validity means that the instruments must be correlated with the endogenous variable, but does not affect the purchase utility in a direct way. Strength means that the instruments must be able to explain a large enough portion of the variation in the endogenous variable.

and purchases. The individual level demand shock can occur due to factors such as a sudden change in her income, a large bonus etc.

We devised several instrument variables to alleviate the aforementioned endogeneity concerns. We first construct an instrument variable using the consumer's lagged channel-specific information stock. We expect the cost of visiting through a specific channel to decrease with information stock accumulated through the channel in the past, and thus, the consumer may visit the website and accumulate more information through the same channel in the current purchase occasion as well. However, this lagged variable should not be systematically co-determined by the future purchase intention.

We further acquired an external data source from the Kantar Media Database (KMD) to address endogeneity. This database tracks all the companies in our analysis on their monthly spending on offline advertising channels including TV, newspaper and radio, as well as two major online advertising channels, search advertising and display ads at both the national and local DMA (Designated Market Area) level. This supply-side ad data enables us to directly control ad impression. The ad spending reflects potential unobserved demand shocks and inclusion of ad spending would directly control for the increase in ad impression due to unobserved random shocks.

We then construct several additional instrument variables using the KMD data. For each firm in a DMA in a month, we compute its competitors' online ad spending on the focal channel (search engine and display/referral) and use this as the IV. Exogeneity of this IV is warranted because it is hard to imagine that competitors would adjust their ad spending according to the searching and purchasing behavior of the focal firm's consumers observed in our data.

Furthermore, when a random demand shock occurs, it is reasonable to expect all other competitors to respond by adjusting their ad spending accordingly, but this activity should not be systematically co-determined by the consumer's purchase utility on our focal website.

At the individual level, a consumer may increase the likelihood to click on ads (e.g. to seek deal information actively) and purchase simultaneously due to a random personal shock, e.g. due to a windfall or a sudden increase in income. To control for this type of possibility, we use each individual's number of clicks on ads of the same channel for other product categories (car rental and hotel) as instrument. We argue that when random personal shock occurs (e.g. income increases), the consumer's interest in ads for other product categories also changes, but this should not be directly affected by her purchase intention on the focal product category.

3.4.4 Likelihood Function

Overall, the joint likelihood function in Equation 10 takes into account the selection of entry site, the formation of the choice set, and purchase decision. It is conditional on the individual-specific preferences for various websites in the two stages. We assume the individual website-specific intercepts α_{ij}, β_{ij}, γ_{ij} follow a normal distribution specified below:

$$\begin{pmatrix} \alpha_{ij} \\ \beta_{ij} \\ \gamma_{ij} \end{pmatrix} \sim N\left(\begin{bmatrix} \overline{\alpha}_j \\ \overline{\beta}_j \\ \overline{\lambda}_j \end{bmatrix}, \Omega_j \right), \text{ where } j = 1, 2, \ldots, J$$

Ω_j is the variance-covariance matrix for the website-specific intercepts. We assume that the correlation between intercepts of different stages remains the same across websites, but allow the variance of each intercept to vary by website and stage to reduce the number of parameters to be

estimated[20]. We estimated the model using the Simulated Maximum Likelihood approach. The details for the estimation procedure are provided in Appendix A.2.

$$L\left(B|\theta\right)=\prod_{i=1}^{N}\prod_{t=1}^{T_i}P\left(F_{it}=j|\alpha_i\right)\times P\left(V_{it}|F_{it},\beta_i\right)\times P\left(B_{it}|F_{it},V_{it},\gamma_i\right) \tag{3.10}$$

3.5. Model Results

3.5.1 Stage 1: The Website Visit Decision

Table 3.5 presents parameter estimates for the first stage in the purchase funnel - website visit decision. This stage is composed of two sub-decisions: 1) choosing an entry site from the full choice set of all relevant websites, and 2) deciding whether or not to visit other remaining websites. The estimates for the choice of entry site are presented in the first column of Table 3.5. We observe that Delta.com and Cheaptickets.com are more likely to be chosen as the entry site compared to the other alternatives, all else being equal. We also find that there is a significant variation in preference for the same website across consumers for all websites except United.com and Cheaptickets.com. The coefficients for cumulative ad spending on offline channels at both the national level and the local DMA level are positive and significant, indicating that offline advertising is still important in generating awareness and goodwill for online consumers. Cumulative ad spending on the local-level display ads also significantly increase the website's likelihood of being chosen as the entry site, but the effects of ad spending on the national-level display ads and search ads are not significant. This reflects the difficulty of converting ad spending into exposure, even more so at the national level.

[20] Due to dimensionality concern, we do not allow for correlation across websites, such as allowing intercepts for expedia.com to be correlated with intercepts for orbitz.com, to maintain model tractability.

Our results show that the coefficients for lag visit, cumulative amount of money spent on the website and channel-specific lag information stock are all negative and significant, indicating that they can increase visit utility by reducing search costs. These results are not surprising as they suggest that the more experience a consumer accumulates with a website, the more likely she will choose it to be the entry site. The cognitive lock-in effects explain it well (Johnson et al. 2003). When the amount of experience with the website increases via repeated visits and purchases, the consumer becomes locked in due to the reduction in costs associated with navigating or understanding the website.

More specifically, our results indicate that information stock accumulated via direct channel (-0.56) is the most effective in reducing cost of visit, followed by information stock accumulated via search engines (-0.33). Although email is the least popular channel according to Table 3.4, information stock accumulated via this channel is still quite effective in making a website the entry site. This could be due to the fact that email subscribers are more loyal to the website compared to those who visit the website via display ads or referral engines. Our result indicates that e-mail is more effective than display ads or referral engines in influencing consumers to choose a website to enter.

Next, we discuss the estimation results for the visit decision beyond the entry site. The estimates for website-specific intercepts represent individual consumers' net preference for visiting the website. In general consumers have a low probability of visiting other websites beyond the entry site, and this is particularly true for the airline companies' own online booking websites and small OTAs. Again, we observe significant variations across individuals in their preferences for visiting a website (standard deviations are statistically significant varying from 1.14 to 2.14,

Table 3.5. Model Estimates: Visit Decision

Parameter	Entry Site		Remaining Sites	
	Mean	Std Dev	Mean	Std Dev
Delta.com	**0.68**	**0.69**	**-1.38**	**1.37**
AA.com	-0.02	**0.53**	**-1.94**	**1.27**
United.com	0.52	0.07	**-1.94**	0.03
Other Airlines	**0.69**	**1.14**	-0.65	**1.46**
Expedia.com	0.07	**1.19**	-0.37	**2.14**
Priceline.com	-0.02	**1.15**	-1.01	**1.89**
Orbitz.com	0.18	**0.73**	-1.11	**1.95**
Travelocity.com	-0.11	**0.76**	**-1.34**	**1.85**
Cheaptickets.com	**1.57**	0.01	**-1.38**	**1.14**
Other OTAs	baseline	baseline	**-1.67**	**1.25**
Cum Ad spending: National Offline	**0.11**		0.01	
Cum Ad spending: Local Offline	**0.07**		0.01	
Cum Ad spending: National Display	0.00		**0.05**	
Cum Ad spending: Local Display	**0.06**		**0.09**	
Cum Ad spending: Search Ads	0.03		**-0.21**	
Lag Visit *j*	**-1.11**		**-1.36**	
Cumulative Spending[21]	**-0.07**		**-0.14**	
Lag Info Stock: Search	**-0.33**		**-0.29**	
Lag Info Stock: Display/Referral	**-0.05**		**-0.10**	
Lag Info Stock: Email	**-0.21**		**-0.31**	
Lag Info Stock: Direct	**-0.56**		**-0.46**	
Log Likelihood	-15,672		-46,865	
AIC	31,401		93,793	

*Bold figures: p-value < 0.05

except for United.com). Offline channel spending does not significantly influence consumers' decisions to visit the remaining websites. The coefficients for lag visit, cumulative amount of

[21] Since the monthly decay rate is not a parameter of central interest, we conducted a grid search from 0.6 to 0.9 rather than estimating it directly in the Simulated Maximum Likelihood Estimator with other parameters. It is set to 0.7 during the estimation because 0.7 yields the best log-likelihood.

money spent on the website and channel-specific lag information stock are found to be significant and negative, indicating that they can lead to lower search costs. Direct channel (-0.46) turns out to be the most effective channel in enticing consumers to visit beyond the entry site, followed by email (-0.31) and search ads (-0.29).

3.5.2 Stage 2: Purchase Decision

3.5.2.1 Control Function Results

We first report the first stage regression results in Table 3.6, which details the effects of the instruments on the information stock for the four channels (the endogenous variables). The R-squares of these four regression models (ranging from 0.76 to 0.91) are high, indicating that our instruments explain a great deal of variations in the endogenous variables. We then employ the J-statistic of Hansen (1982) to test exogeneity of these instruments. The statistic is distributed as chi-square with the degree of freedom equal to the number of over-identifying restriction. The result shows that the Hansen's J of our model is $\chi^2(6) = 9.17$ (p=.164), which implies exogeneity of these instruments. The estimates for the lagged information stock for all four channels are positive and significant, meaning that the individual's past interaction with the same channel is a good candidate to account for *customer selection*. Firm spending on search ads significantly increases a consumer's information stock through search engines. This is unsurprising – higher spending leads to higher ranks in sponsored ads and in turn higher consumer click-through rate (Ghose and Yang 2009). Offline ad spending also increases information stock for the four channels significantly as expected, making it a strong instrument to control for the temporal simultaneity between advertising and unobserved demand shocks.

Table 3.6. Model Estimates: Control Function

	Search Ads	Display /Referral	Email	Direct
Intercept	**-0.01**	0.00	0.00	**-0.03**
Channel-specific browsing history	**0.02**	**0.01**	**0.01**	**0.03**
Channel-specific ad spending (local)	-----	0.00	-----	-----
Channel-specific ad spending (national)	**0.13**	0.00	-----	-----
Offline ad spending (local)	0.00	**0.01**	**-0.01**	0.00
Offline ad spending (national)	**0.04**	**0.11**	**0.21**	**0.13**
Competitor's channel-specific ad spending (local)	-----	0.00	-----	-----
Competitor's channel-specific ad spending (national)	0.03	**0.08**	-----	-----
Channel-specific ad clicks on other categories	**-0.02**	**-0.06**	0.00	**-0.04**
R-square	0.88	0.88	0.91	0.76
F	8,892.2	8,232.5	12,820.7	4,163.5

*Bold figures: p-value < 0.05

We then include the four residuals generated in the first-stage regression as control functions in the purchase utility function. The estimates for these four control functions are positive and significant (ranging from 0.73 to 2.45), suggesting that controlling for endogeneity is necessary in estimating the effectiveness of multi-channel advertising.

3.5.2.2 Model Estimates for Purchase Decision

We present the parameter estimates for the purchase decision of our proposed model in the first and second columns of Table 3.7. First, we notice that among all the competing websites, Cheapticket.com, being a relatively niche OTA, is most preferred conditional on it being searched. Its net preference is the least negative (baseline is the no purchase option). Among the airlines' official websites, United.com is the most preferred. This indicates that once the consumers have visited United.com, their conversion rate is higher compared to other airline websites, everything else being equal. There is significant variation in these website-specific intercepts, indicating high level of heterogeneity across consumers in their preferences for these websites.

The coefficients for the information stocks gained from all four channels are positive and significant, suggesting that all advertising channels are important in generating sales. However, among the four channels, the direct channel (0.56) is the most effective, followed by search ads (0.41) and email ads (0.39). We discuss the implications of these estimates in detail using marginal impact analysis in the next section.

We find that the sequence in website visits also matters. Being either the first or the last website visited during the visit stage significantly increases conversion probability compared to those visited in-between. This finding shows that while consumers usually start with their favorable website, they are also likely to convert on the last-visited website. In addition, we find that consumers exhibit strong inertia when making online purchases. A prior online flight purchase increases the likelihood of a current purchase, especially if the prior purchase is made on the same website. And the more money a consumer has spent on booking flights at a website in the past, the higher her conversion probability will be at the website. Lag visit also increases conversion probability significantly, suggesting a carryover effect on future purchases when the site was included in a consumer's choice set.

3.5.3 Advertising Elasticity of Visit and Demand

While parameter estimates show the relative significance of different channels, it is often more informative to know the extent to which visit and purchase probabilities change in response to a change in information stock. We quantify the impact of information stock of different channels on visit and purchase probability using a marginal impact analysis. The marginal impact is measured as the change in visit and purchase probabilities for each website given a change of the same magnitude in information stock collected through a specific channel for a specific website.

Table 3.7. Model Estimates: Purchase Decision

Parameter	Proposed Model		Baseline Model	
	Mean	Std Dev	Mean	Std Dev
Delta.com	**-7.46**	**1.90**	**-8.23**	**2.25**
AA.com	**-9.04**	**3.07**	**-9.87**	**3.51**
United.com	**-7.08**	**1.76**	**-7.85**	**2.05**
Other Airlines	**-7.70**	**3.29**	**-8.32**	**3.64**
Expedia.com	**-7.76**	**1.77**	**-8.40**	**2.14**
Priceline.com	**-7.88**	**1.73**	**-8.63**	**2.18**
Orbitz.com	**-7.38**	**1.39**	**-8.07**	**1.81**
Travelocity.com	**-8.28**	**1.99**	**-9.16**	**2.53**
Cheaptickets.com	**-5.51**	**1.36**	**-6.53**	**1.62**
Other OTAs	**-10.73**	**3.71**	**-11.86**	**4.28**
Info Stock: Search	**0.41**		**0.44**	
Info Stock: Display/Referral	**0.18**		0.11	
Info Stock: Email	**0.39**		**0.42**	
Info Stock: Direct	**0.56**		**0.59**	
Ad spending: National Offline	0.19		0.14	
Ad spending: Local Offline	**0.04**		**0.04**	
Entry Site	**1.23**		**1.58**	
Last Site	**1.38**		**1.76**	
Lag No Purchase	**-0.59**		**-0.55**	
Lag Purchase on *j*	**0.75**		**0.88**	
Lag Visit on *j*	**0.59**		**0.63**	
Cumulative Spending	**0.07**		**0.08**	
Cumulative Page	**-0.38**		**-0.38**	
Log-likelihood	5,444		5,650	
AIC	10,962		12,080	

*Bold figures: p-value < 0.05

For example, a marginal impact analysis can inform us how probabilities of being chosen as the entry site change for each website if a consumer has collected more information about Expedia.com by browsing more webpages under the influence of search advertising in the past. This marginal impact analysis therefore enables us to compare the effects of ad stock for the same website across different channels and for the same channel across different websites.

We need to simulate the change in choice probabilities because the net preferences for different websites are heterogeneous across consumers and are drawn from a normal distribution. The simulation follows six steps, which is explained in detail in Appendix A.3. The resulting average changes across consumers for the choice of entry site, visit probability of remaining websites, and purchase probability of visited websites are reported in Table 3.8 to 3.10.

Table 3.8 tabulates the own- and cross-impacts on the entry site choice as a result of 10 additional pages of a focal site browsed in the last month through additional ad clicks resulted from the four different channels. The own marginal impacts show that in general information stock collected through direct channel is most effective in raising the probability of being chosen as the entry website, followed closely by search ads. Display/referral ads and email ads also increase the choice probability but the effect is small compared to the other two channels. Every website sees a boost in choice probability due to additional information stock, but this effect size varies significantly across websites. Ad stock is most effective for Expedia.com. For example, increased information stock through search ads due to 10 more pages browsed can increase Expedia.com's probability of being chosen as the entry site by 5.86%, but the same amount of change in information stock only increases United.com's choice probability by 1.77%.

Table 3.8. Impacts of Ad Stock on Choice of Entry Site

Search Ads

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Travelo -city	Cheapti -ckets	Other OTAs
Delta	3.17%	-0.31%	-0.14%	-0.67%	-0.56%	-0.37%	-0.34%	-0.32%	-0.27%	-0.20%
AA	-0.30%	3.40%	-0.15%	-0.70%	-0.61%	-0.41%	-0.37%	-0.37%	-0.27%	-0.22%
United	-0.16%	-0.18%	1.77%	-0.34%	-0.31%	-0.18%	-0.19%	-0.18%	-0.14%	-0.11%
Other Airlines	-0.55%	-0.60%	-0.25%	5.41%	-1.14%	-0.73%	-0.65%	-0.64%	-0.49%	-0.36%
Expedia	-0.53%	-0.59%	-0.25%	-1.27%	5.86%	-0.76%	-0.76%	-0.76%	-0.53%	-0.40%
Priceline	-0.38%	-0.44%	-0.17%	-0.89%	-0.85%	4.43%	-0.53%	-0.51%	-0.36%	-0.29%
Orbitz	-0.37%	-0.42%	-0.18%	-0.86%	-0.91%	-0.56%	4.45%	-0.50%	-0.37%	-0.29%
Travelocity	-0.34%	-0.40%	-0.17%	-0.81%	-0.87%	-0.53%	-0.49%	4.21%	-0.34%	-0.27%
Cheaptickets	-0.29%	-0.31%	-0.13%	-0.64%	-0.63%	-0.39%	-0.38%	-0.35%	3.32%	-0.21%
Other OTAs	-0.24%	-0.27%	-0.12%	-0.52%	-0.54%	-0.34%	-0.32%	-0.31%	-0.24%	2.90%

Display Ads/Referral Engines

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Travelo -city	Cheapti -ckets	Other OTAs
Delta	0.49%	-0.05%	-0.02%	-0.11%	-0.09%	-0.06%	-0.05%	-0.05%	-0.04%	-0.03%
AA	-0.05%	0.54%	-0.02%	-0.12%	-0.10%	-0.06%	-0.06%	-0.06%	-0.04%	-0.03%
United	-0.02%	-0.02%	0.23%	-0.04%	-0.04%	-0.02%	-0.02%	-0.02%	-0.02%	-0.01%
Other Airlines	-0.10%	-0.11%	-0.04%	0.97%	-0.21%	-0.13%	-0.12%	-0.11%	-0.09%	-0.06%
Expedia	-0.08%	-0.09%	-0.04%	-0.20%	0.90%	-0.12%	-0.12%	-0.12%	-0.08%	-0.06%
Priceline	-0.05%	-0.06%	-0.02%	-0.13%	-0.12%	0.62%	-0.08%	-0.07%	-0.05%	-0.04%
Orbitz	-0.05%	-0.06%	-0.02%	-0.12%	-0.12%	-0.08%	0.60%	-0.07%	-0.05%	-0.04%
Travelocity	-0.05%	-0.06%	-0.02%	-0.12%	-0.13%	-0.08%	-0.07%	0.61%	-0.05%	-0.04%
Cheaptickets	-0.04%	-0.04%	-0.02%	-0.09%	-0.09%	-0.06%	-0.06%	-0.05%	0.48%	-0.03%
Other OTAs	-0.03%	-0.03%	-0.01%	-0.06%	-0.07%	-0.04%	-0.04%	-0.04%	-0.03%	0.37%

Email Ads

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Delta	2.29%	-0.22%	-0.10%	-0.50%	-0.41%	-0.26%	-0.25%	-0.23%	-0.18%	-0.14%
AA	-0.22%	2.52%	-0.11%	-0.55%	-0.46%	-0.30%	-0.28%	-0.26%	-0.19%	-0.15%
United	-0.10%	-0.12%	1.18%	-0.22%	-0.21%	-0.12%	-0.12%	-0.12%	-0.09%	-0.07%
Other Airlines	-0.46%	-0.52%	-0.19%	4.44%	-0.94%	-0.59%	-0.54%	-0.52%	-0.39%	-0.28%
Expedia	-0.39%	-0.44%	-0.19%	-0.98%	4.49%	-0.60%	-0.60%	-0.59%	-0.40%	-0.31%
Priceline	-0.26%	-0.30%	-0.11%	-0.63%	-0.62%	3.14%	-0.38%	-0.37%	-0.26%	-0.20%
Orbitz	-0.25%	-0.29%	-0.12%	-0.59%	-0.64%	-0.40%	3.09%	-0.35%	-0.26%	-0.19%
Travelocity	-0.23%	-0.27%	-0.11%	-0.57%	-0.63%	-0.37%	-0.34%	2.93%	-0.23%	-0.18%
Cheaptickets	-0.19%	-0.20%	-0.09%	-0.44%	-0.43%	-0.27%	-0.26%	-0.24%	2.27%	-0.14%
Other OTAs	-0.15%	-0.17%	-0.07%	-0.33%	-0.35%	-0.22%	-0.20%	-0.19%	-0.15%	1.83%

Direct

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Delta	5.24%	-0.51%	-0.21%	-1.05%	-0.94%	-0.61%	-0.59%	-0.54%	-0.44%	-0.35%
AA	-0.47%	5.21%	-0.22%	-1.02%	-0.96%	-0.59%	-0.60%	-0.55%	-0.45%	-0.35%
United	-0.26%	-0.29%	3.11%	-0.56%	-0.54%	-0.34%	-0.33%	-0.32%	-0.26%	-0.21%
Other Airlines	-0.78%	-0.87%	-0.35%	8.10%	-1.70%	-1.11%	-1.00%	-0.96%	-0.75%	-0.57%
Expedia	-0.74%	-0.83%	-0.36%	-1.75%	8.02%	-1.01%	-1.01%	-1.01%	-0.75%	-0.57%
Priceline	-0.54%	-0.58%	-0.25%	-1.29%	-1.16%	6.20%	-0.73%	-0.71%	-0.54%	-0.41%
Orbitz	-0.56%	-0.62%	-0.27%	-1.24%	-1.27%	-0.80%	6.47%	-0.72%	-0.57%	-0.43%
Travelocity	-0.51%	-0.57%	-0.25%	-1.18%	-1.24%	-0.76%	-0.70%	6.14%	-0.53%	-0.41%
Cheaptickets	-0.48%	-0.53%	-0.23%	-1.08%	-1.04%	-0.66%	-0.64%	-0.61%	5.67%	-0.38%
Other OTAs	-0.43%	-0.47%	-0.21%	-0.95%	-0.94%	-0.59%	-0.57%	-0.55%	-0.44%	5.16%

The off-diagonal entries of Table 3.8 show the cross-impact of change in information stock for the website in the row on choice probability for the website in the column. This effect size varies by channels. For example, if the consumer reaches Expedia.com through search engine, Priceline.com's probability of being chosen as the entry site will decrease by 0.76%. However, the same probability will only decrease by 0.12% under the influence of display/referral ads.

Table 3.9 shows the changes in visit probability given an increase in information stock through different channels by websites. Again, information stock collected through direct channel is most valuable in driving up the visit probability. However, information stock gathered through email ads is also very effective, compared to search ads and display/referral ads. For example, for AA.com, if the consumer browsed 10 more webpages in the last month through email ad, the probability that she will visit the website in this purchase occasion increases by 7.54%, more than triple the change in visit probability as a result of display/referral ads. Across websites, Expedia.com is most sensitive to the change in information stock through a channel, followed by other airlines' direct websites, Priceline.com and Orbitz.com.

Table 3.10 tabulates the change in purchase probabilities for every website as a result of 10 more pages browsed through additional ad clicks resulted from different channels on a focal website. The diagonal entries record the own-marginal impacts for each website. In general email ads have a greater impact on conversion probability than any other advertising formats. For example, the change in information stock gathered through email ads increases conversion probability by 4.16% for expedia.com, higher than the 3.23% increase due to more information stock through search ads or 1.79% increase due to more information stock through display/referral ads. While the high effectiveness of search engine advertising or direct visits is expected because

they are consumer-initiated, the larger impact of email ads might be attributed to the higher level of consumer loyalty. Consumers who subscribed to a website's email list are often more loyal than those who are not, so they may also become more responsive to email campaigns. Moreover, if a consumer happens to receive an email ad when she has a purchase need, the coincidence will also increase her purchase probability on the website.

However, even though every website enjoys a boost in conversion probability due to additional ad clicks, this effect size is not uniform across websites. Advertising is most effective for Expedia.com, Orbitz.com, but has barely any impact on small OTAs' conversion probability. The disparate effectiveness of the same advertising channel across competitors suggests the importance of ad content and message strategies. Firms can use these figures as the starting point toward a diagnostic process to improve its own ad messaging and copy strategies. By benchmarking its own ad effectiveness with a competitor's, a firm can determine its standing among the competitive pack.

The off-diagonal entries in Table 3.10 show the cross-marginal impacts of advertising on competitors' conversion probabilities. Each off-diagonal entry in Table 3.10 indicates the change in conversion probability for the column website due to more information stock on the row website. Increasing information stock draws disproportionately from the website's competitors. For example, an increase in information stock gathered through email ads for Expedia.com will decrease the purchase probability for Priceline.com, Orbitz.com and Travelocity.com by more than 0.33%, but will only decrease the purchase probabilities for the other websites by less than 0.2%. This indicates that the competition between these OTAS is more intense. Moreover, the cross-impact of ad stock varies across channels. For example, an increase in information stock through

Table 3.9. Impacts of Ad Stock on Search Probability of the Remaining Websites

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Search	5.89%	5.42%	4.86%	7.94%	8.62%	7.33%	7.61%	6.97%	5.51%	5.11%
Display/Referral	2.14%	2.01%	1.60%	3.08%	3.00%	2.35%	2.40%	2.40%	1.89%	1.58%
Email	7.54%	7.10%	5.87%	10.31%	10.79%	8.98%	9.21%	8.57%	6.82%	6.06%
Direct	9.29%	8.10%	8.19%	12.24%	11.54%	9.60%	10.18%	9.47%	8.63%	7.80%

Table 3.10. Impacts of Ad Stock on Purchase Probability

Search Ads

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Delta	3.23%	-0.06%	-0.08%	-0.10%	-0.14%	-0.14%	-0.14%	-0.10%	-0.08%	-0.02%
AA	-0.06%	3.25%	-0.05%	-0.12%	-0.14%	-0.20%	-0.10%	-0.08%	-0.06%	-0.01%
United	-0.11%	-0.04%	2.82%	-0.09%	-0.11%	-0.07%	-0.07%	-0.20%	-0.02%	0.00%
Other Airlines	-0.08%	-0.12%	-0.09%	4.48%	-0.21%	-0.17%	-0.17%	-0.14%	-0.09%	-0.01%
Expedia	-0.14%	-0.17%	-0.12%	-0.28%	5.12%	-0.25%	-0.29%	-0.24%	-0.15%	-0.03%
Priceline	-0.15%	-0.23%	-0.11%	-0.18%	-0.27%	4.13%	-0.24%	-0.18%	-0.15%	-0.03%
Orbitz	-0.15%	-0.15%	-0.09%	-0.23%	-0.32%	-0.26%	4.71%	-0.20%	-0.20%	-0.04%
Travelocity	-0.13%	-0.11%	-0.22%	-0.15%	-0.30%	-0.18%	-0.21%	4.14%	-0.17%	-0.02%
Cheaptickets	-0.08%	-0.06%	-0.03%	-0.09%	-0.15%	-0.17%	-0.21%	-0.17%	3.65%	-0.01%
Other OTAs	-0.04%	-0.01%	-0.01%	-0.01%	-0.03%	-0.02%	-0.03%	-0.02%	-0.01%	0.77%

Display Ads/Referral Engines

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Delta	1.79%	-0.03%	-0.05%	-0.05%	-0.07%	-0.06%	-0.07%	-0.08%	-0.07%	-0.01%
AA	-0.03%	1.78%	-0.03%	-0.07%	-0.08%	-0.10%	-0.06%	-0.05%	-0.02%	-0.01%
United	-0.04%	-0.03%	1.50%	-0.04%	-0.07%	-0.06%	-0.03%	-0.09%	-0.01%	0.00%
Other Airlines	-0.05%	-0.07%	-0.04%	2.50%	-0.12%	-0.08%	-0.08%	-0.06%	-0.05%	0.00%
Expedia	-0.07%	-0.07%	-0.07%	-0.12%	2.73%	-0.12%	-0.14%	-0.14%	-0.07%	-0.02%
Priceline	-0.06%	-0.10%	-0.03%	-0.08%	-0.12%	2.02%	-0.11%	-0.09%	-0.07%	-0.02%
Orbitz	-0.06%	-0.06%	-0.03%	-0.09%	-0.13%	-0.10%	2.10%	-0.09%	-0.09%	-0.01%
Travelocity	-0.07%	-0.05%	-0.09%	-0.07%	-0.14%	-0.09%	-0.10%	2.05%	-0.07%	-0.01%
Cheaptickets	-0.07%	-0.03%	-0.01%	-0.05%	-0.08%	-0.07%	-0.10%	-0.08%	1.90%	0.00%
Other OTAs	-0.01%	0.00%	0.00%	0.00%	-0.02%	-0.02%	-0.02%	-0.01%	0.00%	0.33%

Email Ads

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Delta	4.16%	-0.08%	-0.10%	-0.14%	-0.19%	-0.17%	-0.17%	-0.18%	-0.16%	-0.02%
AA	-0.08%	4.20%	-0.07%	-0.17%	-0.21%	-0.25%	-0.16%	-0.10%	-0.08%	-0.01%
United	-0.12%	-0.08%	3.72%	-0.10%	-0.17%	-0.14%	-0.09%	-0.21%	-0.04%	0.00%
Other Airlines	-0.13%	-0.18%	-0.10%	5.82%	-0.28%	-0.21%	-0.22%	-0.17%	-0.13%	-0.01%
Expedia	-0.19%	-0.20%	-0.15%	-0.29%	6.42%	-0.33%	-0.37%	-0.34%	-0.19%	-0.04%
Priceline	-0.17%	-0.26%	-0.12%	-0.23%	-0.34%	4.98%	-0.26%	-0.23%	-0.16%	-0.05%
Orbitz	-0.17%	-0.16%	-0.09%	-0.24%	-0.38%	-0.29%	5.43%	-0.24%	-0.26%	-0.04%
Travelocity	-0.19%	-0.11%	-0.20%	-0.18%	-0.37%	-0.21%	-0.24%	4.85%	-0.19%	-0.02%
Cheaptickets	-0.17%	-0.08%	-0.03%	-0.12%	-0.20%	-0.19%	-0.27%	-0.19%	4.45%	-0.01%
Other OTAs	-0.04%	-0.02%	-0.01%	-0.01%	-0.06%	-0.05%	-0.04%	-0.02%	-0.01%	0.91%

Direct

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Delta	2.60%	-0.05%	-0.06%	-0.08%	-0.14%	-0.10%	-0.09%	-0.15%	-0.09%	-0.01%
AA	-0.04%	2.46%	-0.03%	-0.12%	-0.15%	-0.11%	-0.11%	-0.06%	-0.04%	-0.01%
United	-0.05%	-0.08%	2.49%	-0.06%	-0.13%	-0.07%	-0.06%	-0.07%	-0.03%	0.00%
Other Airlines	-0.08%	-0.12%	-0.05%	3.93%	-0.19%	-0.12%	-0.15%	-0.11%	-0.08%	-0.01%
Expedia	-0.15%	-0.11%	-0.13%	-0.17%	4.35%	-0.17%	-0.21%	-0.17%	-0.12%	-0.03%
Priceline	-0.08%	-0.12%	-0.06%	-0.12%	-0.17%	3.08%	-0.10%	-0.12%	-0.10%	-0.02%
Orbitz	-0.13%	-0.11%	-0.07%	-0.15%	-0.25%	-0.19%	3.68%	-0.14%	-0.15%	-0.03%
Travelocity	-0.14%	-0.05%	-0.05%	-0.11%	-0.21%	-0.10%	-0.12%	2.86%	-0.09%	-0.01%
Cheaptickets	-0.09%	-0.07%	-0.01%	-0.06%	-0.11%	-0.08%	-0.17%	-0.08%	2.50%	-0.01%
Other OTAs	-0.01%	-0.01%	-0.01%	-0.01%	-0.03%	-0.02%	-0.03%	-0.01%	-0.01%	0.51%

Table 3.12. Impacts of Ad Stock on Search Probability of the Remaining Websites (Baseline Model)

	Delta	AA	United	Other Airlines	Expedia	Priceline	Orbitz	Traveloc -ity	Cheaptic -kets	Other OTAs
Search	3.07%	2.38%	2.69%	4.07%	4.28%	3.25%	3.93%	3.25%	3.59%	0.44%
Display/Referral	0.93%	0.69%	0.77%	1.23%	1.27%	0.87%	0.96%	0.86%	1.02%	0.10%
Email	4.12%	3.19%	3.64%	5.45%	5.62%	4.11%	4.74%	3.95%	4.54%	0.57%
Direct	2.38%	1.62%	2.27%	3.39%	3.38%	2.23%	2.82%	1.99%	2.41%	0.26%

Expedia.com's email ads (-0.33%) imposes greater damage versus an increase in information stock through its search ads (-0.25%) to the conversion probability of Oritz.com.

3.6 Predictive Analysis and Comparisons

In order to demonstrate the importance and necessity of modeling competition between firms, we compare our competitive analytics approach with a baseline model. The baseline model mimics the scenario where the company only knows its customers' interactions with itself, but not with their competitors. In this case, a consumer's purchase decision is modeled as a binary choice between buying and not buying on a visited website (in contrast to a multinomial choice from all the visited websites in the proposed model). Our comparison is carried out on the purchase stage with a focus on their predictive performances in terms of predicting purchases in the out-of-sample.

We first present the parameter estimates and in-sample model fit statistics of the baseline model in the third and fourth columns of Table 3.7. It is clear that our proposed model outperforms the baseline model in terms of model fitness as measured by the log-likelihood and the AIC criterion.

We then compare the predictive performance of these two models on a holdout sample. The holdout sample is composed of 3,576 purchase occasions made by the same group of consumers for a three-month period between October 2010 and December 2010. Each consumer on average visited 2.41 websites in a purchase occasion, and 950 out of these 3,576 occasions ended up with a purchase, consistent with our estimation sample. The predicted purchase probability is simulated using the same method described in Appendix A.3.

We use two criteria to evaluate the predictive performance. The first one is the true positive (TP) rate. It is defined as the percentage of actual positive events that are predicted as "positive"

by the model. We label the predicted outcome as "positive" if the probability exceeds 0.1, which is close to the actual purchase probability (0.11). The result shows that the TP rate of our proposed model is 63.37%, significantly higher than that of the alternative model, which is only 43.89%. As is shown in Table 3.11, the TP rate of our proposed model is 4.76% to 30.99% higher than that of the baseline model. The proposed model predicts the purchase probability for OTAs especially well, with the TP rates all above 70%.

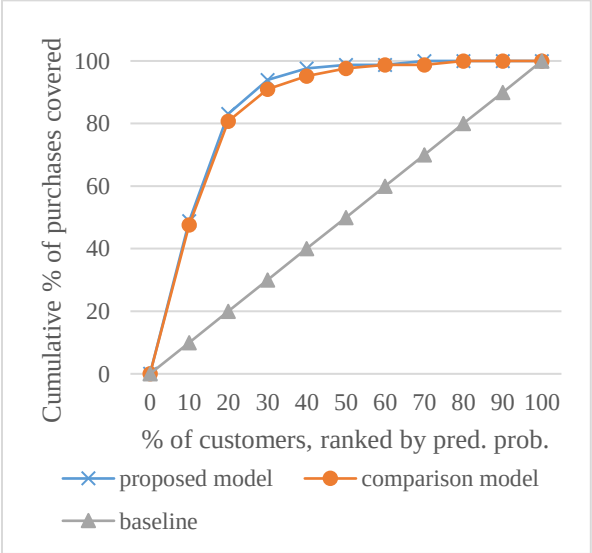
Table 3.11. TP Rate Comparison

	Actual # of purchases	TP rate of proposed model	TP rate of baseline model
Delta.com	90	74.44%	55.56%
AA.com	105	25.71%	14.29%
United.com	37	67.57%	54.05%
Other Airlines	278	52.88%	37.77%
Expedia.com	166	80.72%	59.64%
Priceline.com	68	77.94%	47.06%
Orbitz.com	71	83.10%	52.11%
Travelocity.com	54	72.22%	48.15%
Cheaptickets.com	60	83.33%	55.00%
Other OTAs	21	4.76%	0.00%
Average		63.37%	43.89%

The second criterion we employ is lift, which is a measure of the effectiveness of a predictive model calculated as the ratio between the results obtained with and without the predictive model. Cumulative gains and lift charts are often used to visualize model performance. We rank the observations based on the purchase probability and count the cumulative number of actual positive

events picked up in each percentile and plotted the lift charts over every ten percentiles. Figure 3.2 illustrates the lift chart for two representative websites, Expedia.com and Cheaptickets.com for Airline and OTA sites respectively. Both the proposed model and the baseline model capture the actual purchases faster than a random model represented by the 45-degree line. However, our proposed model outperforms the comparison model as the lift line for our model is consistently above the line for the baseline model. We also find that our proposed model performs especially well for websites with smaller market share, such as Cheaptickets.com. The gap between the two lines is much larger.

Expedia.com



Cheaptickets.com

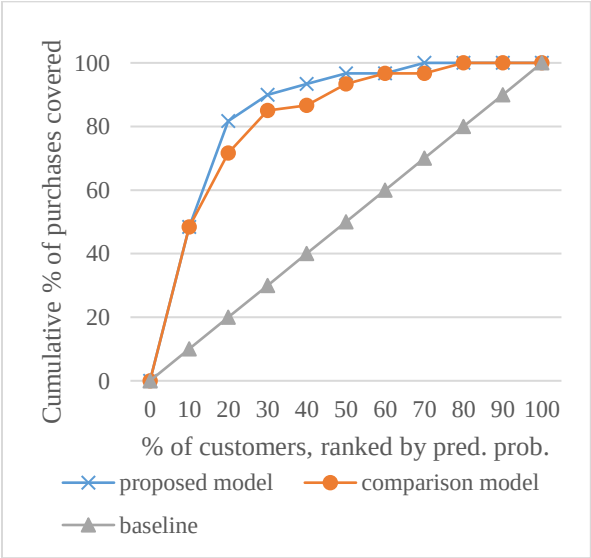


Figure 3.2. Lift charts for Expedia.com and Cheaptickets.com

Lastly, we compare the two models in terms of the marginal impacts of the estimated coefficients. The results indicate that information stock collected through display/referral ads does not contribute to the purchase probability significantly, which is an obvious departure from the insights accrued from our proposed model. The marginal impacts of the other advertising channels

for the baseline model are reported in Table 3.12. By comparing Table 3.12 and the diagonal entries in Table 3.10, we can see that the baseline model underestimates the impact of an increase in information stock on the website's purchase probability. For example, the purchase probability of Orbitz.com will increase 4.71% due to an increase in information stock through search advertising as predicted by the proposed model, but the baseline model entails that the purchase probability will only increase by 3.93%. This demonstrates that failing to factor in competition can lead to biased estimation of the ad effects.

3.7 Conclusions, Limitations, and Future Research

Practitioners have long been calling for better methods to solve the multi-channel attribution problem. Although recent research has made strides when estimating the effects of prior advertising touches, the current ad attribution models still suffer from one fundamental drawback. That is, they focus on analyzing the converting path with respect to one focal firm while largely overlooking the impact of competitors' advertising, leading to biased ad effectiveness estimates. The problem exacerbates in a purchase funnel model because competing advertising not only influences the final conversion stage, but also the earlier product information search and alternative evaluation stage.

We overcome these deficiencies and improvise the current multi-channel attribution models by accounting for competition in a consumer's purchase funnel, estimating both the direct and indirect effect of ad-clicks prior to conversion. Specifically, we develop an integrated multi-stage choice model in order to measure the advertising effectiveness on a consumer's online search and purchase decisions in a multi-channel, multi-touch-point, multi-competitor environment using individual-level advertising response data. Our model considers: (1) consumer choice of entry site,

(2) consumer search decisions concerning the remaining websites that compete in the same industry and (3) subsequent purchases at one of the websites searched. We estimate this model on a unique panel data set covering both search and purchase decisions across multiple websites in the online air ticket booking industry. We then use these parameter estimates to compute the direct and indirect marginal impacts of online advertising channels and predict a consumer's future purchasing behavior based on observed variables. We further compare our proposed competition-centric model against a baseline model using data from a single-firm only. Our model outperforms the baseline model in terms of in-sample model fit as well as out-of-sample predictive performance, necessitating modeling competition in multichannel ad attribution.

Our findings offer managers new insights for evaluating advertising effectiveness. First, we find that in the online air travel industry, information stock collected through search advertising is the most effective ad format in increasing the visit probability. This is consistent with the findings in the prior literature that search ads are the most important channel under all attribution methodologies (Abhishek et al. 2014). The finding also echoes the industry belief that search engine ads are the king of online advertising. However, our result regarding the strong direct effect of email ads is contrary to the conventional belief that email ads are less useful in sealing the deal. This difference demonstrates the strength of our approach because we consider the entire purchase journey and reward all ad clicks prior to conversion, whereas most rule-based models adopted by practitioners only assign weights to touches that directly lead to a purchase; such approaches under-estimate the effect of assisting touches prior to purchase.

Second, we are able to quantify the channel- and stage-specific marginal ad impacts based on our model estimates. Some websites may find that advertising has small effect on improving

its conversion probability directly, but this may not be the case for its competitors. These websites should accordingly benchmark against their competitors and focus more on ad designs. For example, firms may need to choose better keywords and improve their search ad rankings; they may also make display ads more attractive by matching the webpage content or targeting more interested online users. We also find that ad clicks through a specific channel can have either a direct impact on the conversion probability, an indirect impact through its influence on consumers' search decisions during the next purchase session, or both. Additional information stock accrued through a specific channel may still be beneficial because they increase the likelihood of these websites being included in the consumer's consideration set during the next purchase session and indirectly increase the future conversion probability, even if they have a small effect in leading to a higher direct conversion. Firms must consider both the direct and indirect effects of advertising when allocating their budgets.

Multi-channel attribution is a rapidly growing new research area and many aspects await future investigation. Our model spawns at least five interesting directions for further investigation. First, one of the main limitations of our research is that we do not observe the content of webpages visited by the consumers. If page content data is available, future research can take advantage of such data by incorporating product attribute information such as prices and product features into ad attribution models. Second, since we do not observe consumers' ad exposure in our data we cannot model their click-through decisions. It will be a very interesting extension to further model consumers' ad click decisions together with their purchase funnel decisions. Third, another limitation of our estimation sample is that the data was collected in 2010 and several popular ad channels firms in use today were not as developed at the time. This model can therefore be readily

extended to incorporate the effects of more recent ad channels such as social media, mobile ads, etc. Fourth, marketing practitioners are particularly interested in determining how online and offline advertising interact, as well as the spillover effect of online advertising on offline sales and vice versa. Future research may answer this question by incorporating offline advertising and purchase decisions into the model. Finally, our model can be applied to other product categories in order to test these findings regarding the direct and indirect effects of advertising. Different industries have their own characteristics and may differ widely in channels they employ to reach out to customers.

APPENDIX

A.1 Text Classification

In this paper, we use the standard bag-of-features framework to classify text content. The complete machine learning process proceeds as follows. First, two research assistants hand-labeled 1,000 Tweets for training purposes. For each Tweet, three labels ("fact" or "opinion", valence, and topic) are created corresponding to the pre-determined variables. In the second step, we define the "feature bag" to be processed by the classifier. This is done by first tokenizing the Tweets into sets of words (sequence of words is ignored) and filtering out all stopping words and punctuations (except question marks and exclamation points) because they contain little information about the variables. Then, the "feature bag" is constructed using a list of the top 500 most frequently-used words in the overall corpus (plus question marks and exclamation points). In the third step, a feature extractor is applied to the training set to check whether each of words in the "feature bag" is present in a given Tweet. The result of this step is a 502-dimensional binary feature set for each Tweet. In the fourth step, pairs of feature sets and labels are fed into the NBC algorithm to generate a model. Finally, the same feature extractor is used to convert remaining Tweets in our sample to feature sets. These feature sets are then fed into the model which generates the predicted labels.

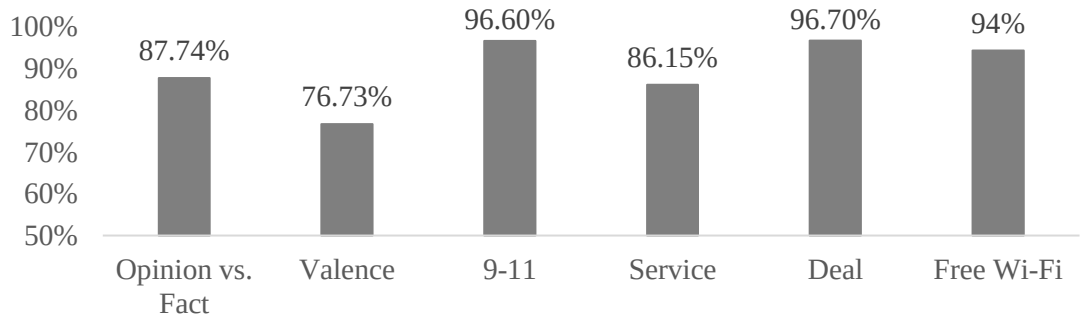


Figure A.1. Accuracy Rate of Naïve Bayes Classifier

A.2 Procedures of Simulated Maximum Likelihood estimation

We assume that the overall attractiveness of visiting or purchasing from website *j* (website-specific intercepts in Equation 3.1, 3.2, and 3.3) can vary over people but being constant over choice situations for each person (Erdem and Keane 1996). The individual website-specific intercepts α_{ij}, β_{ij}, γ_{ij} are specified to follow normal distribution:

$$\begin{pmatrix} \alpha_{ij} \\ \beta_{ij} \\ \gamma_{ij} \end{pmatrix} \sim N\left(\begin{bmatrix} \overline{\alpha}_j \\ \overline{\beta}_j \\ \overline{\lambda}_j \end{bmatrix}, \Omega_j \right), \text{ where } j = 1, 2, \ldots, J$$

The probability conditional on θ_i is

$$L_i(\theta) = \prod_{t=1}^{T_i} L_{it}(\theta) = \prod_{t=1}^{T_i} P\left(F_{it} = j \middle| \alpha_i\right) \times P\left(V_{it} = C_{it} \middle| F_{it}, \beta_i\right) \times P\left(B_{it} = k \middle| F_{it}, V_{it}, \gamma_i\right)$$

The unconditional choice probability is therefore the integral of $L_i(\theta)$ over all possible values of θ_i:

$$P_{it} = \int L_{it}(\theta) f(\theta) d\theta$$

1. Draw θ_i^r from its distribution.

A set of random variables $\eta_i^r = \left(\eta_{j1}^r, \eta_{j2}^r, \eta_{j3}^r\right)$, j = 1, 2, …, J are drawn from i.i.d. standard normal distribution. Then we compute the website-specific intercepts for each person as follows:

$$\alpha_{ij}^r = \overline{\alpha}_j + \sigma_{\alpha_j} \eta_{j1}^r$$

$$\beta_{ij}^r = \overline{\beta}_j + \sigma_{\beta_j} \left(c_{21_j} \eta_{j1}^r + c_{22_j} \eta_{j2}^r \right)$$

$$\gamma_{ij}^r = \overline{\gamma}_j + \sigma_{\gamma_j} \left(c_{31_j} \eta_{j1}^r + c_{32_j} \eta_{j2}^r + c_{33_j} \eta_{j3}^r \right)$$

The benefit of this specification is that the correlation between intercepts of different stages remains the same across websites, thus significantly reducing the number of parameters to be estimated. With this specification, the variance and covariance matrix can be written as follows:

$$\Omega_j = \begin{bmatrix} \sigma_{\alpha_j}^2 & \sigma_{\alpha_j}\sigma_{\beta_j}c_{21} & \sigma_{\alpha_j}\sigma_{\gamma_j}c_{31} \\ \sigma_{\alpha_j}\sigma_{\beta_j}c_{21} & \sigma_{\beta_j}^2\left(c_{21}^2+c_{22}^2\right) & \sigma_{\beta_j}\sigma_{\gamma_j}\left(c_{21}c_{31}+c_{22}c_{32}\right) \\ \sigma_{\alpha_j}\sigma_{\gamma_j}c_{31} & \sigma_{\beta_j}\sigma_{\gamma_j}\left(c_{21}c_{31}+c_{22}c_{32}\right) & \sigma_{\gamma_j}^2\left(c_{31}^2+c_{32}^2+c_{33}^2\right) \end{bmatrix}$$

2. $L_{it}(\theta)$ is calculated for each period, and the product of these $L_{it}(\theta)$'s is taken:

$$L_i^{\,r}\left(\theta\right) = \prod_{t=1}^{T_i} L_{it}^{\,r}\left(\theta\right) = \prod_{t=1}^{T_i} P\left(F_{it}=j\left|\alpha_i^{\,r}\right.\right) \times P\left(V_{it}=C_{it}\left|F_{it},\beta_i^{\,r}\right.\right) \times P\left(B_{it}=k\left|F_{it},V_{it},\gamma_i^{\,r}\right.\right)$$

3. Repeat 1 and 2 for many time, and the results are averaged: $P_i = \frac{1}{R}\sum_{r=1}^{R} L_i^{\,r}\left(\theta\right)$.

4. Calculate the simulated log-likelihood: $SLL = \sum_{i=1}^{N} P_i$.

A.3 Procedures of Simulated Choice Probabilities

We will use the choice of entry site due to a change in information stock through search engine for Expedia.com as an example to illustrate. In step 1, we update the corresponding variable in the dataset, for example by adding 10 pages browsed a month ago through search advertising for Expedia.com to each individual and then re-computing the information stock through search advertising using Equation 3.4. In step 2, we make a draw of net preferences for each website as described in step 1 in Appendix A.2. In step 3, we calculate the probability of being chosen as the entry website for each website using the updated data and the net preferences drawn in step 2. Repeat step 2 and 3 for many times and in step 4, we calculate the simulated probabilities by taking

a weighted average of the results in step 4. The weight is given as $w_{it}^{r} = \frac{P\left(F_{it} \mid \alpha^{r}\right)}{\sum_{r} P\left(F_{it} \mid \alpha^{r}\right)}$ where F_{it} is the actual chosen entry-site for individual *i* in purchase occasion *t*. In step 5, we follow step 2 to 4 to calculate the original choice probabilities without the change in ad stock. In the last step, we calculate the differences in choice probabilities for each website.

REFERENCES

Abhishek, Vibhanshu, Peter S. Fader, and Kartik Hosanagar (2014), "Media Exposure through the Funnel: A Model of Multi-Stage Attribution," *working paper, Carnegie Mellon University, Pittsburgh*

Agarwal, Ashish, Kartik Hosanagar, Michael D. Smith (2011), "Location, Location, Location: An Analysis of Profitability and Position in Online Advertising Markets," *Journal of Marketing Research*, 48(6):1057-1073.

Agarwal, Ashish, Kartik Hosanagar, Michael D. Smith (2015), "Do Organic Results Help or Hurt Sponsored Search Performance?" *Information Systems Research*, 26(4): 695-713.

Amabile, Teresa M. (1983), "Brilliant but Cruel: Perceptions of Negative Evaluators," *Journal of Experimental Social Psychology*, 19 (2), 146-156.

Anderl, Eva (2013), "Analyzing Multichannel Online Customer Journeys for an Online Retailer: A Categorization Approach", *Working Paper*.

Anderl, Eva, Ingo Becker, Florian v. Wangenheim, Jan H. Schumann (2013), "Mapping the Customer Journey: A Graph-based Framework for Attribution Modeling in Managerial Practice," *SSRN*.

Anderson, Eric T. and Duncan I. Simester (2014), "Reviews Without a Purchase: Low Ratings, Loyal Customers, and Deception," *Journal of Marketing Research*, 51 (June), 249-269.

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

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

Berger, Jonah and Eric M. Schwartz (2011), "What Drives Immediate and Ongoing Word of Mouth?" *Journal of Marketing Research*, 48 (October), 869-880.

Berlyne D. E. (1971), *Aesthetics and Psychobiology*, New York, NY: Appleton-Century-Crofts.

Bettman, James R., Mary F. Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25(3): 187-217.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993-1022.

BrightLocal (2016), retrieved from https://www.brightlocal.com/learn/local-consumer-review-survey/.

Brown, Mark R., Roop K. Bhadury, and Nigel K. Pope (2010), “The Impact of Comedic Violence on Viral Advertising Effectiveness,” *Journal of Advertising*, 39(1), 49-65.

Bruce, Norris, B.P.S. Murthi, Ram C. Rao (2016), “A Dynamic Model for Digital Advertising: The Effects of Creative Formats, Message Content and Targeting Engagement,” Working Paper.

Casalboni, Alex (2016), “Google Vision API: Image Analysis as a Service,” retrieved May 15, 2016, from http://cloudacademy.com/blog/google-vision-api-image-analysis/.

Cheema, Amar and Andrew M. Kaikati (2010) "The Effect of Need for Uniqueness on Word of Mouth," *Journal of Marketing Research*, 47 (June), 553-563.

Chen, Zoey, and Nicholas H. Lurie (2013), “Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth,” *Journal of Marketing Research*, 50 (4), 463-476.

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

Cochran, W. G., (1968), “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies”, *Biometrics*, 24 (2), 295-314.

Cyr, Dianne, Milena Head, Hector Larios, and Bing Pan (2009), “Exploring Human Images in Website Design: A Multi-Method Approach,” *MIS Quarterly*, 33(3), 539-566.

De Los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012), “Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior,” *American Economic Review*, 102(6): 2955-2980.

Demirci, Ceren, Koen Pauwels, Shuba Srinivasan, and Gokhan Yildirim (2014), “Conditions for Owned, Paid and Earned Media Impact and Synergy”, *Working Paper*.

Elster, Jon and George Loewenstein (1992), “Utility from Memory and Anticipation,” in *Choice over Time*, ed. George Loewenstein and Jon Elster, New York: Russell Sage Foundation, 213-334.

eMarketer, 2015, “Instagram Will Top 100 Million US Users by 2018,” retrieved May 15, 2016 from http://www.emarketer.com/Article/Instagram-Will-Top-100-Million-US-Users-by-2018 /1012148.

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

Finn, Adam (1988), "Print Ad Recognition Readership Scores: An Information Processing Perspective," *Journal of Marketing Research*, 25(2), 168-77.

Gatignon, Hubert and Thomas S. Robertson (1986), “An Exchange Theory Model of Interpersonal Communication,” *Advances in Consumer Research*, Vol. 13, Richard J. Lutz, ed. Provo, UT: Association for Consumer Research, 534-38.

Ghose, Anindya, and Vilma Todri (2017), “Towards Digital Attribution: Measuring the Impact of Display Advertising on Online Consumer Behavior,” *MIS Quarterly*, forthcoming.

Ghose, Anindya and Sha Yang (2009), “An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets,” *Management Science* 50(10):1605-1622.

Godes, David and Jose C. Silva (2012), “Sequential and Temporal Dynamics of Online Opinion,” *Marketing Science*, 31(3), 448-473.

Goeree, Michelle S. (2008), “Limited Information and Advertising in the U.S. Personal Computer Industry,” *Econometrica*, 76(5): 1017–1074.

Golan, Guy J. and Lior Zaidner (2008), “Creative Strategies in Viral Advertising: An Application of Taylor’s Six-Segment Message Strategy Wheel,” *Journal of Computer-Mediated Communication*, 13(4), 959-972.

Goldfarb, Avi, and Catherine Tucker (2011), "Online Display Advertising: Targeting and Obtrusiveness," *Marketing Science*, 30(3): 389-404.

Goodman, Gail S. (1980), “Picture Memory: How the Action Schema Affects Retention," *Cognitive Psychology*, 12 (4), 473-495.

Greene, William H. (1994), “Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models,” New York University Department of Economics Working Paper EC-94-10.

Greene, William H. (2003), *Econometric Analysis* (5th ed.). New Jersey: Prentice Hall. Chapter 5.6.1. ISBN 0-13-066189-9.

Guo, Shenyang, and Mark W. Fraser (2015), *Propensity score analysis: Statistical methods and applications*, Thousand Oaks, CA: Sage.

Haan, Evert de, Thorsten Wiesel and Koen Pauwels (2014), “The Effectiveness of Different Forms of Online Advertising for Purchase Conversion in a Multiple-Channel Attribution Framework,” *International Journal of Research in Marketing*, 33(3), 491-507.

Hamilton, Rebecca W., Ann Schlosser, and Yu-Jen Chen (2017), "Who's Driving This Conversation? Systematic Biases in the Content of Online Consumer Discussions," *Journal of Marketing Research*, In-Press.

Hansen, Lars P. (1982), “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50 (4), 1029-1054.

Hausman, Jerry, Bronwyn H. Hall and Zvi Griliches (1984), "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52(4), 909-938.

Houston, Michael J., Terry L. Childers, and Susan E. Heckler (1987), "Picture-Word Consistency and the Elaborative Processing of Advertisements," *Journal of Marketing Research*, 24(4), 359-369.

Imbens, Guido W., and Jeffrey M. Wooldridge (2009) "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5-86.

Johnson, Eric J., Steven Bellman, and Gerald L. Lohse (2003), "Cognitive Lock-In and the Power Law of Practice," *Journal of Marketing*, 67, April 2003, 62-75.

Johnson, Garrett A., Randall A. Lewis, and Elmar I. Nubbemeyer (2015) "Ghost Ads: Improving the Economics of Measuring Ad Effectiveness," Simon Business School Working Paper, No. FR 15-21.

Kaplan, Andreas and Michael Haenlein (2010), "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons*, 53(1), 59-68.

Kerin, Roger A., Steven W. Hartley and William Rudelius (2015), *Marketing 12th Edition*, New York, NY: McGrawHill Irwin

KPCB (2015), "Internet Trends 2016 – Code Conference," retrieved May 15, 2016, from http://www.kpcb.com/blog/2016-internet-trends-report.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman and P.K. Kannan (2016), "From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior," *Journal of Marketing*, 80(1), 7-25.

Lee, Eunsun, Jungsun Ahn, and Yeo Jung Kim (2014), "Personality traits and self-presentation at Facebook," *Personality and Individual Differences*, 69 (October), 162-167.

Lee, Yih Hwai and Charlotte Mason (1999), "Responses to Information Incongruency in Advertising: The Role of Expectancy, Relevancy, and Humor," *Journal of Consumer Research*, 26(2), 156-169.

Lewis, David D. (1998), "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," *Proceedings of the European Conference on Machine Learning (ECML)*, 4-15.

Lewis, Randall A., and David H. Reiley (2014), "Online Ads and Offline Sales: Measuring the Effects of Online Advertising via a Controlled Experiment on Yahoo!" *Quantitative Marketing and Economics*, 12(3): 235-266.

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

Li, Hongshuang (Alice) and P.K. Kannan (2014), “Attributing Conversion in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment,” *Journal of Marketing Research*, 51(1):40-56.

Li, Jun, Nelson Granados, and Serguei Netessine (2014), “Are Consumers Strategic? Structural Estimation from the Air-Travel Industry,” *Management Science*, 60(9): 2114-2137.

Li, Yiyi, and Ying Xie (2017), “Is a Picture Worth a Thousand Words? An Empirical Study of Imagery Content and Social Media Engagement,” Working paper.

Loeb, Larry (2016), “Google Cloud Vision API Gives Glimpse into AI, Machine Learning,” retrieved May 15, 2016 from http://www.informationweek.com/software/enterprise-applications/google-cloud-vision-api-gives-glimpse-into-ai-machine-learning/a/d-id/1324400.

Luca, Michael (2011), "Reviews, reputation, and revenue: The case of Yelp.com," HBS Working Paper 12-016, Harvard Business School, Cambridge, MA.

Luca, Michael, and Georgios Zervas (2016), "Fake It till You Make It: Reputation, Competition, and Yelp Review Fraud," *Management Science*, 62 (12), 3412-3427.

Manchanda, Puneet, Peter e. Rossi, and Pradeep K. Chintagunta (2004), “Response Modeling with Non-Random Marketing Mix Variables,” *Journal of Marketing Research*, 41(3): 467-478.

McCallum, Andrew Kachites (2002), "MALLET: A Machine Learning for Language Toolkit," http://mallet.cs.umass.edu.

Miniard, Paul, Sunil Bhatla, Kenneth Lord, Peter Dickson and H. Rao Unnava (1991), “Picture-Based Persuasion Processes and the Moderating Role of Involvement,” *Journal of Consumer Research*, 18(1), 92-107.

Mitchell, Andrew A. and Jerry C. Olson (1981), “Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude?” *Journal of Marketing Research*, 18(3), 318-332.

Moe, Wendy and David Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31 (3), 372-386.

Moe, Wendy and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (June), 444-456.

Moorman, Christine (2016), “CMO Survey Report: Highlights and Insights”, retrieved in April 2016 from https://cmosurvey.org/wp-content/uploads/sites/11/2016/02/The_CMO_Survey-Highlights_and_Insights-Feb-2016.pdf.

Muchnik, Lev, Sinan Aral, and Sean J. Taylor (2013), "Social Influence Bias: A Randomized Experiment," *Science*, 341 (6146), 647-651.

Nerlove, Marc, and Kenneth J. Arrow (1962), “Optimal Advertising Policy under Dynamic Conditions,” *Economica*, 29(144):129-142.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), “Mine Your Own Business: Market-Structure Surveillance Through Text Mining,” *Marketing Science*, 31 (3), 521-543.

Palmer, Stephen E. (1999), *Vision Science: Photons to Phenomenology*. Cambridge, MA: A Bradford Book/MIT Press.

Pang, Bo, and Lillian Lee (2008), “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, 2 (1-2), 1-135.

Park, Young-Hoon, and Fader, P. S. (2004), “Modeling Browsing Behavior at Multiple Websites,” *Marketing Science*, 23(3): 280-303.

Peters, Kay, Yubo Chen, Andreas M. Kaplan, Björn Ognibeni and Koen Pauwels (2013), “Social Media Metrics — A Framework and Guidelines for Managing Social Media,” *Journal of Interactive Marketing*, 27(4), 281-298.

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

Petty, Richard E. and John T. Cacioppo (1986), "The Elaboration Likelihood Model of Persuasion," in *Advances in Experimental Social Psychology*, Vol. 19, ed. Leonard Berkowitz, New York: Academic Press, 123-205.

Petty, Richard E., John T. Cacioppo and David Schumann (1983), “Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement,” *Journal of Consumer Research*, 10(2), 135-146.

Pieters, Rik and Michel Wedel (2004), “Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects,” *Journal of Marketing*, 68(2), 36-50.

Pieters, Rik, Michel Wedel, and Rajeev Batra (2010), “The Stopping Power of Advertising: Measures and Effects of Visual Complexity,” *Journal of Marketing*, 74(5), 48-60.

Porter, Lance and Guy J. Golan (2006), “From Subservient Chickens to Brawny Men: A Comparison of Viral Advertising to Television Advertising,” *Journal of Interactive Advertising*, 6(2), 4-33.

Purnawirawan, Nathalia, Martin Eisend, Patrick De Pelsmacker, and Nathalie Dens (2015), “A Meta-Analytic Investigation of the Role of Valence in Online Reviews,” *Journal of Interactive Marketing*, 31, 17-27.

Raman, Kalyan, Murali K. Mantrala, Shrihari Sridhar, and Yihui (Elina) Tang (2012), "Optimal Resource Allocation with Time-varying Marketing Effectiveness, Margins and Costs," *Journal of Interactive Marketing*, 26:43-53.

Rao, Adithya, Nemanja Spasojevic, Zhisheng Li, and Trevor Dsouza (2015), "Klout Score: Measuring Influence across Multiple Social Networks," in *Proceedings of 2015 IEEE International Big Data Conference* - Workshop on Mining Big Data in Social Networks.

Rey, Benjamin (2014), "Your Tweet Half-Life is 1 Billion Times Shorter than Carbon-14's," retrieved May 15, 2016 from http://www.wiselytics.com/blog/tweet-isbillion-time-shorter-than-carbon14/.

Rosenbaum, Paul R., and Donald B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70 (1), 41-55.

Rossi, Peter E. (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science*, 33(5):655-672.

Rutz, Oliver J., and Randolph E. Bucklin (2012), "Does Banner Advertising Affect Browsing for Brands? Clickstream Choice Model Says Yes, for Some," *Quantitative Marketing and Economics*, 10(2): 231-257.

Schlosser, Ann E. (2005), "Posting Versus Lurking: Communicating in a Multiple Audience Context," *Journal of Consumer Research*, 32 (September), 260-265.

Schweidel, David A. and Wendy W. Moe (2014), "Listening in on Social Media - A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (August), 387-402.

Seiter, Courtney (2015), "The Secret Psychology of Facebook: Why We Like, Share, Comment and Keep Coming Back," retrieved May 15, 2016 from https://blog.bufferapp.com /psychology-of-facebook.

Shao, Xuhui and Lexin Li (2011), "Data Driven Multi-Touch Attribution Models," *KDD* '11: 258-264

Shugan, Steven M. (1980), "The Cost of Thinking," *Journal of Consumer Research*, 7, 2 (September), 99-111.

Sundaram, D.S., Kaushik Mitra, and Cynthia Webster (1998), "Word-of-Mouth Communications: A Motivational Analysis," *Advances in Consumer Research*, Vol. 25, Eric J. Arnould and Linda M. Scott, eds. Provo, UT: Association for Consumer Research, 527-531.

Sysomos (2009), "An In-Depth Look Inside the Twitter World," retrieved May 15, 2016 from https://sysomos.com/inside-twitter.

Tirunillai, Seshadri and Gerard J. Tellis (2012), "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," Marketing Science, 31(2), 19-215.

Tornquist, Stefan and David Tradewell (2012), "Marketing Attribution: Valuing the Customer Journey," Econsultancy, Retrieved July 13, 2015.

Toubia, Olivier and Andrew T. Stephen (2013), "Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?," *Marketing Science*, 32(3), 368-392.

Tucker, Catherine E. (2015), "The Reach and Persuasiveness of Viral Video Ads," *Marketing Science*, 34(2), 281-296.

Unnava, H. Rao and Robert E. Burnkrant (1991), "An Imagery-Processing View of the Role of Pictures in Print Advertisements," *Journal of Marketing Research*, 28(2), 226-231.

Vavrek, Adam (2012), "Image SEO: Pictures can Increase Your Readership [Photo from Research]" retrieved May 15, 2016 from http://www.skyword.com/contentstandard/marketing /image-seo-pictures-can-increase-your-readership-photo-from-research/.

WebLiquid (2011), "Facebook-Reach Analysis: Does Post Engagement Create Reach?," retrieved May 15, 2016 from http://www.emarketer.com/Article/What-Encourages-Facebook-Engagement/1008677.

Wedel, Michel and Rik Pieters (2015), "The Buffer Effect: The Role of Color When Advertising Exposures Are Brief and Blurred," *Marketing Science*, 34(1), 134-143.

Weitzman, M., (1979), "Optimal Search for the best Alternative," *Econometrica*, 47 (May), 641-54.

Wiesel, Thorsten, Koen Pauwels, and Joep Arts (2011), "Marketing's Profit Impact: Quantifying Online and Offline Funnel Progression," *Marketing Science*, 30(4):604-611.

Wilson, Theresa, Janyce Wiebe and Paul Hoffmann (2005), "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 347-354.

Winkelmann, Rainer (2008), *Econometric Analysis of Count Data*, Springer.

Wojnicki, Andrea C. and David Godes (2013), "Word-of-Mouth as Self-Enhancement," HBS Marketing Research Paper No. 06-01.

Xiao, Li, and Min Ding, "Just the Faces: Exploring the Effects of Facial Features in Print Advertising," *Marketing Science*, 33(3), 338-352.

Xu. Lizhen, Jason A. Duan and Andrew Whinston (2014), "Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion," *Management Science*, 60(6):1392-1412.

Yelp Inc. (2012), “A Picture’s Worth 1,000 Yelps,” retrieved in June 2017 from https://www.yelpblog.com/2012/06/a-pictures-worth-1000-yelps.

Zantedeschi, Daniel, Eleanor M. Feit, and Eric T. Bradlow (2016), “Measuring Multi-Channel Advertising Response Using Consumer-Level Data”, *Management Science*, Articles in Advance.

BIOGRAPHICAL SKETCH

Yiyi Li was born in Shanghai, China. After completing her schoolwork at Qibao High School in Shanghai in 2007, Yiyi entered Fudan University in Shanghai, China, and received a Bachelor of Business Administration with a major in Marketing in June 2011. She then entered the Management Science program of Naveen Jindal School of Management, at The University of Texas at Dallas in August 2011 to obtain her PhD in Marketing.

CURRICULUM VITAE

YIYI LI

EDUCATION

Ph.D.	Marketing, Naveen Jindal School of Management, The University of Texas at Dallas	**2017**
B.BA.	Marketing, School of Management Fudan University	**2011**

RESEARCH INTERESTS

Social Media, User Generated Content, Multi-channel Attribution, Online Advertising

DISSERTATION

Committee: Ying Xie (Chair), Brian T. Ratchford, Dmitri Kuksov, Zhiqiang (Eric) Zheng
Proposal Defended: May 12, 2016
"Essays on Consumer Response to User Generated Content and Online Advertising"

Essay 1: Is a Picture Worth a Thousand Words? An Empirical Study on Imagery Content and Social Media Engagement (Job market paper)
Essay 2: Imagery Content and Star Rating in Online Reviews
Essay 3: Modeling Multi-Channel Advertising Attribution across Competitors

WORKING PAPERS

Li, Yiyi, Ying Xie, and Zhiqiang (Eric) Zheng, "An Integrated Choice Model of Multi- Channel Advertising Effectiveness in Consumer Purchase Funnel," Revise and Resubmit at *MIS Quarterly*.

Li, Yiyi and Ying Xie, "Is a Picture Worth a Thousand Words? An Empirical Study on Imagery Content and Social Media Engagement," Revise and Resubmit at *Journal of Marketing Research*.

WORK IN PROGRESS

Li, Yiyi and Ying Xie, "The Power of Match: How Does Social Media Content Affect Sales?"

Li, Yiyi and Ying Xie, "Imagery Content and Star Rating in Online Reviews"

CONFERENCE PRESENTATIONS

"Imagery Content and Social Media Engagement"

38th ISMS Marketing Science Conference, Shanghai, China 2016

Discussant, 10th UT Dallas Frank M. Bass FORMS Conference, Dallas 2016

"Multi-Channel Advertising Effectiveness in Consumer Purchase Funnel"

10th CSWIM, Dalian, China 2016

UH Marketing Doctoral Symposium, Houston 2015

37th ISMS Marketing Science Conference, Baltimore 2015

5th Theory + Practice in Marketing (TPM), Atlanta 2015

HONORS AND AWARDS

Fellow, AMA-Sheth Foundation Doctoral Consortium 2015

Fellow, Marketing Science Doctoral Consortium 2015

Fellow, UH Marketing Doctoral Symposium 2014, 2015

Full scholarship at The University of Texas at Dallas 2011- 2016

Outstanding Undergraduate Thesis, Fudan University 2011

People's Scholarship (First Class), Fudan University 2010, 2011

TEACHING EXPERIENCE

Instructor, Principles of Marketing, UTD 2015, 2016

Evaluations: average 4.7/5.0

Teaching Assistant, Capstone Course in Marketing, UTD 2013, 2014

Teaching Assistant, Principles of Marketing, UTD 2012- 2014

Teaching Assistant, Capstone Marketing Decision Making, UTD 2013

Teaching Assistant, Marketing Analytics Using SAS, UTD 2012

TEACHING INTERESTS

Principles of Marketing, Internet Marketing, Social Media, Marketing Research

COMPUTER LANGUAGES

Data Analysis: SAS, Matlab, GAUSS Web

Scraper: Perl

Nature Language Processing: Python