FOLLOWING THE CYBERSPACE “BREADCRUMBS”: MODELING OPTIONS AND INTERACTIONS AMONG CONSUMERS, ADVERTISERS AND SEARCH ENGINE PROVIDERS

by

MING CHENG

A Dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Marketing Science Major

written under the direction of

Dr. S. Chan Choi

and approved by

____

____

____

____

Newark, New Jersey

May, 2015

ABSTRACT OF THE DISSERTATION

FOLLOWING THE CYBERSPACE “BREADCRUMBS”: MODELING OPTIONS AND INTERACTIONS AMONG CONSUMERS, ADVERTISERS AND SEARCH ENGINE PROVIDERS

By Ming Cheng

Dissertation Director: Professor S. Chan Choi

Sponsored search-engine advertising, as one of the dominant online marketing paradigms, has been widely adopted by many advertisers in the contemporary marketplace. Its success can be largely attributed to search engine providers’ ability to facilitate a connection between an individual consumer and firms advertising their products and services. In this dissertation, we develop two essays to examine two major aspects of the interactions between consumers and advertisers, as facilitated by a leading China’s search engine company.

In the first essay, we focus on examinations of the impact of various types of branded keywords on consumer click-through rates. Using a two-stage joint probability model, we find that various types of advertisers experience significantly different performance in the field of sponsored search advertising. The difference is not merely drawn from the search engine’s various ranking decisions, but also by consumers who differentiate from one kind to another. On the basis of these observations, we further drill into the explorations of branded keywords strategies applied to each type of

advertisers. The empirical insights generated under this theme could shed lights on advertisers with respect to branded keyword selections.

In the second essay, we study the impact of displayed paid ads assortment size and composition on consumer click-through behavior in sponsored search advertising. We apply a three-stage joint probability modeling approach to examine these relationships. Our empirical findings show that when the number of displayed ads increases, consumer's average click-through rate on individual paid ads decreases. Interestingly however, as the displayed ads list grows longer, consumer's click-through rate on "well-known" brands will significantly increases. With regard to the ads composition appeared in the search results, we find that the more perceivably attractive paid ads displayed at prominent ranks, the less likely that consumer will click on paid ads that shown below them. Our empirical results suggest that, in addition to their popularity with consumers, advertisers need to understand the effect of keywords on assortment size and click-through rate.

Preface

The thesis entitled “Following the Cyberspace ‘Breadcrumbs’: Modeling Options and Interactions among Consumers, Advertisers and Search Engine Providers” is prepared by Ming Cheng through her Ph.D. program from 2009 to 2015, at the Department of Supply Chain Management and Marketing Sciences (later changed as Department of Marketing) at Rutgers, the State University of New Jersey.

Acknowledgement

I would like to take this opportunity to express my deepest gratitude to my advisor, Dr. S. Chan Choi, and my committee members, Dr. Christopher K. Anderson, Dr. Lei Wang, and Dr. Sungjoon Nam, who have been devoting enormous amount of time and effort in guiding me through my entire Ph.D. study, giving me generous support, great advice and encouragement. They have set such great examples and their influence will go a long way into my career and life.

I would also like to express my sincere appreciation to Dr. Rohit Verma, who offered me a visiting scholar opportunity so that I can conduct my research at Cornell University for more than three years. In addition, I want to thank Dr. Vithala R. Rao for his generous help and great inputs, which formed the foundation of the dissertation.

Finally, I want to thank my family and all of my dearest friends for always being by my side, loving me and supporting me for all the time they've been in my life.

Table of Contents

List of Tables

List of Illustrations

CHAPTER 1 INTRODUCTION

Sponsored search advertising, also called paid search, is one of the leading online advertising approaches in the contemporary marketplace and has been widely adopted by a tremendous amount of advertisers. It is usually undertaken by advertisers paying certain amount of money to search engines for ads to be displayed in the search results webpage.

The proliferation of this online advertising scheme is mostly driven by the unique value provided by search engines, which act to intermediate between consumers and online advertisers. On one side, search engines remove information asymmetry and are viewed as one of the most prevalent information sources by consumers (Liaw and Huang 2003; Xiang, Wöber and Fesenmaier 2008). Take Google as an example, on average, it processes approximately 40,000 consumer searches per second, which translates to over 3.5 billion search queries per day and about 1.2 trillion search queries per year worldwide. On the other side, by directly and appropriately responding to each explicit consumer query, search engines provide advertisers with a powerful mechanism to connect targeted consumer segments in the marketplace. According to a report from Statista[1], in 2015, the total amount of spending on paid search advertising worldwide is expected to reach $ 54.9 billion US dollars.

Three entities are involved in the sponsored search advertising activities: consumers, advertisers, and search engine providers. Generally, each search is initiated when a consumer typed a series of text strings, defined as “search query”, in the search box located in the middle of a search engine’s front page. Paid ads displayed in response to

[1] Statista is an online statistics company, which offers survey results and statistics from more than 18,000 sources.

consumer's specific query oftentimes are shown on the right side of the screen or on the left side of the screen right above the organic search results section (See Figure 1.1). Once a consumer clicked on a link of a paid ad, s/he will be redirected – by search engines – to the advertiser's designated website.

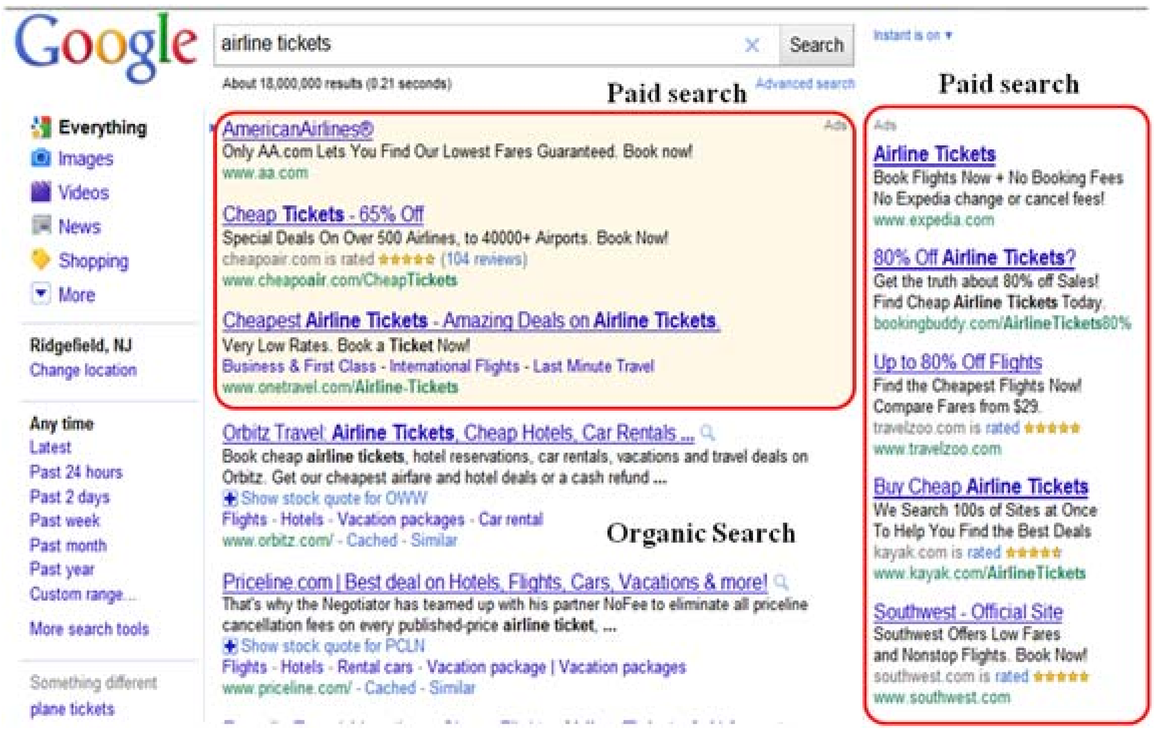

Figure 1.1 Screen Layout for Paid Search and Organic Search Sections

An advertiser, on the other hand, has to deliberately consider two major issues for its sponsored ads eventually being shown in the search result webpage. One critical issue is to select appropriate keyword sets. Usually, each advertiser has a unique set of keywords. Each keyword can be designed as a single word, a combination of words or a short phrase, which is tied to a specific landing page and is pre-set into one of the three major keyword match types (e.g. accurate, phrase or broad type, see definition in Table 1.1).

Table 1.1 Definition of Keyword Match Type

Match Type	**Definition**	**Example**
Accurate Keyword	Advertiser's paid ads will display in the search results only when consumer's search query exactly matches advertiser's selected keyword – with same words (or phrases) and in the same order.	Advertiser set keyword "cheap economical hotel" as accurate keyword match, the corresponding ad will only display in the search results <u>if and only if</u> a consumer enters the exact same phrase "cheap economical hotel".
Phrase Keyword	Advertiser's paid ads will display in the search results so long as consumers' search queries cover the entire keyword. The only condition is that the overlap between consumer query and keyword should be in the exact same order.	Advertiser set keyword "cheap economical hotel" as phrase keyword match, the corresponding ad will display in the search results when a consumer enters query such as "cheap economical hotel near the airport", or "where can I find a cheap economical hotel". However, the paid ad will <u>NOT</u> display if consumers enter queries in different orders, such as "economical hotel cheap".
Broad Keyword	Advertiser's paid ads will display in the search results so long as there is at least one word included in the consumer's query matches advertiser's keyword.	Advertiser set keyword "Cheap economical hotel" as broad keyword match, the corresponding ad will display in the search results when a consumer enters query such as "Cheap hotel", or "economical hotel", or "hotel".

Another important issue for advertisers to consider is to effectively plan for the bidding schemes for participations in a so-called "generalized second-price" auction process[2], through which advertiser's paid ads will be assigned to the search results webpage and ads ranking positions will be determined by the search engine. Advertisers

[2] "Generalized second-price" auction is an auction process where an advertiser is charged based upon the next bidder's bidding price.

only pay fees to search engines when their displayed paid ads get clicked by consumers – a payment mechanism called "pay per click".

Search engines, which act as intermediaries between consumers and advertisers, employ complex algorithms to determine which advertisers' sponsored ads, how many ads and in which order that ads will be shown in the search result webpage. Due to the "pay per click" payment mechanism, search engines need to appropriately and promptly return paid ads in an effort to stimulate consumer click-through and approach to the goal of revenue maximization. Therefore, bids, ads quality, and the level of "match" between keywords and consumer queries become three major concerns when search engines make decisions regarding ads display and ranking positions. Meanwhile, search engines keep track of detailed consumer search "footprints", such as collecting consumer search queries, cookies, and clickstream information. It also needs to provide performance reports to individual advertiser on a daily basis (e.g. Google Analytics). Click-through rate[3] (CTR) and conversion rate[4] are two important indicators which are regularly measured by advertisers to evaluate the effectiveness of their sponsored search advertising campaigns.

Although several studies have been developed to explore the inter-relationships among these three actors from different perspectives using various methodological tools (e.g. Ghose and Yang 2009; Song and Mela 2011), there are still many aspects left unexplored which are considered crucial issues when it comes to advertiser's decision-

[3] Click-through rate: it is an indicator which can be calculated as the ratio of the total number of clicks to the total number of impressions.

[4] Conversion rate: it is an indicator which can be calculated as the ratio of the total number of purchases happened at an advertiser's website to the total number of visits to advertiser's websites

making with regard to its sponsored search advertising campaigns. For instance, how would advertisers employ different keyword branding strategies to attract more customer traffic to their own websites? When seeing a list of paid ads displayed in the search result webpage, how would consumers react to those ads – in terms of the click-through? And how would advertisers transform these insights into managerial tactics? In other words, what kind of strategies advertisers could develop such that it could help advertisers themselves benefit from adopting search engine advertising? These are all important questions that have not been thoroughly studied. And to solve them, a deeper and clearer understanding regarding the inter-plays among consumers, advertisers and search engine providers is indispensable. After all, a search is initiated and ended by a consumer. Consumers are the ones who get to decide which displayed paid ads to click. To attract consumers to click on certain ads, advertisers involved in sponsored advertising campaigns can only depend upon 1) making higher bids – in the hope that ads shown in the search results might be placed at more prominent rankings (e.g. closer to the top of the screen) and might get more of the customers' attention, which would eventually lead to click-through; or 2) choosing "right" keywords, which would help advertisers connect to the "right" consumers who might show greater interests and would prefer to click on certain advertisers' ads. Finally, a search engine is in a leverage position where it needs to provide benefits to both consumers (e.g. return information as consumer requested) and advertisers (e.g. connect to preferred customer segments) (e.g. Song and Mela 2011), because its revenue drawn from the "pay-per-click" mechanism largely depend upon 1) consumers' actual clicks, which determines the number of clicks that a search engine receives; and 2) the number of advertisers that actually participated in a search engine's auction process, which determines the expensiveness for bidding on specific keywords.

The dissertation is devoted to examining the interactions among consumers, advertisers and search engine providers through analyzing the "breadcrumbs" – which is defined as online search and click-through information generated by consumers in the sponsored search advertising domain. The goal of this dissertation is to develop empirical insights into the dynamic interactions in sponsored search advertising area and to provide advertisers with actionable managerial tactics to succeed in competitive search engine campaigns. In particular, we focus on advertisers' keywords branding and design issue in the sponsored search advertising context. Two essays are developed under this theme and are conducted based upon analysis and examinations using a unique dataset collected from one of China's leading search engines.

In the first essay, we focus on examinations of the impact of various types of branded keywords on consumer click-through rates. Using a two-stage joint probability model, we find that various types of advertisers experience significantly different performance in the field of sponsored search advertising. The difference is not merely drawn from the search engine's various ranking decisions, but also by consumers who differentiate from one kind to another. On the basis of these observations, we further drill into the explorations of branded keywords strategies applied to each type of advertisers. The empirical insights generated under this theme could shed lights on advertisers with respect to branded keyword selections.

In the second essay, we study the impact of displayed paid ads assortment size and composition on consumer click-through behavior in sponsored search advertising. We apply a three-stage joint probability modeling approach to examine these relationships. Our empirical findings show that when the number of displayed ads increases, consumer's average click-through rate on individual paid ads decreases, while increases on paid ads that are top ranked or belong to well-known brands. We also find that the more perceivably attractive paid ads displayed at prominent ranks, the less likely that consumer will click on paid ads that shown below them. Our empirical results suggest that, in addition to their popularity with consumers, advertisers need to understand the effect of keywords on assortment size and click-through rate.

The dissertation contributes to the body of literature in three aspects: First, we fill the theoretical gap by looking into two topics: 1) examining keyword branding issue and 2) analyzing the displayed paid set's impact on consumer click-through and its impact on advertiser's keyword design issue (e.g. choose more specific keywords versus popular keywords), which have not been investigated or discussed in the previous search engine marketing literature. Second, unlike most of the empirical studies which mainly focus on and draw conclusions based upon a single advertiser, we focus on our analysis in hospitality industry and conduct studies which entail examinations across hundreds of advertisers. In a sense, we could paint a more holistic picture with respect to the diverse search engine performance across the entire industry and provide advertiser with different strategies based upon its varieties. Moreover, our analysis are conducted using disaggregate consumer query-level data, which may to some extent remove system bias generated from using aggregate-level data as highlighted by Abhishek, Hosanagar and Fader (2012). Finally, given that the dataset used in this dissertation is a collection from a search engine company, it compiles advertiser's search engine portfolios as well as consumers' clickstream data. That said, for each consumer search query, we are able to observe all displayed advertisers' ads listed in the search results – a snapshot of the entire paid search section, which includes a completed "picture" with respect to the keywords, their related attributes (e.g. match type, bids, ads quality), and consumer click information. Thus, it provides us with a unique opportunity to investigate advertiser's potential search engine strategies while taking into account other co-existed advertisers' performance. All of these perspectives differentiate our studies from the previous literature.

The remainder of the dissertation is structured as below: Chapter 2 provides an overview of the related sponsored search advertising literature. The literature is unfolded into two mainstream developed in this field: empirical and analytical studies. Chapter 3 provides an overview of search engine provider and the dataset we used for conducting the dissertation. Chapter 4 is a composition for the first essay – a study in which we focus on analyzing advertiser keyword branding issue. Chapter 5 is a composition of the second essay, in which we conduct a study to analyze the displayed paid ads set's impact on consumer click-through and its impact on advertiser keyword design/selection strategies. In Chapter 6, we conclude and provide a discussion of potential future research projects which will be conducted along this path.

CHAPTER 2 LITERATURE REVIEW

In this chapter, we provide an overview of the related sponsored search advertising literature. Two sections are included. In Section 2.1, we discuss the related literature from two branches: empirical research and analytical research. In Section 2.2, we examine the common characters embedded in the literature and discuss how this dissertation would differentiate, as well as contribute to the search engine marketing's body of knowledge.

2.1 Literature in Sponsored Search Advertising

In this section, we bring an overview of the related sponsored search advertising literature from two mainstreams. Two sub sections are developed accordingly. In Section 2.1.1, we summarize the related literature from the empirical field. In Section 2.1.2, we provide a summary of the search engine marketing literature from the analytical aspect.

2.1.1 Empirical Research

In the empirical world, studies related to sponsored search advertising can be classified into three categories: 1) the impact of rank (ads displayed location in search results) on consumer click-through and conversion rates; 2) Examination of the inter-relationship between paid ads and organic search results; 3) The role of keywords and ads description's impact on consumer click-through and conversion rates.

1) The impact of rank on consumer click-through and conversion rates

In sponsored search advertising, ads placement plays an important role which determines the actual consumer click-through that an advertiser could receive through search engine platforms. Several research papers have shown and validated that consumer's click-through rate is exponentially decreased as sponsored ads appeared from the topmost position to the bottom (Goodman 2006; Aggarwal, Feldman and Muthukrishnan 2007; Feng et al. 2007). Thus, traditionally, both researchers and practitioners have treated ads placement (or ads ranking) as a critical indicator when evaluating the effectiveness of advertiser's sponsored search campaigns, which even formed a role of thumb in the industry, stated as "bid all the way to the top", because practitioners believed that the top-most position is equivalent to the profit-maximizing position. However, the belief of "bidding to the top" or "location as the most important key element in sponsored advertising" has been proved to be less efficient or inaccurate by Ghose and Yang (2009), Agarwal, Hosanagar and Smith (2011) and Jerath et al. (2012).

Ghose and Yang (2009) proposed a set of simultaneous equations and employed the hierarchical Bayesian methodology to examine the interactions among search engine ranking decision, consumer click-through and conversion decision. They found that, consistent with the traditional literature, consumer's click-through rate is negatively associated with the ads positions. However, they also highlighted a key finding in which ads displayed at more prominent locations may not be the most profitable ones from the company's perspective, whereas ads displayed in the middle position – rather than the top or the bottom – turned to be the more profitable ads.

Consistent with Ghose and Yang's findings, Agarwal, Hosanagar and Smith (2011) conducted a similar study using data generated from a field experiment of an online retailer ads campaign. They found that, although consumer click-through rate is decreased with positions, the conversion rate moves in the opposite direction as ranks go from the topmost to the bottom. They have also highlighted the importance role played by ads' keyword, as they stated that consumer conversion rate is even higher for some specific keywords.

Jerath et al. (2012) further pushed the boundary by proposing a "position paradox" issue in sponsored search advertising, where they categorized companies into two kinds: superior and inferior companies. In the paper, they found that inferior companies are more demanding in locating at more prominent positions thereby they may act more aggressively in the auction process, whereas for superior companies, given that their click-through rates can be maintained at a high level even in the middle positions, therefore they are are more reluctant in devoting more money resources in the bidding process.

Recent literature has shifted its focus from the examination of impact of rank on advertisers' conversion into the analysis of rank impact on search engine's revenue, as exemplified by Ghose, Ipeirotis and Li (2012 & 2014). Ghose Ipeirotis and Li (2012) designed a new ranking system for travel search engines by taking into account consumers' multidimensional preferences for certain products or services which has been overlooked in the traditional ranking system design. Moreover, on the methodology side, the study also set a good example by integrating the information system techniques into the marketing analytics domain – using text-mining approach to explore unstructured user-generated and crowdsourced content while examining consumers' multidimensional preferences. Ghose Ipeirotis and Li (2014) examined three kinds of search engine ranking mechanisms and they found that product search engines should 1) incorporate "signals" from social media in ranking system design, which would generate a positive impact on its revenue; and 2) control the amount of information provides to its consumers due to the fact that information overload may lower the consumers purchases, which may further drive down search engines' revenue.

2) The role of keywords and ads description's impact on consumer click-through

Given that ads placement is not the only influential factor which determines the effectiveness of advertiser's ads campaigns, another branch of the empirical literature has emerged wherein researchers mostly focus on examinations of the role of keywords and its impact on consumer click-through and conversion rates (Rutz and Bucklin 2011; Rutz and Trusov 2011; Jerath et al. 2014).

Rutz and Bucklin (2011) developed a dynamic linear model to examine the impact of keywords on consumer click-through rate using a dataset collected from a major lodging chain. They found that in sponsored search advertising environment, there is an asymmetric spillover effect existed between two types of keywords categorized as "branded" and "generic". This asymmetric spillover effect can be demonstrated as generic keyword search may increase the probability of future branded search activities, but not vice versa.

Aside from analyzing keywords through its "branded vs. generic" character, Jerath, Ma and Park (2014) took keywords popularity into consideration and examined its impact on consumer search and click behavior. In the paper, they found that ads with less popular keywords equipped tend to receive higher consumer click-through rate than ads embedded with popular keywords in the sponsored search advertising environment. They also pointed out that consumers who initiated searches with less popular keywords might be closer to their purchases therefore shall be labeled as preferred customers in sponsored search.

Rutz and Trusov (2011) moved beyond the analysis of the impact of keywords (branded versus generic, more popular versus less popular) and they explored into the issue of how ads description appeared in the search results would affect consumers' click-through and conversion rates. To do that, they built a two-stage consumer level model and examined the model using a dataset from a ringtone company and they developed a novel framework to analyze elements (or key attributes) of sponsored ads.

3) The inter-relationship between paid ads and organic search results

The third branch of empirical research conducted under sponsored search advertising theme is the analysis of inter-relationships between paid ads and organic search results, given that these two sections are generally appeared concurrently. However, since it is usually difficult to collect data displayed in the organic section, the number of research developed in this field is quite limited. Yang and Ghose (2010) examined the interdependence between paid search and organic search results using the hierarchical Bayesian Monte Carlo method. They found that consumer click-through on organic listings is positively interdependent on click-through on paid search, meaning that with the appearance of paid section, consumer's click-through rate on organic listing is significantly higher, compared with the situation where there is no paid ads displayed alongside the organic listing. Jerath et al. (2014) studied the impact of keyword popularity on consumer click-through behavior in these two sections and they found that consumers who seek information using more popular keyword terms tend to click on organic search results rather than paid search ads, which lead to significantly low click-through rates in the paid search section.

2.1.2 Analytical Research

In the analytical world, research developed in sponsored search advertising field is evolved through three stages: 1) analysis of position-based auction, bidding mechanism, equilibrium and price competition, 2) examination of the impact of click fraud on search engine's revenue, and 3) exploration of the two-sided market in sponsored search advertising.

1) Position-based auction, bidding mechanism, equilibrium and price competition

The emergence of the search engine literature is actually stemmed from the studies conducted in this field – which focused on examining different types of auction and bidding mechanisms involved in sponsored search advertising process and finding optimal pricing equilibrium under different circumstances. Especially given the unique character in which analytical studies don't heavily rely on empirical data, a majority of the analytical studies have been developed in this field (Aggarwal, Feldman and Muthukrishnan 2007; Varian 2007; Edelman, Ostrovsky and Schwarz 2007; Athey and Ellison 2008; Katona and Sarvary 2010; Chen and He 2011; Yalcin and Ofek 2011; Xu, Chen and Whinston 2011).

Aggarwal, Feldman and Muthukrishnan (2007) studied and presented a pricing mechanism in which advertisers can not only specify their bids but also post location-specific requirement before participating in the search engine auction process. They've also pointed out the unique characters embedded in the proposed pricing mechanism, compared with the traditional auction mechanism called "Vickrey-Clarke-Groves" (VCG) mechanism. Edelman, Ostrovsky and Schwarz (2007) further drilled down into this area by comparing and contrasting the VCG and “generalized second price” mechanisms (which is widely employed by many search engines nowadays), and they highlighted that “generalized second price” mechanism contains several features which are not included in the VCG mechanism. Following the similar thread, Varian (2007) examined this “generalized second price auction” mechanism – the paid ads auction mechanism which is undertaken by Google and Yahoo! and found that the equilibrium of ad auction can be explicitly calculated. Katona and Sarvary (2010) examined the auction process by incorporating two extra components: 1) the interaction between organic listing and paid search ads; and 2) the intrinsic differences in click-through between sites in their model and they pointed out that these two factors may generate great impact on advertisers’ bidding behaviors, which further affect the equilibrium of prices of paid search ads.

Up until this point, the studies described above are all “auction process” driven. Taking into consideration the real “object” (the consumers) that search engines always need to interact with, Athey and Ellison (2008) conducted a study which specifically addressed the issue of how the paid search auction may affect the overall welfare, and how this surplus would be distributed among consumers, advertisers and search engines. Chen and He (2011) also developed a study along the similar line. The difference, compared with Athey and Ellison (2008), is that consumers are uncertain about the value of the product provided by the companies as they initiated their search and can learn progressively throughout costly search. The study provided some interesting findings in which the equilibrium shows that search engines tend to assign more relevant paid ads to more prominent positions. At the same time, advertisers have more incentives to bid higher prices to get to the topmost locations if consumer's search is more relevant.

2) Click fraud on search engine's revenue

One of the critical issues that is oftentimes raised by many advertisers is that what percentage of the clicks generated from search engine platforms are fraudulent – clicks that are done by the advertiser's competitors or some third parties. Because when deceptive clicks happen, it won't create any value towards advertiser's conversion, rather it might end up costing advertisers pay a significant amount of money to search engines. Usually, Advertisers tend to think that search engines might be reluctant to prevent this type of "click fraud" from happening or might even prefer it, because advertisers believe that search engines may actually benefit from the phenomena of "click fraud" due to its "pay per click" payment mechanism.

To test this statement, Wilbur and Yi (2009) specifically explored in this field and they have found that whether or not search engines would benefit from "click fraud" depends upon the level of competition existed in the sponsored search advertising domain. When the market is more competitive, advertisers might tend to lower their budget constraints, which would further drive down search engine's revenue if "click fraud" do happen. However, if the marketplace is less competitive, search engine will actually get a "bonus" from other entities conducting fake clicks – therefore, they suggested that there should be a third party to audit the actual click sources.

3) Two-sided market in sponsored search advertising

There is a stream of search engine marketing literature emerged which bridges in between empirical and analytical fields as exemplified by Weber and Zheng (2007) and

Song and Mela (2011). These studies are mainly focusing on examination of the effect of the two-sided market phenomena as intermediated by search engines. Weber and Zheng (2007) developed a two-stage model which captures both demands from consumers and advertisers. They also discussed several issues with respect to search engine's design mechanism for selling more proper ads to companies, and characterized advertisers' bidding strategies based upon consumers' equilibrium search behaviors. Song and Mela (2011) developed a dynamic model to capture the interactions between consumers, advertisers and search engines.

2.2 Discussion

Since the studies we conducted in the dissertation both fall into the empirical category, in this section, we discuss how our studies differentiate with previous empirical literature and could contribute to the sponsored search advertising's body of literature.

To begin with, we create a table which summarizes the empirical literature in the paid search field (See Table 2.1). As can be seen from the table, most of the empirical studies were conducted using aggregate-level data collected from a single advertiser. If we further extract the main research theme by creating a flow chart which captures the key activities undertaken in the paid search domain (as illustrated in Figure 2.1), we can see that most of the research is developed by targeting at examinations of consumer click-through and conversion rate – the second half of the flow chart, rather than investigating how consumers might respond to different ads based upon different queries and ads appearance – which is the first half of the flow chart.

Table 2.1 A Summary of Empirical Literature in Sponsored Search Advertising

Author	Year	Field of Analysis	Methodology	Data Used
Ghose and Yang	2009	Consumer click-through and conversion rates	Hierarchical Bayesian Modeling Markov Chain Monte Carlo (MCMC) estimation	Aggregate-level data from a retailer
Yang and Ghose	2010	Interdependence between paid search and organic search	Hierarchical Bayesian Modeling	Aggregate-level data from a retailer
Agarwal, Hosanagar and Smith	2011	Consumer click-through and conversion rates	Hierarchical Bayesian Modeling	Field experiment from ads campaigns
Ghose, Ipeirotis and Li	2012	Design ranking system	Text-mining Crowdsourcing	Data from travolocity.com
	2014	Search engine's ranking systems	Hierarchical Bayesian Modeling MCMC estimation	
Rutz and Buckin	2011	keyword effect (branded vs. generic)	Dynamic linear model Hierarchical Bayesian	Aggregate-level data from a major lodging chain
Rutz and Trusov	2011	Impact of ads description on click-through and conversion rates	Hierarchical Bayesian Modeling	Disaggregate consumer-level data from a ringtone company
Jarath, Ma and Park	2014	Role of keyword popularity	Choice Model Bayesian Approach MCMC estimation	Data from a search engine company

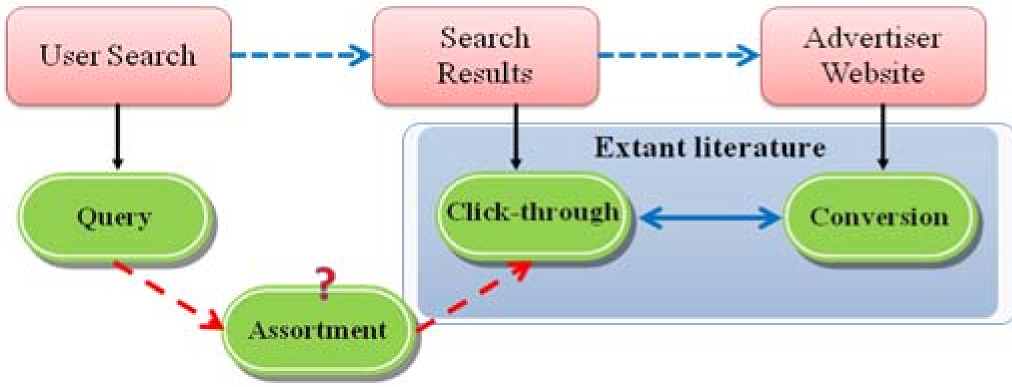

Figure 2.1 A Flow Chart of the Key Activities undertaken in Sponsored Search Advertising

Therefore, our studies are developed to fill the theoretical gap in the literature. By analyzing data collected from a leading search engine company in China, we are able to overcome some of the limitations that cannot be examined in the literature. For instance, we may to some extent reduce the system bias generated by employing aggregate-level data. Moreover, we are able to compare consumer's different click behavior across different types of advertisers and explore how consumers might respond paid ads given the variation of the ads shown in the search results.

CHAPTER 3 PROBLEM SETTING AND DATASET

In this chapter, we introduce the problem setting and dataset we used for conducting the two essays that we will discuss in the following two chapters (Chapter 4 & 5). Four sections are included in this chapter. In section 3.1, we briefly introduce the search engine provider, which is the source of the dataset we used for examinations of the two studies included in the dissertation. In section 3.2 and section 3.3, we drill down into some detailed explanations with regard to the industry we selected and specific location of displayed paid ads (e.g. ads displayed on the left-side of the screen versus right-side of the screen) we used in our analysis. In section 3.4, we discuss the detailed data generating process and set definitions for most of the attributes embedded in the modeling process.

3.1 The Search Engine Provider

The studies are conducted using a large-scale dataset collected from one of the leading China's search engine companies ("Baidu"). According to online reports published during 2013 and 2014 by CNNIC and iResearch, the search engine controls a great proportion of China's online search market, which has a market share up to 60% of the entire market. Among the 490 million online searchers existed in China's online search market, 90% of the users have experience using Baidu to retrieve information. Meanwhile, 70% of all the online searchers prioritize Baidu as their first choice when it comes to decide which search engines to use. Therefore, it is considered as a search giant with equivalent influential power as Google in China's online search market.

There are some similarities and fundamental differences embedded in Baidu, compared with other search engines which have often been discussed in the literature (e.g.

Google). First, with regard to the sponsored search layout, similar to Google, Baidu provides two sections ("paid search" and "organic search") with similar locations – where paid ads appear either on the right-side of the screen or on the top left-side of the screen. The only difference in terms of the layout is that, unlike search engines such as Google, Yahoo! and Bing which generally apply conspicuous colors or signs to differentiate paid ads from organic search section, the search engine in question planted a footnote called "sponsored ad" planted at the corner of each displayed sponsored ad (See Figure 3.1 and Figure 3.2).

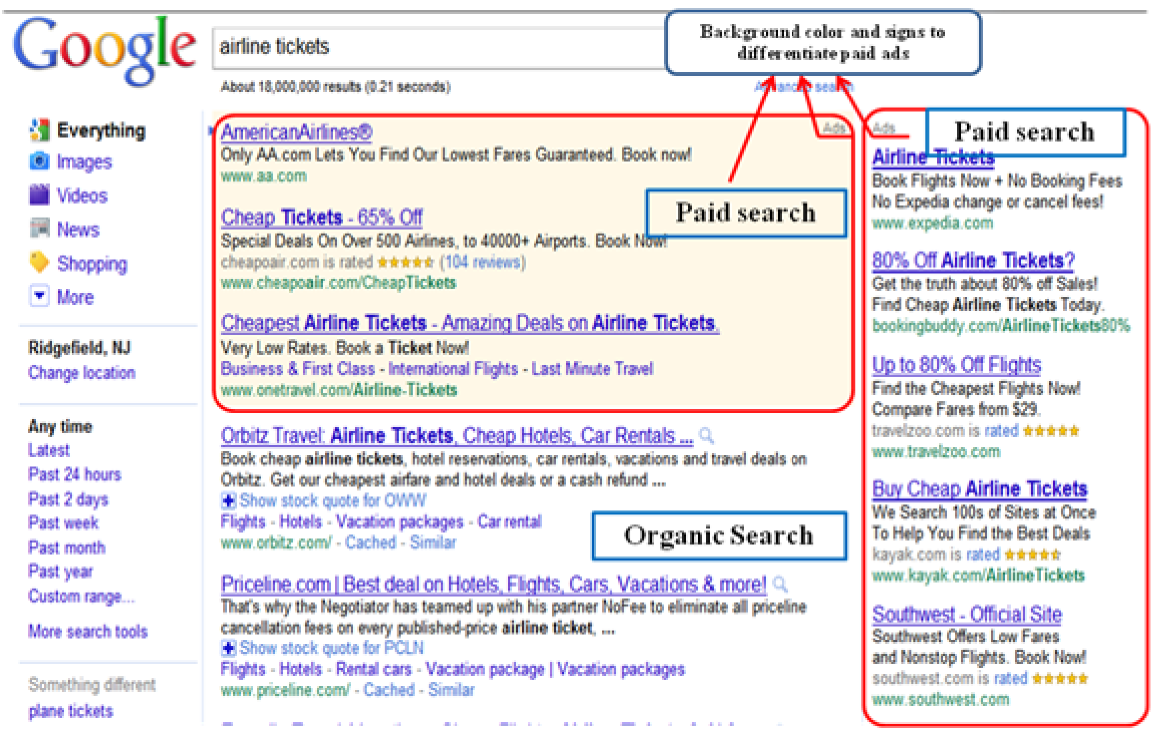

Figure 3.1 Paid Ads Screen Layout in Search Results Webpage (Example of Google)

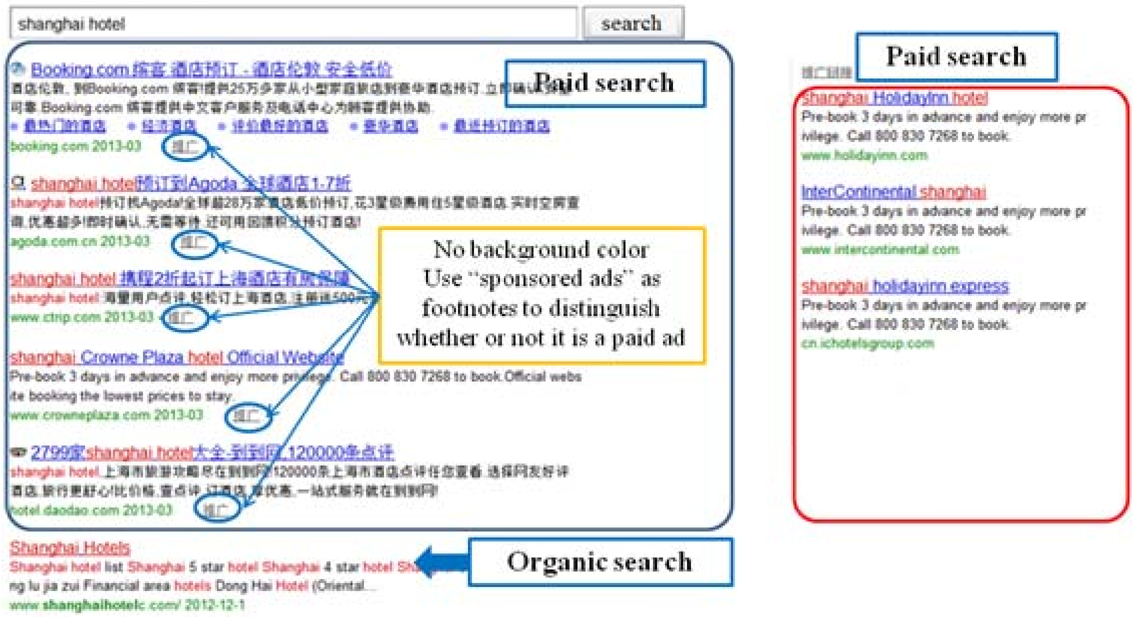

Figure 3.2 Paid Ads Screen Layout in Search Results Webpage (Example of Baidu)

Second, with regard to the number of ads displayed in the search results, similar to Google, Baidu has no restrictions on the number of ads displayed on the right-side of the screen. That said, paid ads shown on the right-side of the screen can be listed starting from page 1 to page n, which mostly depend upon how many advertisers participated in bidding process for certain keywords. However, the situation is quite different for ads displayed on the left-side of the screen, compared with Google and Baidu. For instance, Google generally provides three slots at maximum to the left-side of the screen and this is quite consistent for each page (starting from page one). Whereas, for Baidu, ads displayed on the left-side of the screen have up to ten slots and only appear within the first page. The reason of this configuration is because of the significant low consumer click-through rates generated from second page onwards.

3.2 Paid Ads Displayed on the "Left-Side"

The samples we used for conducting the two studies are ads displayed on the left-side of the screen. As described in Section 3.1, each ad is assigned to a slot numbered from one to ten with locations ranging from the topmost to the lower-middle of the screen. Our justification for this confinement lies in three aspects: First of all, as it supported by Lorigo et al. (2008), research has shown that customers devote most of their attentions to the left-side of the screen. In addition, many consumers nowadays are aware that search results shown on the right-side are sponsored ads. As stated by Raman (1997), consumers are reluctant to click on links that are developed for the purpose of advertising. Second, unlike the search engines been studied in the previous literature (e.g. Google, Yahoo! and Bing), the search engine in question substitutes conspicuous colors or signs – which is used for differentiating paid ads section and organic section – with a footnote called "sponsored ad" planted at the corner of each displayed ad's description. Under this circumstance, most consumers are not aware that the clicked search results are actually paid ads (Vuylsteke et al. 2010). So, it may to some extent reveal consumers' "real" click behaviors. Finally, there is a fundamental difference embedded in the slot settings for different search engines. For Google, it usually assigns eleven ads at maximum on the first search results webpage, with three ads appeared to the left and eight ads displayed on the right. The fact that many studies solely consider ads on the right-side of the screen is stemmed from the limited number of ads displayed on the left-side (with only 3 slots). In comparison, due to the "pay-per-click" mechanism and given that consumers prefer to click on search results appeared on the left-side than the right-side, the search engine examined in our study assigned ten slots at most to the left-side, leaving only two ads on average to the right-side, which has the slot configuration that is exact the opposite as Google's. Thus, all of these stated above contribute to our decision for collecting ads displayed on the left-side of the screen.

3.3 Choosing Data from "Hospitality Industry"

We conduct the dissertation based upon examinations of advertisers and sponsored search activities that are undertaken in the China's hospitality industry. Several considerations were made during the "industry selection" decision-making process. Aside from considering the actual computation capability that a personal computer could carry, we also consider conducting research while we could get references from the literature where similar situations might be discussed. From the search engine marketing literature, we have found several studies focusing on solving sponsored search advertising issues within hospitality industry (e.g. citations). And, as we further expand our analysis and drill down into this industry, it turns out that it is not a coincidence for many researchers choosing hospitality industry while developing search engine marketing related research. Here is why:

First, customers heavily rely on search engines while searching and collecting information in the hospitality or tourism industry. As it supported by the work done by Xiang, Wöber and Fesenmaier (2008), search engines are playing increasingly important role in facilitating information exchange between online consumers and online tourism domain. Search engines many to some extent ease and shorten consumer information search procedure by providing customers with many options and connecting customers with different travel agents, such as tripadvisor.com, priceline.com, where customers could get all kinds of travel-related information spanning from airline ticket, hotel booking to car rentals. Not to mention that search engines could also facilitate the transactions between customers and travel agents, as highlighted in the study developed by Anderson (2011), more than 80% of travel related online purchases are proceeded by some form of search.

Another important reason that we chose hospitality industry to conduct our sponsored advertising studies is driven by the maturity of online platforms developed in other industries (e.g. commodity, regular merchandise industry). A number of well-developed online retail platforms (e.g. amazon.com, eBay.com, taobao.com) already exist in the marketplace. That said, when a customer is thinking about buying a book, s/he might prefer directly go to amazon.com and search for the book rather than going to a search engine, unless s/he couldn't remember the exact web link of the website.

3.4 Data Generating Process

The dataset contains advertiser sponsored search engine performance and consumer click-stream data for two-month period (50 days) starting from January 2012. As we described in Section 3.2 and 3.3, the dataset is collected from the leading Chinese search engine and related to sponsored search advertising activities within hospitality industry. In this dataset, two types of advertisers are involved: 1) hotel website, such as Hilton.com or Marriott.com, which is defined as a combination of individual hotels, hotel management companies, brands and ownership groups. It runs its own official websites and uses search engine as one of the advertising channels to attract customer visits to its own websites; and 2) online travel aggregator (OTA), such as booking.com or expedia.com, which is defined as a travel search engine that aggregates travel information (e.g. airline tickets, car rental, or hotel reservation) from multiple online sources. In this research, we narrow this definition to travel aggregators which attempt to sell hotel room (for commissions) on behalf of multiple hotel companies.

The data generating process of sponsored advertising activities starts when a consumer typed a hospitality related search query (e.g. "Hilton hotel in Shanghai"). Followed by the customer's search query, a keyword (sometimes more than one keyword) gets triggered and a list of paid ads will be shown in the search results. A "snapshot" which captured all the listed paid ads provided by the search engine is collected in the dataset. One thing we need to highlight here is that each "snapshot" captured in the dataset only includes information related to paid ads and there is no organic search results involved.

We define "impression" as the display of a paid ad in the search results. Thus, in the dataset, each row of records represents a paid ad impression with data collected in four aspects: 1) Advertiser information, including advertiser ID, name and a link to its website; 2) Displayed ad information, including ranking position ranging from one to ten, click status and ad quality, where ad quality is a score given by the search engine (although the search engine does not disclose the exact algorithm of calculating ad quality, it asserts that it is a function of the ad landing page quality, the previous click-through rates and how well the product or service matches the description written in the paid ads); 3) Keyword information associated with the displayed paid ad, including keyword content, match type, the amount of money that the advertiser bids on the keyword, and the amount of money (cost-per-click, CPC) that the advertiser spent when the ad gets clicked; and 4) Consumer search information, in which we capture the search query information, including the exact search terms that caused the appearance of the paid ads, ip address and sometime cookie is captured as well. Figure 3.3 shows an example of the dataset with some of the key attributes described above.

search id	ip	query	advertiser	Section	rank	click	Keyword match	Ads quality	price ($)	bid ($)
1	258.163	Shanghai Hotel	booking .com	L	1	1	AC	3.5	1.8	3
1	258.163	Shanghai Hotel	agoda .com	L	2	0	PH	2.3	0	1.8
1	258.163	Shanghai Hotel	ctrip.com	L	3	1	AC	2.5	0.78	1.3
1	258.163	Shanghai Hotel	crownplaza.com	L	4	0	PH	2.0	0	0.78
1	258.163	Shanghai Hotel	tripadvisor .com	L	5	0	BR	1.1	0	0.50
1	258.163	Shanghai Hotel	Holiday Inn	R	1	0	AC	2.0	0	0.45
1	258.163	Shanghai Hotel	Intercontinental	R	2	0	PH	1.5	0	0.40

Figure 3.3 An Example of the Dataset

As illustrated in Figure 3.3, a consumer typed a "Shanghai Hotel" search query. The search engine provided seven paid ads in the search results. Of all the sponsored ads, five of them are displayed on the left-side of the screen. And the rest of them are shown on the right-side of the screen. Keyword match, bid and ad quality reflect each displayed ad's features.

In this example, we could also see that this consumer click on the first and third paid ads on the left-side, whereas there is no click generated from the right-side of the screen. For those ads being clicked by the consumer, the attribute "price" shows the actual money that the advertiser pays to the search engine, and it follows the "generalized second price" mechanism.

CHAPTER 4 ESSAY ONE: EXAMINING KEYWORD PERFORMANCE IN SPONSORED SEARCH ADVERTISING

In this chapter, we conduct an empirical study to analyze advertiser keyword branding issue involved in the sponsored search advertising field. Five sections are included in this chapter. In Section 4.1, we briefly introduce the background and discuss the motivation of conducting this study from both academic and managerial perspectives. In Section 4.2, we discuss the methodology, including data analysis and modeling approach for solving the research questions. In Section 4.3, we present the empirical findings and the results of several robustness tests. Section 4.4 explains the managerial implications of the findings. In Section 4.5, the conclusion is made.

4.1 Introduction

As we described in Chapter One, there are two major issues that an advertiser always needs to consider in order to effectively manage its paid search campaigns. One is to effectively plan for the bidding schemes for participations in a so-called “generalized second-price” auction process. As summarized in the previous chapter, there are a number of analytical studies developed to examine the auction mechanism behind the sponsored search advertising (Edelman, Ostrovsky and Schwarz 2007; Varian 2007; Athey and Ellison 2011).

Another important issue is to select appropriate keyword sets. Doing so could help advertisers accomplish several goals. First of all, it could assist advertisers to reach out to advertisers’ potential customer segments and attract consumers to visit advertisers’ websites (Rutz and Trusov 2011). Second, it could help advertisers deter competitors which are co-existed in the sponsored search environment (Desai, Staelin and Shin 2014).

Finally, it could facilitate the occurrence of transactions within advertisers' websites (Ghose and Yang 2009).

Generally, advertisers create keyword sets by extracting potential words or short phrases from a variety of social media websites, such as product reviews or personal blogs. Under the rapid development of text-mining techniques, the number of potential keywords that can be mined by advertisers is increasing dramatically. In reality, a single advertiser could manage as much as hundreds of keywords on a daily basis. However, due to the budget constraints, researchers have found that many advertisers invest most of budgets on a much smaller portion of keywords (Rutz and Trusov 2011). Therefore, identifying the “right” keywords becomes a challenging task for advertisers to achieve.

Literature related to keyword selection topic has shown some descriptive insights. For example, researchers have concentrated on examining the characteristics of keywords and its impact on consumer click-through rate and conversion rate (Ghose and Yang 2009; Rutz and Bucklin 2011). However, since the findings derived from the studies are examined using data collected from a single advertiser, it might be too arbitrary to generalize findings into broader fields. In reality, it is quite possible that consumers might respond to paid ads differently based upon advertiser's variety (e.g. hotel website versus online travel aggregators), which may eventually lead to different keyword strategies to different types of advertisers. To that end, potential research with regard to the development of advertiser keyword strategy shall be studied by taking into account the intrinsic differences among advertisers. Also, analysis is expected to be conducted while incorporating the complicated situations involved in the search engine advertising environment (e.g. customer search heterogeneity). On the basis of that, research developed to examine keyword branding – an important topic in sponsored search advertising – has not yet been thoroughly explored.

In this study, we investigate keyword branding issue in the paid search advertising environment. We focus on examining the impact of various types of branded keywords on consumer click-through rates, as our dataset collected from the search engine doesn't keep track of the consumer conversions which happened inside of the advertisers' websites. The goal of the study is to address the following research questions: How would various types of branded keywords influence consumer click-through rates in the sponsored search advertising environment? Is this influence consistent across different types of advertisers? If not, how could different advertisers select branded keywords to improve paid search campaigns? We employ a two-stage joint probability approach to address these research questions. We start with an examination to understand the general paid ads performance in the industry. We then conduct a Swait and Louviere test to examine whether consumers differentiate paid ads from one type of advertisers to another, and thereby respond to ads differently – especially in terms of click-through rates. Finally, we study the impact of various branded keywords on consumer click-through rates for each type of advertisers described above.

Our empirical results show several insights. First, the empirical results show that paid ads from various types of advertisers experience significantly different performance in the sponsored search advertising domain. More specifically, online travel aggregators' ads receive more prominent ranking positions and higher click-through rates than hotels' paid ads. Second, we find that consumers differentiate paid ads between online aggregators and hotels, and they tend to click on aggregators' ads rather than hotels' ads when controlling for other influential factors. Third, with respect to the branded keyword selections, our findings suggest that both aggregators and hotels should consider selecting own brands into the keyword portfolios. When considering which other advertisers' brands can be included in the keyword settings, our empirical results suggest that online aggregators should choose "across-category" branded keywords rather than "within-category" branded keywords. In other words, online aggregators would be better off to choose hotel branded keywords rather than other aggregator branded keywords. With regard to hotel advertisers, they would benefit from choosing "within-category" rather than "across-category" branded keywords. Meanwhile, choosing "across-category" branded keywords will not create significant increase in consumer click-through rates compare with general generic keyword settings.

The study contributes to the body of literature in two aspects: First, on the basis of the previous literature, we not only discuss the impact of different keyword characters on consumer click-through rate (e.g. branded vs. generic keyword), but also expand the branded keywords into more detailed examinations (e.g. advertiser's own brand, "across-category" and "within-category" branded keywords), in which we study each specific group of brands impact on consumer click-through rates and its corresponding keyword branding strategy for different types of advertisers. Second, we add to the literature by examining paid ads through different types of advertisers. Particularly in our context, we study two major advertisers – online travel aggregators and hotels. The examinations across different types of advertisers within the paid search advertising domain give us a chance to understand how differently the consumers and the search engine might respond to their paid ads – in terms of click-through rates and ads placements. From practical standpoint, the empirical findings provide valuable information to facilitate advertiser's keyword selection process. It could also shed lights on new adopters (advertisers with similar characters as we examined or in the same industry) regarding their prospective performance in sponsored search. The limitations of this study are mostly drawn from the empirical data. Since the dataset is a collection of information from the search engine's website, we are unable to keep track of consumer purchases occurred in the advertisers' websites, which precludes us from using consumer conversion rate as a measurement for evaluating the effectiveness of advertisers' branded keywords. Under this circumstance, the branded keyword selection suggestions from our study are developed for the purpose of increasing consumer click-through rates or attracting more customer visits to the advertisers' websites, which may not directly contribute to the advertiser profitability with respect to the sales of products or services.

4.2 Methodology

In this section, we first provide descriptive data analysis to illustrate the characteristics of the dataset we used for conducting this study. Then, we build the model framework using joint probability approach. Two sets of models are presented to examine the research questions.

4.2.1 Data Analysis

To address the research questions raised at the beginning of this chapter, we use a dataset which includes one-day advertiser keyword impressions and consumer clickstream information collected from the search engine provider. The dataset contains 1,440,660 impressions across 62,253 distinct keywords from 1,150 advertisers. Of all the impressions, 183,654 clicks were generated, which leads to 12.75% aggregate-level click-through rate. All paid ads collected in the dataset are displayed on the left side of the screen. Table 4.1 illustrates an aggregate-level advertiser click-through performance. It summarizes click-through rate (CTR), cost-per-click (CPC), bids and ad quality for all advertisers in our dataset. As shown from Figure , the search engine places higher quality ads near the top, with advertisers paying for the prominent position with elevated bids. The situation in which CPCs are less than bids reflects the nature of second price auction mechanism. The CTRs diminish as rank goes from 1 to 10.

Table 4.1 Summary of Rank Specific Performance

Rank	**CTR**	**CPC**	**Bid**	**Ad Quality**
1	22.9%	3.33	3.9	1.68
2	15.8%	2.18	2.51	1.65
3	10.7%	1.68	1.83	1.57
4	7.7%	1.43	1.55	1.48
5	5.9%	1.31	1.41	1.42
6	4.2%	1.24	1.32	1.35
7	3.2%	1.11	1.23	1.30
8	2.7%	1.07	1.14	1.25
9	2.2%	0.98	1.06	1.21
10	2.4%	0.88	0.97	1.17
ALL	12.7%	2.47	2.32	1.55

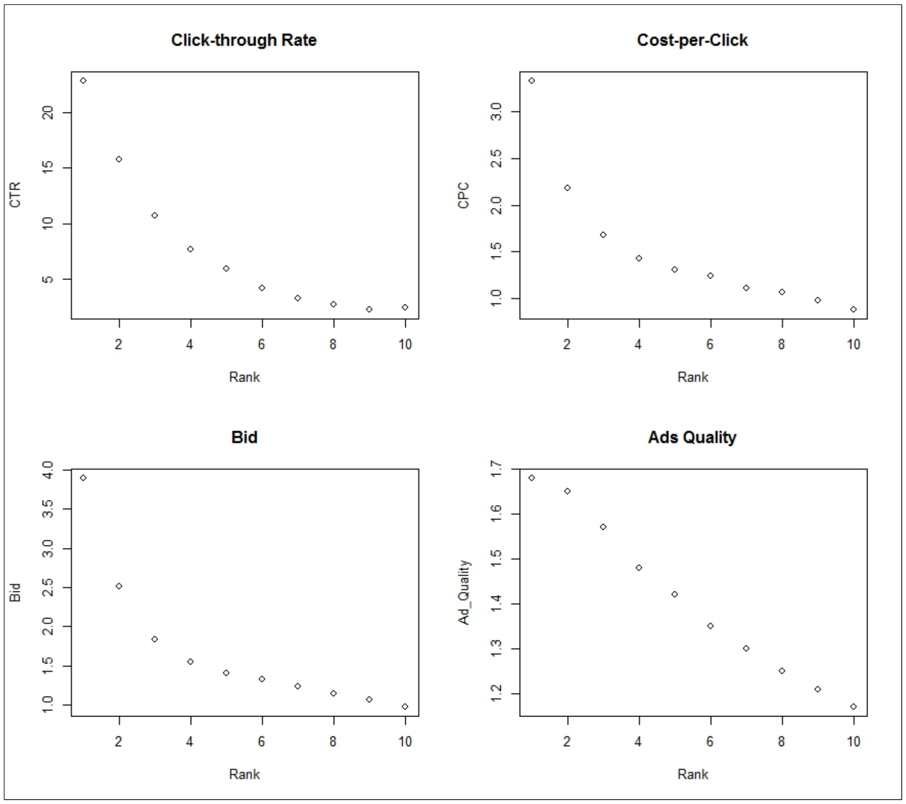

Figure 4.1 Scatter Plots of Rank Specific Performance

Among all the advertisers, 434 are hotel websites and 716 are online travel aggregators (OTA). As summarized in Table 4.2, OTAs experience slightly higher rankings than hotels, where OTA's average ranking is 3 while hotels' average ranking is 4. With regard to the click-through rate, OTAs receive much higher click-through than Hotels (13% and 8% respectively).

Table 4.2 Summary of Two Types of Advertisers

Advertiser Type	**Variables**	**Mean**	**Std.dev.**	**Min**	**Max**
OTA (716)	Rank	3.13	1.95	1	10
	Ads Quality	1.56	0.86	0	9.51
	Bid ($)	0.40	0.44	0.05	5.88
	Click	0.13	0.34	0	1
Hotel (434)	Rank	3.89	2.24	1	10
	Ads Quality	1.45	0.78	0	9.11
	Bid ($)	0.31	0.31	0.07	6.67
	Click	0.08	0.28	0	1

Table 4.3 illustrates the click-through performance from two advertiser categories. As described in Figure 4.2, the CTRs, bid and CPC generated from online travel aggregator at prominent ranks are much higher than the CTRs generated from hotel websites.

Table 4.3 Summary of Rank Specific Performance through Two Types of Advertisers

Rank	*OTA*				*Hotel*			
	CTR	**CPC**	**Bid**	**Ad Quality**	**CTR**	**CPC**	**Bid**	**Ad Quality**
1	23.4%	3.38	4.05	1.68	18.1%	2.65	2.82	1.70
2	16.3%	2.19	2.55	1.66	10.7%	2.08	2.14	1.56
3	11.1%	1.67	1.82	1.58	7.8%	1.77	1.88	1.48
4	8.0%	1.42	1.55	1.49	6.0%	1.56	1.61	1.42
5	6.0%	1.31	1.39	1.43	5.0%	1.40	1.48	1.36
6	4.4%	1.24	1.30	1.37	3.6%	1.26	1.42	1.30
7	3.2%	1.09	1.21	1.31	2.9%	1.19	1.28	1.27
8	2.7%	1.07	1.12	1.27	2.5%	1.09	1.21	1.20
9	2.1%	0.95	1.05	1.23	2.5%	1.00	1.11	1.17
10	2.2%	0.94	0.97	1.18	2.9%	0.76	0.96	1.16

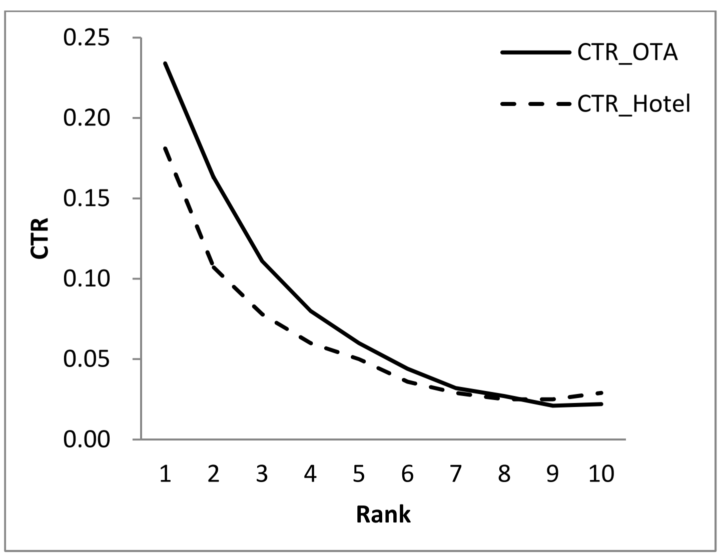

Figure 4.2 Click-Through Rates Comparison (OTA vs. Hotel)

We expand the existing dataset by introducing attributes from two perspectives. First, we constructed dummy variables to capture individual keyword characteristics. Particularly, we focused on categorizing brand information that is embedded in each keyword. Four variables are added under this situation: *Brandself Keyword*, *Brandother Keyword*, *Brandother_OTA Keyword*, and *Brandother_Hotel Keyword*. When an advertiser includes its own brand name into the keywords (e.g. Hilton.com selects "Hilton hotel in Newark" as one of its keywords), *Brandself Keyword* is marked as 1. If an advertiser includes other brand names to its keyword portfolio, *Brandother Keyword* is marked as 1. We further partition advertiser choosing other brand names into two scenarios. One scenario is that when an advertiser chooses other online travel aggregator brand name, we assign *Brandother_OTA Keyword* as 1. It can be exemplified as Hilton.com or Booking.com selects "expedia.com" as one of its keywords. The other scenario is that when an advertiser selects other hotel brand name, we label *Brandother_Hotel Keyword* as 1. For instance, Hilton.com or Booking.com selects "Marriott.com" as one of its keywords. When *Brandself Keyword* and *Brandother Keyword* are both equal to 0, it means that the character of the keyword is generic.

Second, we add a dummy variable to classify advertiser into two groups based upon the two advertise types (Hotels vs. OTAs). When the displayed paid ad belongs to an online travel aggregator, the dummy variable is marked as 1 and 0 otherwise. Third, we classify consumer search queries into two categories (branded vs. generic). If the search query includes brand information (e.g. a consumer search for "Hilton Hotel in Newark"), *Brand Query* is labeled as 1. Otherwise, it is recorded as 0. *Brand Match* captures the "match" between consumer branded queries and advertiser brands. Taking the previous example, if the search query contains “Hilton Hotel” and one displayed paid ads in response to this query is owned by the Hilton hotel group, *Brand Match* is recorded as 1. If, however, none of the displayed ads belong to Hilton Hotel, instead, the displayed ads returned from “Hilton Hotel” queries are owned by Marriott.com or Travelocity.com, *Brand Match* is recorded as 0. Table 4.4 reports the click-through rates across different advertisers under different keyword characters, keyword match types and consumer queries.

Table 4.4 Summary of CTRs by Keyword Match Types and Query Types

Category	***ALL*** **(%)**	***OTA*** **(%)**	***Hotel*** **(%)**
Keyword Match Type:			
Accurate	13.5	13.7	10.4
Phrase	12.4	13.4	8.0
Broad	9.4	10.2	7.3
Branded Keywords:			
Brandself Keyword	21.3	41.8	20.0
Brandother Keyword	13.1	13.3	6.8
Query Type:			
Branded Query	14.1	13.5	18.1
Branded Match	26.2	44.5	24.4
Generic Query	12.4	13.3	6.2

As illustrated in Table 4.4, Accurate Keyword, compare with two other keyword match type categories, generates the highest click-through rate for both aggregators and hotel websites. Similarly, Brandself Keyword receives higher click-through rates than Brandother Keyword. As for various types of consumer query, Branded Query experiences higher click-through rates than Generic Query. And, the click-through rates of Brand match for either aggregators or hotels are at least twice as much as the generic query.

4.2.2 Model Framework

Paid search advertisers consider click-through rate one of the critical measurements when evaluating the effectiveness of sponsored advertising campaigns. Examining click-through rate in an effective manner could further assist advertiser decision making, especially when it needs to decide which keywords can attract more consumer traffic, what specific content should be included (e.g. branded or generic, include its own brand name or other advertiser's brand name).

In this section, we introduce the model framework for estimating consumer click-through rate. We employ a joint probability modeling approach to capture the behavior generated from three entities: advertisers, consumers and the search engine. The section is begun with a binary logit model, which is to model consumer click behavior given individual ad display. We then introduce an ordered logit model, which is used as a proxy to model the ranking decision made by the search engine. We expand our model into the joint probability format by multiplying these two independent models. Finally, we build two sets of model to address the research questions raised at the beginning of the study.

4.2.2.1 Click Model

When evaluating consumer click-through rate, advertisers oftentimes would rely on its own paid ads historical data, given the facts that they generally don't have other advertisers' detailed paid advertising information (e.g. keyword portfolio, bids, paid ads quality) and that the search engine dynamically manage all paid ads. Meanwhile, advertisers couldn't collect detailed consumer information from the search engine, such as consumer user id and cookie information, due to the search engine's privacy nondisclosure policy. Under this circumstance, advertisers are unable to identify those consumers, who search for and click on their paid ads, and any sequential clicks generated from the search engine site. Even assuming that advertisers could somehow capture the consumer's IP address information, it would still be too arbitrary to conclude that clicks collected from the identical IP address are done by the same person. Two independence assumptions are made based upon the situations described above. First, for each displayed paid ad, we assume that consumer's click decision is independent. Second, we assume that consumer's each time click decision is independent, if there is any sequential click behavior involved. In other words, an individual's click decision is not influenced by his or her prior click decisions.

We use a binary logit model to estimate consumer click-through rate, because it is consistent with the literature (e.g. Ghose and Yang 2009; Rutz and Trusov 2011). Another important reason we choose binary logit model, rather than multinomial logit model (MNL), is due to the advertiser's lack of detailed paid ads information of other co-listed advertisers, especially under the circumstance where the search engine actively arranges displayed paid ads. We denote *i* as consumer query, and *j* as advertiser paid ads. We assume that when a consumer's perceived utility, denoted as U_{ij}^{C}, is greater than 0, the individual will click on paid ad *j* under query *i*. We use random utility function to represent an individual's perceived utility, which is specified as $U_{ij}^{C} = V_{ij}^{C} + \varepsilon_{ij}^{C}$, where V_{ij}^{C} is the deterministic portion of the consumer perceived utility and it can be specified as a linear function of X_{ij}^{C}. X_{ij}^{C} is a set of covariates which influence an individual's click decision. ε_{ij}^{C} is defined as the stochastic error term which is assumed to be identically and independently distributed with the extreme value distribution. Thus, the click model can be specified as:

$$P(Y_{ij} = 1) = \frac{\exp(V_{ij}^{C})}{1 + \exp(V_{ij}^{C})} \tag{1}$$

In equation (1), Y_{ij} is a dummy variable which represents consumer click status on paid ad *j* under query *i*. When a consumer clicks on ad *j*, Y_{ij} is assigned as 1, and 0 otherwise. $P(Y_{ij} = 1)$ denotes the probability of consumer clicking on paid ad *j* under query *i* with consumer's perceived utility or propensity to click, V_{ij}^{C}. As described before, V_{ij}^{C} can be specified as a linear function of X_{ij}^{C}, which consists a set of explanatory variables that could represent the displayed paid ad *j*'s characters (ad quality, bid, ranking position, keyword characters) and a set of control variables (query character, advertiser type). Therefore, the representative utility V_{ij}^{C} can be written in the format of equation (2):

$$\begin{aligned} V_{ij}^{C} &= \beta_0 + \beta_1 AdQuality_{ij} + \beta_2 Bid_{ij} + \beta_3 Rank_{ij} + \beta_4 Brandself\_Keyword_{ij} \\ &+ \beta_5 Brandother\_Keyword_{ij} + \beta_6 Accurate\_Keyword_{ij} + \beta_7 Phrase\_Keyword_{ij} \\ &+ \beta_8 Brand\_Query_{ij} + \beta_9 Brand\_Match_{ij} + \beta_{10} OTA_{ij} \end{aligned} \tag{2}$$

One of the issues in modeling click behavior is the dependence of click-through rates upon search engine controlled ad position as shown in Table 4.1, where click-through rate dramatically decreases with lower ad position. Table 4.5 summarizes two binary logit models of estimating consumer click-through rate, one with variable Rank and the other without *Rank*.

Table 4.5 Parameter Estimations of Click Model with & without Rank

Parameter	***With Rank***			***Without Rank***		
	estimates	**S.E.**	**p-value**	**estimates**	**S.E.**	**p-value**
Intercept	-2.4628	0.0135	<0.0001	-3.6052	0.013	<0.0001
Brandmatch	0.4485	0.0210	<0.0001	0.7353	0.0207	<0.0001
Brandquery	0.0802	0.0076	<0.0001	0.178	0.00746	<0.0001
OTA	0.356	0.0111	<0.0001	0.571	0.0109	<0.0001
Accurate Keyword	0.4441	0.0113	<0.0001	0.2832	0.0116	<0.0001
Phrase Keyword	0.3839	0.0115	<0.0001	0.324	0.0119	<0.0001
Ad Quality	0.6042	0.0027	<0.0001	0.6951	0.0027	<0.0001
Bid	0.0184	0.00101	<0.0001	0.0853	0.0009	<0.0001
Rank	-0.3446	0.0019	<0.0001			

As illustrated in Table 4.1, consumer click-through rate decreases as rank goes from topmost to bottom positions indicates *Rank* to be a strong determinant for estimating consumer click-through rate. However, assigning paid ads into different slots is a decision made only by the search engine. Thus, it makes little sense to include *Rank* in the click model, because rank is actually outside the control of advertisers. But if we exclude variable *Rank* from the click model, as shown in Table 4.5, the parameter estimates are biased. For instance, with the removal of *Rank*, *Accurate_Keyword* results in lower click propensity (and lower probability of click) than *Phrase_Keyword*, where appropriate relative values of parameter estimates result with *Rank* in model. Similarly, the variable *Bid* is also affected by the exclusion of *Rank*, this could be illustrated by looking at the odds ratio impact of *Bid*, where the odds ratio impact can be calculated as the exponential of the parameter estimates. With exclusion of rank, the odds ratio increase to 1.0890 from 1.0186 due to the correlation between *Bid* and *Rank* (-0.35, $p <$

0.0001) and the direct result of the search engine utilizing *Bid* as a key driver in paid ad display position. It is this need to include paid ad rank in our modeling of probability of click combined with the control of rank by the search engine that leads us to jointly model consumer click and rank, where the probability of paid ad rank becomes the output of an ordered logit model and the input of the binary logit model – in essence including the importance of *Rank* by realizing its exogenous value as controlled by search engine but influenced consumer's click-through behavior. Taking variable *Rank*'s endogenous property into account, the function V_{ij}^{C} can be rewritten as:

$$\begin{aligned} V_{ij}^{C} = {} & \beta_0 + \beta_1 Accurate\_Keyword_{ij} + \beta_2 Phrase\_Keyword_{ij} + \beta_3 Brandself\_Keyword_{ij} \\ & + \beta_4 Brandother\_Keyword_{ij} + \beta_5 Rank_{ij} + \beta_6 Brand\_Query_{ij} + \beta_7 Brand\_Match + \beta_8 OTA_{ij} \end{aligned} \tag{3}$$

4.2.2.2 Rank Model

We employ the ordered logit modeling approach to estimate the search engine's ranking decision. As we described in the previous section, advertisers depend upon limited information when estimating click-through rate. Similarly, when it comes to make conjectures regarding the search engine's ranking decision, advertisers do not have full information with respect to the comprehensive algorithm used by the search engine for ranking calculations and other advertisers' detailed paid ads information. Therefore, advertisers still need to rely on their own paid ads information. The fact that individual advertiser couldn't gather other co-listed advertiser's paid ads information also eliminates the possibility of using ordered rank logit model (Beggs, Cardell and Hausman 1981) in estimating the search engine's ranking decision.

We denote R_{ij} as the actual ranking that paid ad *j* received under consumer query *i*, and we assume that the actual rankings are known by advertisers. R_{ij}^{*} is denoted as a latent variable, which is unobserved by advertisers and is an index that reflects the real ranking situations. The latent variable R_{ij}^{*} can be specified in the random utility function, $R_{ij}^{*} = V_{ij}^{R} + \varepsilon_{ij}^{R}$, where ε_{ij}^{R} are the stochastic error terms assumed to be identically and independently distributed with the extreme value distribution. V_{ij}^{R} can be expressed as a linear function of X_{ij}^{R}. X_{ij}^{R} is a set of variables that affect the search engine's ranking decision.

Consider the situation where a paid ad is chosen to be displayed in the search results webpage. In this position allocation process, instead of randomly assigning the selected paid ad to a slot, the search engine would most likely prioritize the ad to a more prominent ranking position if the ad itself has a higher quality or if the keyword embedded within the paid ad better matches consumer search queries. Doing so would increase the chance of the paid ad being clicked by the inquired consumers, which in return will raise the search engine's revenue based upon the "pay-per-click" mechanism. Thus, we consider X_{ij}^{R} includes variables from three perspectives: displayed ads quality, bids, and match between consumer query and keyword. We also control for advertiser type by adding a dummy variable OTA to the model. Therefore, the representative utility V_{ij}^{R} can be elaborated as:

$$\begin{aligned} V_{ij}^{R} = {} & \alpha_1 AdQuality_{ij} + \alpha_2 Bid_{ij} + \alpha_3 AccurateKeyword_{ij} + \alpha_4 PhraseKeyword_{ij} \\ & + \alpha_5 OTA_{ij} \end{aligned} \tag{4}$$

Given that the ranking positions in our dataset range from 1 to 10, we denote k as any positive integer between 1 and 10, $k \in [1,10]$, representing the actual ranking that paid ad j received given query i. Let M represents a set of cut-offs with nine elements included, $\mathrm{M} = \{\mu_1, \mu_2, ..., \mu_9\}$. The probability of paid ad *j* being located at rank *k* can be specified as:

$$P(R_{ij} = k) = P(\mu_k < R_{ij}^* \le \mu_{k-1}) = P(\mu_k - V_{ij}^R < \varepsilon_{ij}^R \le \mu_{k-1} - V_{ij}^R) = \Lambda(\mu_{k-1} - V_{ij}^R) - \Lambda(\mu_k - V_{ij}^R) \quad (5)$$

where: $\Lambda(\mu_k - V_{ij}^R) = \dfrac{\exp(\mu_k - V_{ij}^R)}{(1 + \exp(\mu_k - V_{ij}^R))}$

4.2.2.3 Joint Probability Model

The joint probability model is built based upon the click and rank models presented in the previous sections. Since the decision made by the search engine on ranking allocation is independent from the current click decision made by consumers, the joint probability model can be expressed as the multiplication of the click model and the rank model. Let *N* represents the total number of consumer search queries, the joint probability model can be written as:

$$P(Y_{ij} = 1) = \sum_{i=1}^{N} P(Y_{ij} = 1 \mid R_{ij} = k) \cdot P(R_{ij} = k) \quad (6)$$

We use traditional maximum likelihood estimation (MLE) to estimate the joint probability model. The likelihood[5] as a result of multiplying each internet users click decision is:

$$LH = \prod_j \prod_i P(Y_{ij} = 1 \mid R_{ij} = k)^{Y_{ij}} P(Y_{ij} = 0 \mid R_{ij} = k)^{1-Y_{ij}} \cdot P(R_{ij} = k) \quad (7)$$

Resulting in log-likelihood:

$$LLH = \sum_{i=1}^{N} [Y_{ij} \cdot \log P(Y_{ij} = 1 \mid R_{ij} = k) + (1 - Y_{ij}) \cdot \log P(Y_{ij} = 0 \mid R_{ij} = k) + \log P(R_{ij} = k)] \quad (8)$$

---

[5] The detailed steps of the derivation of the ordered logit model (rank model) and the log likelihood function are included in Appendix A.

We then maximize this log-likelihood function (equation 8) to determine parameter estimates. Due to the complexity of the log-likelihood function we numerically maximize using the quasi-Newton Broyden Fletcher Goldfarb Shanno (BFGS). We can approximate standard errors for the parameter estimates by inverting the approximated Hessian in the BFGS routine, with the diagonals of the inverted Hessian being the standard errors of the estimates.

4.2.2.4 Model Setup

Using the joint probability model framework, two sets of model are developed to examine the research questions. Model 1(a) is to provide general understandings regarding how consumer click-through rates might change under different paid ads' keyword configurations and ranking positions. In Model 1(b) and Model 1(c), we separate the dataset into two categories (OTAs and Hotels) and run the click model to test whether consumers differentiate displayed paid ads from one kind to another. Model 2(a) and Model 2(b) are built to explore the contribution of keyword configurations to consumer click-through rates in the sponsored search advertising environment. The detailed empirical results are discussed in Section 4.3.

4.3 Empirical Analysis

In this section, we discuss the empirical results generated from the two sets of model presented in the methodology section. Robustness tests are conducted to verify the consistency of the estimations from the joint probability model framework.

4.3.1 Parameter Estimation

Table 4.6 and 4.7 summarize the parameter estimates generated from the two sets of models. Estimates from the click model are included in Table 4.6 and estimates from the rank model are presented in Table 4.7.

Table 4.6 Parameter Estimation from Joint Probability Models -- Click Model

Parameter	**Model 1(a)**	**Model 1(b)**	**Model 1(c)**	**Model 2(a)**	**Model 2(b)**
	ota & hotel combined	*ota only*	*hotel only*	*ota only*	*hotel only*
Click Model:					
Intercept	-1.893*** (0.014)	-1.460*** (0.012)	-1.879*** (0.025)	-1.460*** (0.012)	-1.882*** (0.025)
Brand Query	-0.106*** (0.014)	-0.116*** (0.015)	-0.035 (0.052)	-0.116*** (0.015)	-0.036* (0.052)
Brand Match	0.793*** (0.039)	1.057*** (0.065)	0.928*** (0.066)	1.057*** (0.065)	0.930*** (0.066)
Accurate_Keyword	0.716*** (0.011)	0.794*** (0.012)	0.384*** (0.027)	0.794*** (0.012)	0.387*** (0.027)
Phrase_Keyword	0.535*** (0.011)	0.626*** (0.013)	0.064* (0.027)	0.625*** (0.013)	0.067* (0.027)
Brandself_Keyword	0.402*** (0.037)	0.382*** (0.076)	0.273*** (0.043)	0.382*** (0.076)	0.274*** (0.043)
Brandother_Keyword	0.036** (0.016)	0.042** (0.016)	0.153* (0.077)	--	--
Brandother_Hotel Keyword	--	--	--	0.044** (0.016)	0.159* (0.077)
Brandother_OTA Keyword	--	--	--	-0.219* (0.110)	-0.135 (1.045)
Rank	-0.397*** (0.002)	-0.410*** (0.002)	-0.285*** (0.005)	-0.410*** (0.002)	-0.285*** (0.005)
OTA	0.480*** (0.011)	--	--	--	--

*** $p<0.001$; ** $0.001<p<0.01$; * $0.01<p<0.05$

Table 4.7 Parameter Estimation from Joint Probability Models -- Rank Model

Parameter	**Model 1(a)**	**Model 1(b)**	**Model 1(c)**	**Model 2(a)**	**Model 2(b)**
	ota & hotel combined	*ota only*	*hotel only*	*ota only*	*hotel only*
Rank Model:					
Ad Quality	0.516*** (0.002)	--	--	0.502*** (0.002)	0.637*** (0.006)
Bid	0.671*** (0.001)	--	--	0.676*** (0.001)	0.634*** (0.004)
Accurate_Keyword	0.239*** (0.006)	--	--	0.280*** (0.007)	0.158*** (0.013)
Phrase_Keyword	0.293*** (0.006)	--	--	0.346*** (0.007)	0.140*** (0.012)
OTA	0.789*** (0.005)	--	--	--	--
Cutoffs:					
U1	5.189*** (0.008)	--	--	4.445*** (0.009)	4.989*** (0.020)
U2	4.018*** (0.008)	--	--	3.263*** (0.008)	3.956*** (0.018)
U3	3.202*** (0.008)	--	--	2.443*** (0.008)	3.185*** (0.016)
U4	2.520*** (0.007)	--	--	1.759*** (0.008)	2.521*** (0.016)
U5	1.890*** (0.007)	--	--	1.127*** (0.008)	1.906*** (0.015)
U6	1.245*** (0.007)	--	--	0.474*** (0.008)	1.295*** (0.015)
U7	0.503*** (0.007)	--	--	-0.276*** (0.008)	0.574*** (0.015)
U8	-0.648*** (0.008)	--	--	-1.449*** (0.009)	-0.515*** (0.016)
U9	-14.966** (5.609)	--	--	-15.694* (6.395)	-19.838 (131.07)

**** p<0.001; ** 0.001<p<0.01; * 0.01<p<0.05*

As illustrated in Table 4.6 and 4.7, in Model 1(a), the coefficients of *OTA* generated from click model and rank model (0.480 and 0.789 respectively, $p<0.001$) are positive and significant, indicating that, on average, paid ads belong to online travel aggregators receive higher click-through rate, as well as more prominent rankings. The estimate of *Brand Query* is negative (-0.106, $p<0.001$), whereas *Brand Match* is positive (0.793, $p<0.001$), showing that when the search query is branded, the average click-through rate on a paid ads is lower. However, when the brand of a displayed paid ad matches the consumer's branded query, the click-through rate is significantly increased. This could be demonstrated by *Brand Match* compensating the negative effect generated by *Brand Query* (0.793-0.106 = 0.687) and positively contributing to click-through rates.

Second, the coefficients of *Accurate_Keyword* and *Phrase_Keyword* are both positively associated with click-through rates (0.716 and 0.535 respectively). To examine the significance of the difference between *Accurate_Keyword* and *Phrase_Keyword*, we conducted a Wald test (See details in Appendix B). The result ($32.79 > Z_{0.001} = 3.09$) shows that the estimate of *Accurate_Keyword* is significantly greater than the estimate of *Phrase_Keyword*, indicating that consumers would mostly favor those displayed paid ads with accurate keyword embedded, which exactly match their search queries. The least option that individuals might choose would be ads with broad keyword match equipped. With respect to the impact of branded keywords on consumer click-through rate, the coefficient of *Brandself_Keyword* is positively associated with click-through rate (0.402, $p<0.001$), showing that paid ads with advertiser's own brand included receives higher click-through rate than ads with generic keyword embedded. Although *Brandother_Keyword* is positive and significant (0.036, $p<0.001$), the value is close to 0, meaning that having ads equipped with other advertiser brand information would bring limited increase to consumer click-through rates. This can also be proved by a Wald test ($20.42 > Z_{0.001} = 3.09$), in which the estimate of *Brandself_Keyword* is significantly greater than the estimate of *Brandother_Keyword*. As expected, the coefficient *Rank* is negative and statistically significant (-0.397, $p<0.001$). *Ad Quality* and *Bid* are both positively associated with the ads ranking (0.516 and 0.671 respectively), which are consistent with the previous literature.

Next, to further understand whether or not consumers differentiate displayed paid ads from online travel aggregators to hotel websites, we divide the dataset into two subgroups based upon advertiser types and conduct a test developed by Swait and Louviere (1993) using parameter estimates generated from the click model (See results from Model 1(b) & (c) and details in Appendix C). The result from the test yields λ_A = 4136.96 > $\chi^2(9)$ = 23.59, indicating that consumers do distinguish ads from different advertisers when making click decisions. Adding to the previous discussion of the coefficient of *OTA* (0.480, p<0.001) in Model 1(a), we could infer that consumers prefer to click on aggregators' paid ads than hotels' ads, when controlling for other factors.

Given consumers' different reactions toward displayed ads, Model 2(a) and 2(b) (representing OTAs and hotels respectively) are developed to separately investigate the relationships between various keyword configurations (e.g. brand content, match type) and consumer click-through rates. To examine the impact of various types of branded keywords on click-through rates in a detailed manner, we use variables *Brandother_Hotel Keyword* and *Brandother_OTA Keyword* as substitutes for *Brandother_Keyword*. As shown in Table 4.6, the coefficients of *Brandself_Keyword* generated from Model 2(a) and Model 2(b) are positive and significant (0.382 and 0.274 respectively), implying that both aggregators and hotels should consider including their own brand names as keyword options.

Second, the coefficient of *Brandother_Hotel Keyword* in Model 2(a) is positively associated with click-through rates (0.044, p = 0.007), whereas the coefficient of *Brandother_OTA Keyword* is negatively related to click through rates (-0.219, p = 0.047), indicating that when aggregators include hotel brand names in their keywords, the average click-through rate is significantly greater than having generic keywords equipped. On the other hand, when including other aggregator's brand names, aggregators will receive lower click-through rates than having generic keywords embedded. These results suggest that when considering the option of adding other brands into the keyword portfolios, aggregators should favor hotel brand names rather than other aggregator brands for the purpose of receiving higher click-through rates.

Third, the coefficient of *Brandother_Hotel Keyword* in Model 2(b) is positively related to click-through rates (0.159, p = 0.04), while the coefficient of *Brandother_OTA Keyword* is negative but not significant (-0.135, p = 0.89). The results imply that when designing to include other brand information into keyword portfolios, hotel advertisers should choose other hotel brand names rather than aggregator brands in order to obtain higher click-through rates. To further capture the different influence made by these three types of keyword configurations (with own brand, other hotel brands, other aggregator brands), we conducted a series of Wald tests for both aggregators and hotel advertisers (See Table 4.8).

Table 4.8 Results from Wald Tests on Model 2(a) & Model 2(b)

Model	**Null Hypothesis**	**Test Statistics**		**Result at Level 0.1% $Z_{0.05}$ = 1.65**
		Wald b_1-$b_2\leq0$	Wald* b_1/b_2-1≤0	
Model 2(a)	*Brandself ≤ Brandother_Hotel*	5.98	3.93	Rejected
	Brandother_Hotel ≤Brandother_OTA	2.39	11.36	Rejected
	Accurate ≤ Phrase	29.41	24.86	Rejected
Model 2(b)	*Brandself ≤ Brandother_Hotel*	1.74	1.99	Rejected
	Brandother_Hotel ≤Brandother_OTA	0.28	0.24	Cannot Rejected
	Accurate ≤ Phrase	16.20	3.88	Rejected

From aggregator's perspective, the influence generated by these three types of keyword brand configurations on consumer click-through rate, from positive to negative with generic keyword as the baseline, would be *Brandself_Keyword* > *Brandother_Hotel Keyword* > *Generic Keyword* > *Brandother_OTA Keyword*. As for hotel advertisers, the order can be elaborated as *Brandself_Keyword* > *Brandother_Hotel Keyword* > *Generic Keyword*. Given the result that *Brandother_OTA Keyword* is not significant in Model 2(b), we didn't include it in the order list. However, the negative estimate of *Brandother_OTA Keyword* may somewhat reveal the consequence of including aggregator brands in hotel advertiser's keywords, which is getting lower click-through rates compare with the situations where paid ads are tied to generic keywords.

With regard to the relationship between keyword match and consumer click-through rate, as illustrated in Table 4.6, the coefficients of *Accurate_Keyword* and *Phrase_Keyword* are both positive and significant. Results from the Wald test also show that for both aggregators and hotel advertisers, *Accurate_Keyword* would obtain highest click-through rate, then *Phrase_Keyword*, and *Generic_Keyword* as the baseline has the least click-through rate.

4.3.2 Robustness Check

The robustness check includes three additional tests to examine the consistency of the results generated from the joint probability model in estimating consumer click-through rate. First, we collect advertiser displayed paid ads and consumer clickstream data from another day which is randomly selected. Advertisers chosen for the consistency tests are identical with the advertisers selected in the previous analysis. Three robustness tests are then conducted, and the results are presented in Table 4.9. In the first test, we fit the test data to the joint probability model described before. As can be seen from Table 4.9, the parameter estimates are consistent with the results discussed in the previous sections. In the second test, holding all the variables the same, we introduce a quadratic form of *Rank* (*Rank*^{2}) to the click model to capture the non-linearity effect generated by rankings. The coefficient of *Rank*^{2} is 0.043 (with p < 0.001), showing the exponential decrease of consumer click-through rate as rank goes from one to ten. All the other parameter estimates are still consistent with results presented before. In the third test, we conduct a likelihood ratio (LR) test to investigate the necessity of introducing query related attributes (*Brand Query* and *Brand Match*) while estimating consumer click-through rates. To statistically test the significance of coefficients related to consumer query, we create a reduced model by eliminating the two variables mentioned above from the click model: *Brand Query* and *Brand Match*. As illustrated from Table 4.9, the removal of these two variables leads to biased estimation, where *Brandother Keyword* turned to be negative and significant (-0.124, p < 0.001). Meanwhile, the result of LR test yields 308 $> \chi^2(2) = 13.82$ (with p < 0.001), suggesting that consumer query related variables are not redundant and should be controlled for when estimating consumer click-through rates.

Table 4.9 Empirical Results for Robustness Check

Parameter	**Test 1**	**Test 2**	**Test 3**
Click Model:			
Intercept	-1.828***	-1.561***	-1.787***
	(0.021)	(0.022)	(0.020)
Brand Query	-0.148***	-0.139***	--
	(0.018)	(0.018)	
Brand Match	0.882***	0.455***	--
	(0.050)	(0.050)	
Accurate_Keyword	0.643***	0.677***	0.644***
	(0.015)	(0.015)	(0.015)
Phrase_Keyword	0.462***	0.494***	0.456***
	(0.016)	(0.016)	(0.016)
Brandself_Keyword	0.311***	0.302***	0.961***
	(0.050)	(0.050)	(0.026)
Brandother_Keyword	0.013	0.014	-0.124***
	(0.020)	(0.020)	(0.010)
Rank	-0.425***	-0.694***	-0.427***
	(0.003)	(0.009)	(0.003)
$Rank^2$	--	0.043***	--
		(0.001)	
OTA	0.458***	0.869***	0.416***
	(0.016)	(0.016)	(0.016)
Rank Model:			
Ad Quality	0.384***	0.384***	0.384***
	(0.002)	(0.002)	(0.002)
Bid	0.151***	0.151***	0.151***
	(0.001)	(0.001)	(0.001)
Accurate_Keyword	-1.229***	-1.229***	-1.229***
	(0.009)	(0.009)	(0.009)
Phrase_Keyword	-0.827***	-0.827***	-0.827***
	(0.009)	(0.009)	(0.009)
OTA	0.665***	0.665***	0.665***
	(0.007)	(0.007)	(0.007)

*** $p<0.001$; ** $0.001<p<0.01$; * $0.01<p<0.05$

Finally, we calculate the values of the pseudo R square of the joint probability model. As summarized in Table 4.10, four methodologies have been applied in the calculation. $\ln(\hat{L}(M_{Full}))$ is the estimated log likelihood of the full model, where $\ln(\hat{L}(M_{Full}))$ = -3,202,836. $\ln(\hat{L}(M_{Intercept}))$ is the estimated log likelihood of the reduced model with only intercept parameters included, where $\ln(\hat{L}(M_{Intercept}))$ = -3,511,153. K is the number of parameters that are excluded from the full model. N is the number of observations included in the dataset, which is N = 1,440,660. Likelihood Ratio (LR) Test yields a result which is 616,634 > $\chi^2(14) = 29.14$, indicating that we can reject the null hypothesis, where the reduced form model – with only intercept parameters included – is sufficient.

Table 4.10 Pseudo R Square of Joint Probability Model – Study I

Approach	Function	Value
McFadden	$R^2 = 1 - \frac{\ln(\hat{L}(M_{Full}))}{\ln(\hat{L}(M_{Intercept}))}$	0.08781
McFadden (adjusted)	$R^2 = 1 - \frac{\ln(\hat{L}(M_{Full})) - K}{\ln(\hat{L}(M_{Intercept}))}$	0.08782
Cox & Snell	$R^2 = 1 - \left(\frac{\ln(\hat{L}(M_{Intercept}))}{\ln(\hat{L}(M_{Full}))} \right)^{2/N}$	0
Likelihood Ratio (LR)	$LR = -2\left[\ln(\hat{L}(M_{Intercept})) - \ln(\hat{L}(M_{Full})) \right]$	616,634

4.4 Managerial Implications

The empirical insights generated from the study provide several managerial implications to paid search advertisers. First, our empirical findings show that paid ads owned by different types of advertisers will be prioritized differently by both consumers and the search engines. As we described before, we focus on examining two types of advertisers (online aggregators, hotels) in this study. We find that compare with hotels' paid ads, online travel aggregators' ads not only receive more prominent rankings but also get higher consumer click-through rates. From practical standpoint, it may give advertisers, which have similar characters to either type of the advertisers we examined, some sort of idea about their potential performance when adopting sponsored advertising approach. For online travel aggregators, having higher consumer click-through rates is not merely driven by ads located at higher rankings (ads appeared at the top of the screen). We find that when making click decisions, consumers would differentiate displayed ads between aggregators and hotels, and they are inclined to click on ads owned by aggregators. Consumers react to ads in this way is mostly due to the intrinsic difference between aggregators and hotels. As Ratchford, Talukdar and Lee (2001) discussed, aside from the amount of time to spend, consumer's propensity of choosing a certain type of information source depend upon the quantity and quality of the information that they could obtain. Online travel aggregator (e.g. expedia.com) collects information from a variety of products and services, from airline ticket bookings, hotel reservations to car rentals. The products or services provided are under various price schemes ranging from regular price to deeply discounted price. It also contains "consumer-generated content" sections (e.g. customer reviews) which are viewed as critical components for reducing consumer's perceived risks (Hung and Li 2007). In comparison, hotel websites provide customers with information which usually only related to their own hotel chain information, which might be too limited to satisfy consumer's information seeking requirement.

Meanwhile, consumer's brand awareness also plays an important role in sponsored search advertising. Our empirical results suggest that when consumers' search queries contain brand names, they tend to click on ads with specific brand that match the brand they searched. Also, the size of the paid ads set considered by the consumers is smaller than when consumer search generic queries. That is said, when a consumer enters a branded search query (e.g. "Hilton hotel near Newark airport"), s/he might tend to click on Hilton' ads and are less likely to click on the co-listed ads from Marriott.com. This finding is consistent with the traditional literature from consumer information search (Johnson and Russo 1984; Brucks 1985), where experienced customers search more efficiently and have the tendency to ignore irrelevant information. In the sponsored search settings, our results again highlight the importance of fostering and reinforcing consumers brand awareness towards advertisers' own brand, which, in return will lead to better search engine advertising performance.

Second, our empirical results provide insights into advertiser's keyword configurations. When deciding which specific brand content should be included in the keyword portfolios for the purpose of increasing consumer click-through rate or attracting more consumer visits to advertiser websites, one thing in common for both aggregators and hotels to consider is to select advertisers' own brands. It appears from our analysis that paid ads with the presence of advertisers' own brand names receive highest consumer click-through rate compare with other alternatives (e.g. generic, other branded names included in keywords). On the basis of that, if we define two additional ways of keyword branding: "across-category" branded keywords as aggregator chooses hotel brands in its keywords or vice versa, and "within-category" branded keywords as aggregator chooses other aggregator brands or hotel chooses other hotel brands as its keyword options, our empirical findings suggest that aggregators should employ "across-category" rather than "within-category" branded keywords, whereas hotel advertisers would be more beneficial to use "within-category" instead of "across-category" branded keywords.

We could also consider hotels as service suppliers which manage their own websites providing certain types of service (have the equivalent position as manufacturers), and online aggregators as service retailers which act as intermediaries collecting information from various suppliers and offering information and special packages to customers. In this sense, the advised keyword branding strategy for service retailers, aside from encouraging choosing own brand names, would be to select supplier brand-specific keywords rather than service retailer-specific keywords. This result is somewhat inconsistent with Ghose and Yang (2009), in which they suggested that retailers should favor retailer-specific keywords (e.g. “walmart.com”) rather than manufacturer brand-specific keywords (e.g. “Nautica bedding sheets”) and they pointed out that brand-specific keywords actually led to the decrease of consumer click-through rates. This difference is driven by three aspects. First, our examination is related to advertisers and consumers in the hospitality industry, which is quite different from the situation of merchandise discussed in Ghose and Yang’s paper. Finding a resort place and searching for bedding sheets could involve two completely different thinking and searching processes for consumers. In our settings, consumers require a lot more information before making purchase decisions (Gretzel and Yoo 2008), which could lead to different consumer click behavior among different industries. Second, unlike merchandise industry where manufacturers mostly depend upon retailers and don’t often run their own websites, the suppliers in question have their own websites. Under this situation, service retailers’ ads may not only co-list with other retailers’, but also might co-appear with service suppliers’ ads in the search results. The different combinations of advertisers in the keyword market – one with mostly retailers and the other is the co-existence of retailers and suppliers – may also lead to various keyword selection strategies. Third, given that hospitality industry is information-intensive, advertisers should take into account the characteristics of consumer information search, aside from considering the size and the level of competition of keyword market as described in Ghose and Yang (2009). Suppose a service retailer selected a supplier-specific keyword, for instance, expedia.com chose “Hilton hotel” as its keyword. It is possible that the ad from expedia.com might be co-listed with the ad of Hilton.com. If the consumer’s intention is to compare among various hotel alternatives, expedia.com may experience click-through by the inquired consumer given its nature of containing a lot more information than individual hotel website.

With respect to the keyword match settings, our results show that the presence of accurate keyword experience highest consumer click-through rates compare with other keyword match types (e.g. phrase or broad keyword match), indicating that advertisers – both aggregators and hotels – should focus more on developing keywords that could accurately match consumer search queries in the sense that it may help increase consumer click-through rates. However, given that search queries are created by consumers and are mostly out of the control of advertisers, to successfully set up accurate keywords, advertisers need to really understand and study their targeted customers’ search preference, words and phrases individuals might use in the search procedures. Or, they could depend upon third parties, such as Google Keyword Planner, to design the appropriate keywords to use.

Finally, our empirical results about keyword strategies may also provide some insights into advertiser bidding strategies involved in paid advertising campaigns. One of the possible approaches that advertisers could consider is that when balancing the amount of money spent on individual keyword, advertiser could refer to the direct contribution (e.g. click-through rate) of each keyword. As we described before, the presence of advertiser's own brand or keywords set to be accurate match receives superior click-through performance than generic keywords and broad match keyword. From advertiser's standpoint, it would be a beneficial tactics to make higher bids on keywords (e.g. accurate keyword, keyword with own brand) which have greater potential to obtain higher consumer click-through rate than keywords (e.g. generic keyword, broad matched keyword) that are less likely to be clicked. However, we do acknowledge that the appropriateness of this approach lies in the goals that advertisers plan to achieve throughout search engine campaigns. That said, it might make more sense for an advertiser to execute the suggested approach if its goal is to attract more consumer visits to its own website. However, the approach discussed above might not directly contribute to advertiser's sponsored search campaigns, if the advertiser's goal is to increase its products' conversion rates or to deter other competitors' ads from showing in the search results page.

4.5 Conclusion

Search engine has become one of the presiding online advertising channels and its paid advertising has been widely adopted by advertisers across a variety of industries. From advertiser's viewpoint, to successfully manage paid search campaigns, advertisers need to come up with strategies that are adapted to their own characteristics. The formation of paid search strategies requires advertisers to comprehensively understand the "terrain", which not only includes a bunch of bidding schemes, but also involves ways of keyword selections and basic knowledge about paid search activities and performance undertaken in the related industries.

We conduct this study which mainly focuses on examining the impact of various types of branded keywords on consumer click-through rates in the sponsored search domain, and comparing and contrasting the paid search activities undertaken by two major types of advertisers in the hospitality industry. The empirical insights generated from the joint probability model have been discussed, which can be directly utilized by individual advertiser for developing its own keyword campaigns. The empirical findings regarding different advertisers' paid ads performance may also shed some lights on those companies – which are either inside of the same industry or having similar characters as the advertisers we examined – with respect to their potential performance when entering into the search engine advertising field. Finally, the angle and the approach employed in the study can also be applied to other industries (e.g. merchandise industry).

Our study has several limitations. First, since the dataset is collected from a search engine, there is no way for us to know whether or not a consumer made a purchase, the exact products that have been purchased, or the amount of money spent after the consumer left the search engine and landed on advertisers' websites. Also, it is difficult for us to make any assumptions about the products that consumers purchased, as one of the advertisers examined in this study is online aggregators which contain great variations in prices. Simply calculating the average prices and assigning the numbers to individual advertiser may distort the true facts. Therefore, we had to give up estimations of conversion so that we were unable to build the model following the traditional analytical path (Ghose and Yang 2009; Rutz and Trusov 2011; A. Agarwal, K. Hosanagar and M. D. Smith 2011). In the absence of analyzing keywords' contribution towards conversion rates, our suggestion about advertiser keywords selection is only for the purpose of improving advertisers' paid ads click-through rates performance, or bringing more customer traffic into advertiser's website through the search engine. If possible, future research can be developed by combining data from different sources (search engine and advertiser websites) to examine whether our conclusions regarding keyword selections are still hold.

Second, our analyses about advertiser's keywords selection are mostly focusing on the branded information that different advertisers should consider while operating keywords campaigns. We realize that sometimes consumers may generate queries with identical brand information but have different intentions behind those queries. Consider two queries in which the details are "Hilton hotel" and "Hilton hotel near Newark airport". For the former case, the consumer's intention might be to look for the direct link of Hilton or to acquire some general information about this hotel brand. For the latter case, the consumer specifically mentioned a location name. Combined with the brand name that has also been pointed out, the query might imply that the consumer wants to find if there is a Hilton hotel near the airport and to make a reservation. That is to say, it is important for advertisers to know how to compartmentalize consumer queries and be able to extract patterns for the purpose of developing more comprehensive keywords strategies. A separate paper will be conducted to explore this issue.

Finally, the research is conducted under the hospitality industry, and among the keywords selection strategies we presented, we've already found some differences compare with literature developed in the merchandise industry. That said, future research can be conducted to explore similar topics in other industries. Moreover, one thing that we found interesting from our raw dataset is that there is a small amount of advertisers –

which are not within the hospitality industry – using “cross-industry” keywords strategy to appear along with paid ads that are owned by online aggregators or hotels. Future research can also drill down into this path to examine if it is worthwhile to consider this type of keywords selection approach. We hope that the empirical insights generated from this study could help practitioner better employ keyword selection strategies in sponsored search advertising activities.

CHAPTER 5 ESSAY TWO: EXAMINING THE IMPACT OF ASSORTMENT SIZE AND PAID ADS COMPOSITION IN SPONSORED SEARCH ENGINE ADVERTISING

In this chapter, we study the impact of displayed paid ads assortment size and composition on consumer click-through behavior in sponsored search advertising. Five sections are included in this chapter. In Section 5.1, we briefly introduce the background and discuss the motivation of conducting this study from both academic and managerial perspectives. In Section 5.2, we discuss the methodology, including data analysis and modeling approach for solving the research questions. In Section 5.3, we present the empirical findings and the results of several robustness tests. Section 5.4 explains the managerial implications of the findings. In Section 5.5, the conclusion is made.

5.1 Introduction

As we described in the previous chapter, the reason that sponsored search advertising becomes so intriguing to many advertisers is that it could facilitate a connection between an advertiser and its preferred customers through certain keywords. In the search engine environment, every search is initiated by a consumer typing a series of text string (“search queries”) into the search box. Given the message included in the consumer’s search query, the search engine will then scrutinize its ads pool and provide the consumer with a list of ads displaying in the search results webpage. Paid ads returned by the search engine may be reviewed by the consumer, who will decide which ads to click, based upon the pertinence of displayed ads and his or her search intention. Once the consumer clicks on a web link, s/he will be redirected to the advertiser’s designated website. Perhaps, at the end of his or her visit, the consumer might make a reservation (e.g. book a hotel room) at this advertiser’s website. From the advertiser’s standpoint, this would be considered as a successful sponsored search advertising case, where the advertiser managed successfully in terms of its keyword configuration as well as its bidding schemes and win against other co-listed advertisers' ads by getting a click from the consumer and converting the click into a final purchase at its own website.

In the literature, we have found a number of related studies focusing on one single advertiser (e.g. a national-wide retailer, a major lodging chain) (Ghose and Yang 2009; Rutz and Bucklin 2011; Rutz et al. 2012). One thing that is highly identical regarding these studies is that the analytical path usually starts from consumers seeing the advertiser's ad display, making click decisions based upon paid ads' own characteristics, to patronizing products or services at advertiser's website. Interestingly however, the research that focuses on examining the impact of the number of paid ads shown in the search results on consumer click-through rates, as well as and the impact of composition of displayed paid ads on consumer click-through rates has barely been developed. Though, as it supported by Ariely (2009), sometimes consumers' decision making is relative, meaning that consumer makes decisions base upon the options provided to them. Back to the online search situation discussed in this study, it is highly plausible that when a consumer seeing a list of ads displayed in the search results, his or her click decision might be affected by other co-listed advertisers' ads. Moreover, as pointed out by Rutz and Trusov (2011), the ideal click-through rate estimation should incorporate all alternatives displayed in the search results in the sponsored advertising environment. On the basis of that, traditional retail marketing literature has already spotlighted the importance of integrating assortment size and composition of alternatives in consumer decision making estimation process (Chernev and Hamilton 2009). Therefore, in this study, we take into account the assortment size of displayed paid ads, which is defined as the number of paid ads displayed in the search results, and the composition of displayed ads while estimating consumer's click-through rates with sponsored search advertising environment. We focus on addressing the following research questions: First, will consumer click-through rate be affected by displayed ads' assortment size, or ads composition? More specifically, as the number of displayed paid ads increases, will consumer click-through rate on a single paid ad increase, decrease or remain the same? How does ads composition affect consumer click-through rate? When taking into account the impact of displayed ads size on consumer click-through rate, what kind of keyword strategies that advertisers could develop to improve their sponsored advertising performance in a way that could bring more customer visits to advertisers' own websites (e.g. higher click-through rates)?

We employ a three-stage joint probability modeling approach to examine these research questions. As mentioned in Chapter 3, we focus our empirical analyses in the hospitality industry. Our empirical results show several insights which haven't been explored in the previous literature. First, we find that intense keyword competition, meaning a large number of advertisers participating in a bidding process, will increase the probability of search engine providing a longer list of paid ads in the search result. As the number of displayed ads increases, on average, the probability of a consumer clicking on an individual paid ad will decrease. Interestingly, however, under the same scenario, consumer's average click-through rate on well-known (or top ranked) brands' paid ads will be significantly increased. A notable exception occurs under the condition where consumers enter branded search queries. In this case, the impact of displayed ads size on consumer click-through rate becomes less effective, as the number of displayed ads increases. Second, we find that displayed ads composition also influence consumer average click-through rate. In particular, we examine consumer click-through rate under the condition where attractive ads (e.g. top ranked brand or "accurate match" ads) are placed in a clustered manner at prominent ranks. Our empirical results reveal that the more attractive paid ads accumulated and displayed at relatively top ranks, the less likely that a consumer will click on ads that are shown below them.

Our research contributes to the search engine marketing body of knowledge in the following aspects: First, we consider all paid ads listed in the search results while estimating consumer click-through rates and to the best of our knowledge, this is the first study that empirically examines the relationships between displayed paid ads assortment size and consumer click-through rate. Second, we take into consideration the endogeneity issue generated by displayed ads size and ads ranking and apply a three-stage joint probability modeling approach to simultaneously estimate the covariates. When estimating each individual paid ads click-through rate, we incorporate variables that not only delineate one single advertiser's paid ads own characteristics, but also capture the surrounding paid ads' features (e.g. brand popularity, keyword match). Finally, by referencing to the traditional retail marketing and behavioral marketing literature while demonstrating the implications of our empirical findings, we theoretically bridge the gap between search engine marketing literature and traditional marketing literature.

5.2 Methodology

In the first half of this section, we provide descriptive data analysis to illustrate the dataset we used for conducting this study. In the second half of this section, we present the model framework – the three-stage joint probability model to examine the research questions raised in the previous section.

5.2.1 Data Analysis

In this study, we randomly create a 10,000-search sample data set within one day which leads to 49,510 impressions generated by 11,668 keywords from 676 online advertisers. Since our goal is to examine displayed ads assortment size impact on consumer click-through behavior, each search we selected has at least one click made by a consumer[6]. In total, there are 13,568 clicks collected in the sample dataset. The average click-through rate for each keyword, which is defined as the ratio of total number of clicks to total number of impressions in individual keyword-level, is 19.7%[7]. Similar to the previous study, we purely focus on the paid ads displayed on the top left-side of the screen ranging from rank 1 to 10. Adding to the justification of this confinement which has been made in Chapter 3, we also found that it is impossible for us to tease out the "noise" – customer's unwillingness to click on an ad rather than the size of the displayed ads set actually plays an role in consumer's click-through – directly from the dataset, because whether or not consumers are willing to click on an ad is entirely out of the

[6] The justification of randomly selecting data records which has at least one click is shown in Robustness Check section.

[7] We notice that the average click-through rate is higher than it is from the literature. The reasons of having high click-through rate is that: first, unlike the ads location examined in the literature, which is right-sided ads, we examine ads display on the left-side of the screen. Our preliminary analysis shows that click-through rates are significantly different compare ads displayed on both sides. Second, we eliminate the cases where there is no click generated given a consumer search query, which drives the average click-through rate even higher.

scope of our observations. Moreover, Greenspan (2004) stated that consumers are reluctant to make clicks on ads shown on the right-side of the screen given that they are aware that all ads displayed in that specific locations are sponsored ads, the click-through rates of ads displayed on the right-side are generally significantly low. For instance, the average click-through rate for a given ad with keyword “Holiday Inn” displayed on the left side is 8.3%, compare with the average click-through rate, which is 3.9%, for the same ad displayed on the opposite side. Therefore, we conduct analysis by only looking at ads displayed on the top left.

Table 5.1 and Figure 5.1 illustrate a brief summary of the dataset. As can be seen from Figure 5.1, starting from size equals to 2, the average click-through rate in aggregate-level consumer clickstream is decreased as displayed paid ads size grows. Meanwhile, the average advertiser bid is increased by 40 cents for each keyword as displayed ad size becomes larger (from 4 to 10), indicating that that the level of competition among advertisers is more intense. With regard to ads quality, although ads quality first goes down (between 2 to 4) then up a little bit (between size = 4 to 6), the main trend of ads quality is decreased as well. Finally, from the diagram, the size has a “bell” shape distribution in our sample set.

Table 5.1 Summary of Size Specific Performance

Size	total impression	total click	average bid	average ads quality
1	707	707	3.56	2.15
2	1866	1053	2.23	1.78
3	3624	1531	2.03	1.66
4	5896	1975	1.96	1.63
5	7550	2154	2.04	1.65
6	8718	2088	2.30	1.69
7	9086	1920	2.54	1.73
8	6656	1274	2.52	1.71
9	3987	646	2.44	1.65
10	1420	220	2.59	1.59

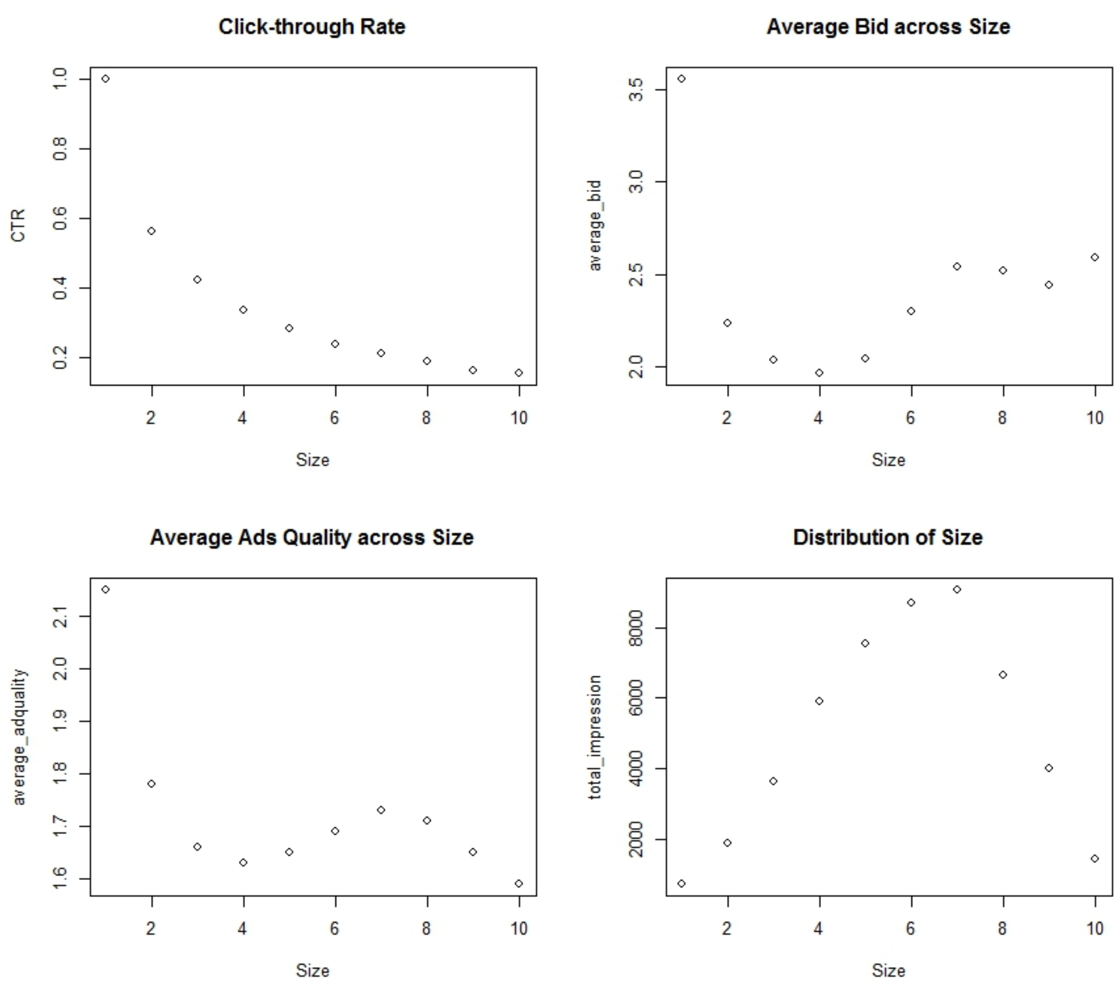

Figure 5.1 Plots of Size Specific Performance

We further decompose the dataset into two categories: OTAs and Hotels. Table 5.2 summarizes some of the basic descriptive analysis for both categories.

Table 5.2 Summary of Two Types of Advertisers (OTA vs. Hotel)

Advertiser Type	Variables	Mean	Std.dev.	Min	Max
OTA (341)	Rank	3.36	2.01	1	10
	Ads Quality	1.71	0.93	0.12	9.51
	Bid ($)	2.36	2.62	0.40	28.04
	Click	28%	0.45	0	1
Hotel (335)	Rank	4.51	2.26	1	10
	Ads Quality	1.49	0.87	0.27	9.11
	Bid ($)	1.87	1.91	0.40	25.58
	Click	19%	0.39	0	1

As shown in Table 5.2, OTAs on average experience a slightly superior ranking performance than Hotel websites. Although OTAs also have higher quality paid ads (1.71, compared with 1.49 of hotels' ads), the amount of bids made by OTAs is approximately 26% higher than Hotels on an average level, which may also to some extent reflect its emphasis in sponsored search advertising field. Finally, OTAs on average receive higher consumer click-through rates than Hotels.

We also provide two tables with summaries of size specific performance for each type of advertisers (See Table 5.3 and 5.4). As shown in both tables, the performance of two types of advertisers is quite consistent as what we observed in Table 5.2, even when we examine each type advertisers size specific performance in terms of the average bid, ads quality and click-through rates.

Table 5.3 Summary of Size Specific Performance of OTA's Ads

size	total impression	total click	average bid($)	average ads quality
1	631	631	3.57	2.17
2	1698	961	2.23	1.79
3	3316	1396	2.06	1.63
4	5379	1813	1.99	1.62
5	6822	1999	2.07	1.67
6	7845	1945	2.35	1.71
7	8033	1779	2.62	1.76
8	5771	1183	2.61	1.76
9	3381	611	2.56	1.71
10	1192	204	2.67	1.67

Table 5.4 Summary of Size Specific Performance of Hotels' Ads

size	total impression	total click	average bid($)	average ads quality
1	76	76	3.48	2.01
2	168	92	2.27	1.67
3	308	135	1.73	1.90
4	517	162	1.60	1.69
5	728	155	1.73	1.52
6	873	143	1.87	1.49
7	1053	141	1.93	1.47
8	885	91	1.92	1.38
9	606	35	1.74	1.29
10	228	16	2.15	1.20

5.2.2 Model Framework

The effectiveness of click-through rate estimation is usually a major concern to many online advertisers, since estimating consumer click-through rate in an effective manner can help advertisers determine their bidding strategies, as well as decide which keywords should be included in the keyword portfolios. However, advertisers are commonly facing challenges stem from the incapability of observing other competitors' movement and search engine's proactive management of ads positions. These challenges restrain each individual advertiser from estimating consumer click-through rate using a broader range of attributes, for example, taking assortment size impact into estimation process. Using the disaggregate-level data collected from the search engine, we take into account the assortment size impact – which is unobserved by individual advertisers – to estimate consumer click-through rate. In order to account for the behavior from three entities – advertiser, search engine and consumer, we apply a three-stage joint distribution modeling approach to estimate consumer click-through probabilities. We begin with a count model to estimate the probabilities of the number of paid ads to be displayed in the search results. Followed by a ranked-ordered logit model, we then discuss the rank order decision made by the search engine. We use a binary logit model to estimate consumer's click probabilities which are conditional upon the given size and rank order of the paid ads. The multiplication of the three model segments constructs our three-stage joint distribution model and we use maximum likelihood estimation to calculate the parameter coefficients.

5.2.2.1 Size Model

When the search engine receives a search query sent by a consumer, first of all, it needs to decide how many ads to be displayed in the search results and where to locate each individual ad (e.g. left-sided versus right-sided display). Assume that the search engine decides to show a group of left-sided ads, we define the term "size" as the number of paid ads displayed in the left region. The range of the available slots for paid ads within this area is from 1 to 10. To model the "size" decision made by the search engine, we employ a Poisson regression model (PRM), a standard approach for modeling the situation where the dependent variable is a count (Winkelmann 2008). Assume that S_i is the actual size determined by the search engine in response to consumer query i and it follows Poisson distribution $s_i \sim Poisson(\lambda_i)$, $s_i \in S$ where S is a set of all possible sizes span from 1 to 10. The size model is specified in equation (1) and (2):

$$P(S_i = m) = \frac{\exp(-\lambda_i)\lambda_i^{\ m}}{m!} \tag{1}$$

$$\lambda_i = \exp(\alpha_1 + \alpha_2 QueryPopularity_i + \alpha_3 KeywordCompetition_i) \tag{2}$$

In equation (2), λ_i is specified as an exponential function of a linear index of two explanatory variables: query popularity and keyword competition. An examination of the correlation between these two variables is presented in Table 5.5.

Table 5.5 Correlation between Query Popularity and Keyword Competition

Pearson Correlation Coefficients, N = 10,000 Prob > |r| under H0: Rho=0		
	Query Popularity	Keyword Competition
Query Popularity	1.00000	0.22693 <.0001
Keyword Competition	0.22693 <.0001	1.00000

Query popularity is a discrete variable which captures the historical data of the search volume made by consumers for each individual query in a given day. *Keyword competition* measures the total number of advertisers bid for each specific consumer query. Depend upon how much information is carried in a consumer query, sometimes the search engine may return paid ads with multiple keywords. For example, when a consumer searched for “pet friendly cheap hotel”, the search engine might simultaneously show ads with keyword “pet friendly hotel” and ads with keyword “cheap hotel” in the search results – in which case, there are two distinct keywords being triggered under the consumer search and the “keyword competition” variable becomes the summation of the number of advertisers bid for keyword “pet friendly hotel” as well as the number of advertisers bid for keyword “cheap hotel”.

5.2.2.2 Rank Model

After the search engine decided the number of left-sided ads to be displayed in the search results, it will subsequently determine what and how to rank order the selected ads. Without knowing the actual ads selection and rank order algorithm implemented by the search engine, we apply a Rank-Ordered Logit Model (ROL) introduced by Beggs et al. (1981), as an approximate approach to model the rank order decision made by the search engine. Except for the ease of computation, the probability model specification is capable of capturing the entire ordering process and providing empirical insights regarding how the search engine might evaluate each of the attributes, which are characterized from a set of paid ads. Assume that the search engine is a rational utility maximizer and is facing a rank-order decision of a set of paid ads being generated by the search engine in response to consumer query *i*. The size of the paid ads set is denote as S_i ($1 \le S_i \le 10$), which consists of J_i paid ads. Let X_{ij}^{SE} represents a vector of observed attributes of paid ad *j* given consumer query *i*. The random utility function is specified as: $U_{ij}^{SE} = V_{ij}^{SE} + \varepsilon_{ij}^{SE}$, where ε_{ij}^{SE} are the stochastic error terms assumed to be identically and independently distributed with the extreme value distribution. V_{ij}^{SE} is the representative utility and can be specified as a linear function of X_{ij}^{SE}. Consistent with the literature (Ghose et al. 2014; Ghose and Yang 2009), we consider bid and ads quality as two important attributes that constructs X_{ij}^{SE}. Adding to the literature, we then introduce a third attribute which captures the level of “match” between query *i* and the keyword embedded in paid ad *j*, denote as dummy variables *Accurate Keyword* and *Phrase Keyword*. We also control for the brand popularity[8] of paid ad *j*, denote as *Topbrand*, a dummy variable which captures whether or not the displayed ad belongs to a top-ranked brand group, and paid ad *j*'s advertiser type, denote as *OTA*, a dummy variable representing whether a displayed ad belongs to an online travel aggregator (*OTA* = 1) or a hotel company (*OTA* = 0). Therefore, the representative utility function V_{ij}^{SE} is elaborated as:

$$\begin{aligned} V_{ij}^{SE} &= \beta_1 Bid_{ij} + \beta_2 AdQuality_{ij} + \beta_3 AccurateKeyword_{ij} + \beta_4 PhraseKeyword_{ij} \\ &\quad + \beta_5 Topbrand_{ij} + \beta_6 OTA_{ij} \end{aligned} \tag{3}$$

In the position allocation process, even though U_{ij}^{SE} are unobserved, we assume that the search engine will give ad *j* a more prominent rank than ad *k* when $U_{ij}^{SE} \geq U_{ik}^{SE}$. Let R_{ij} denotes the observed ranking for ad *j* given query *i*. Let $\delta_{ijk} = 1$ when $R_{ik} \geq R_{ij}$ and 0 otherwise. Let C_i represents the arrangement and the composition of the selected paid ads. Therefore, according to Chapman and Staelin (1982), the rank-order function is specified in equation (4)[9]:

$$P(U_{i1} \geq U_{i2} \geq \cdots \geq U_{ij}) = P(C_i \mid S_i) = \prod_{j=1}^{J_i} \left[\frac{\exp(V_{ij}^{SE})}{\sum_{k=1}^{J_i} \delta_{ijk} \exp(V_{ij}^{SE})} \right] \tag{4}$$

---

8 We collected brand popularity information from a secondary data source provided by the biggest online research institute in China (CNNIC). Brand popularity is measured as a list of brands that are ordered in a descending way based upon the historical daily search volume made by consumers. We select the top ten online advertisers (five online aggregators and five hotel brands) and denote them as a top-ranked brand group. When an advertiser falls into this group, the covariate *Topbrand* is marked as 1 and 0 otherwise.

9 The derivation of the reduced form of the rank ordered logit model can be found in Appendix D.

5.2.2.3 Click Model

Conditional upon the search results, which is given as a certain number of paid ads displayed on the left side of the screen in a specific order, a consumer will start looking at each individual ad and make click decisions. A common perspective regarding consumer's online browsing pattern is that consumers scan through the search results from the topmost to the bottom of the screen (Athey and Ellison 2008; Chen and He 2011; Feng et al. 2007). When a consumer is going through the search results webpage, she might be relating each ad to the search query that she entered at the beginning of the search and deliberating if there is any ad that satisfies her search requirement. If there is one, she might tend to click on the related web link. We assume that a consumer will click on paid ad *j* given query *i* when her perceived utility of clicking on ad *j*, denote as U_{ij}^{C}, is greater than the perceived utility of not clicking on ad *j*, which is 0. The perceived utility generated by consumers can be represented in a random utility function, specified as $U_{ij}^{C} = V_{ij}^{C} + \varepsilon_{ij}^{C}$, where ε_{ij}^{C} are the stochastic error terms assumed to be identically and independently distributed with the extreme value distribution. V_{ij}^{C} is the deterministic portion and can be specified as a linear function of X_{ij}^{C}. X_{ij}^{C} consists attributes from three main aspects: 1) paid ad *j*'s own characteristics: including paid ad *j*'s keyword match type, keyword content, ranking position, brand popularity and corresponding advertiser type; 2) assortment size: the number of the displayed paid ads when paid ad *j* was shown in the search results; and 3) the attractiveness of the surrounding ads which are simultaneously displayed with and laid out in front of paid ad *j*. Therefore, the representative utility V_{ij}^{C} can be elaborated in equation (5):

$$\begin{aligned} V_{ij}^{C} &= \gamma_0 + \gamma_1 AccurateKeyword_{ij} + \gamma_2 PhraseKeyword_{ij} + \gamma_3 BrandselfKeyword_{ij} \\ &+ \gamma_4 BrandotherKeyword_{ij} + \gamma_5 Rank_{ij} + \gamma_6 Topbrand_{ij} + \gamma_7 OTA_{ij} \\ &+ \gamma_8 Size_{ij} + \gamma_9 Size_{ij}^2 + \gamma_{10} Rank_{ij} \times Size_{ij} + \gamma_{11} Topbrand_{ij} \times Size_{ij} \\ &+ \gamma_{12} Topbrand\_cum_{ij} + \gamma_{13} Accurate\_cum_{ij} + \gamma_{14} OTA\_cum_{ij} \\ &+ \gamma_{15} BrandQuery_{ij} \end{aligned} \tag{5}$$

In equation (5), paid ad *j*'s own characteristics are captured using the following variables: *Accurate Keyword* and *Phrase Keyword* are dummy variables representing the level of match between ad *j*'s keyword and consumer query *i*. *Brandself Keyword* and *Brandother Keyword* are dummy variables indicating how paid ad *j*'s keyword is branded. When the advertiser of ad *j* includes its own brand name in the keyword, *Brandself Keyword* equals to 1, otherwise it is 0. Similarly, when the advertiser of ad *j* embedded its competitor's brand name in the keyword, *Brandother Keyword* equals to 1 and 0 otherwise. *Rank* represents paid ad *j*'s ranking position under consumer query *i*. *Topbrand*, as described in the previous section, captures whether ad *j* is a top-ranked brand (*Topbrand* = 1) or a regular brand (*Topbrand* = 0). *Size* captures the number of the displayed paid ads when paid ad *j* was shown. We also consider a quadratic term, denote as $Size^2$, to capture the potential non-linearity impact generated by the covariate of *Size*. To capture the attractiveness of the surrounding paid ads that are laid out in front of ad *j*, we introduce two discrete variables: *Topbrand_cum* and *Accurate_cum*, based upon the related behavioral assortment size literature (Chernev and Hamilton 2009). *Topbrand_cum* represents the number of paid ads that are top-ranked brands displayed in front of ad *j* under consumer query *i*. *Accurate_cum* measures the number of paid ads with "accurate match" keywords shown in front of ad *j*. Also, consider the situation where paid ads provided by two distinct types of advertisers (e.g. online travel aggregators versus hotel) may also arouse different consumer's perception of attractiveness, we add a third variable, denote as *OTA_cum*, which captures the number of paid ads which belong to online travel aggregators displayed in front of ad *j*.

Based upon the assumption that consumer browsing pattern is from top to bottom, we don't take into account the attractiveness of paid ads displaying below ad *j*, since it is possible that when a consumer is trying to decide whether or not to click on ad *j*, she may not even notice the ads shown below. Moreover, many consumers believe that search engines provide search results from the most attractive (or most relevant) option to the least attractive option. We also examine the heterogeneity of ad ranking and brand popularity in different assortment sizes by adding two additional interaction variables ($Rank_{ij} \times Size_{ij}, Topbrand_{ij} \times Size_{ij}$). Lastly, we control for consumer query *i*'s character. *Brand Query* represents whether consumer query *i* is a branded query (*Brand Query* = 1) or a generic query (*Brand Query* = 0).

Consider that consumers oftentimes hold different intentions as they initiate their search procedure (Muramatsu and Pratt 2001), and unlike the situation of making a purchase decision, clicking on a paid ad can be easily achieved which doesn't require any money spending, sometimes there are multiple clicks occurred within a single consumer search. In this research, since our goal is to analyze assortment size impact on consumer click-through rate rather than examine consumer sequential click behavior based on consumer learning, for multiple-click cases, we assume that consumer's each time click within a single search are independent. Therefore, conditional upon the given paid ads size S_i and the ads arrangement and composition C_i, let Y_{ij} represents consumer actual click status on ad j (Y_{ij} = 1 clicked and 0 otherwise), we apply a simple binary logit model to model consumer click decision, which is specified in equation (6):

$$P(Y_{ij} = 1 | C_i, S_i) = \frac{\exp(V_{ij}^C)}{1 + \exp(V_{ij}^C)} \tag{6}$$

5.2.2.4 Three-Stage Joint Probability Model

Given the three model segments described in the previous sections: 1) size model – capturing the search engine's decision on the number of ads to be displayed; 2) rank-ordered model – examining the search engine's ads ranking decision conditional upon the given size; and 3) click model – estimating consumer click probability on each alternative ad based upon the existing ads arrangement and the given size, and each segment is independent from others, we construct the joint distribution model by multiplying the three segments, which is specified in equation (7):

$$P(Y_{ij}, C_i, S_i) = P(Y_{ij} | C_i, S_i) \cdot P(C_i | S_i) \cdot P(S_i = m) \tag{7}$$

Based upon the specification of the joint distribution function, the likelihood and log-likelihood function is developed in equation (8) and (9), where n represents the number of consumer searches[10]. We use maximum likelihood estimation approach to simultaneously estimate the coefficients included in the model, the empirical findings are discussed in the next session.

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m} P(Y_{ij} = 1, C_i, S_i)^{Y_{ij}} \cdot P(Y_{ij} = 0, C_i, S_i)^{1-Y_{ij}} \tag{8}$$

[10] The derivation of the log-likelihood function can be found in Appendix D.

$$Log(L)=\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\left(Y_{ij}\cdot\log(\frac{\exp(V_{ij}^{C})}{1+\exp(V_{ij}^{C})})+(1-Y_{ij})\cdot\log(\frac{1}{1+\exp(V_{ij}^{C})})\right)+\sum_{j=1}^{m}\left(V_{ij}^{SE}-\log\left(\sum_{k=1}^{m}\delta_{ijk}\exp(V_{ik}^{SE})\right)\right)+\log\left(P(S_i=m)\right)\right) \quad (9)$$

5.3 Empirical Analysis

We start this section by briefly describing the empirical results estimated from the three-stage joint distribution model we discussed in the previous section. Then, we show several robustness tests that we conducted in order to examine the consistency of the main empirical findings generated from our model.

5.3.1 Parameter Estimation

In this section, we separately discuss our main empirical results from three perspectives: paid ads size, ads rank and consumer click-through rate.

5.3.1.1 Paid Ads Size

As shown in Table 5.6, *Keyword Competition* has a positive and significant impact on paid ads size, indicating that when there are more advertisers participating in a bidding process for a specific consumer search query, the probability that the search engine will provide larger number of displayed paid ads in the search results will be higher. However, the coefficient of *Query Popularity* is positive but not significant, showing that even though a consumer query is popular, it may not necessarily increase the probability of search engine returning larger paid ads set in the search results.

Table 5.6 Coefficient Estimate on Paid Ads Size

	Mean (s.e.)	*P-value*
Intercept	**1.165** **(0.010)**	<0.001
Query Popularity	0.002 (0.004)	0.541
Keyword Competition	**0.204** **(0.004)**	<0.001

Intuitively, the search engine should prefer to provide more alternative paid ads to consumers when the related queries are popular. One plausible explanation for *Query Popularity* being insignificant on paid ads size is that, although advertisers might extract keywords from user-generated content, such as product reviews or blogs (Dhar and Ghose 2010), and user-generated content oftentimes can facilitate the formation of popular consumer queries (McCarthy 2010), advertisers will also need to take into consideration of keywords that could highlight product or service features (e.g. "economical hotel with $59.99 per night") and identify potential customer segments (e.g. "economical hotel for backpackers"). Under this circumstance, the selected keywords may not be directly drawn from popular consumer queries (e.g. "economical hotel" with approximately 14,500 consumer searches each day). If a consumer entered a popular query with a limited number of advertisers involved, there might a lower probability that the search engine will show larger paid ads set in the search results.

5.3.1.2 Paid Ads Rank

Table 5.7 shows the empirical results on paid ads rank order decision made by the search engine. Our analysis reveals that, all four covariates, including *Bid*, *Ads Quality*, *Accurate Keyword* and *Phrase Keyword*, are positive and significant. First, the positive impact of *Bid* on ads ranking positions indicates that the more amount of money that an advertiser bids on keywords, the higher probability that the search engine will assign advertiser's paid ads at more prominent ranks in the search results. Next to *Bid, Ads Quality* shows its significance in the search engine's ranking process as well, suggesting that the search engine intends to place higher quality ads at more prominent locations. Moreover, the significance of both covariates, *Accurate Keyword* and *Phrase Keyword*, indicates that the search engine also takes into consideration the relevance between displayed paid ads keywords and consumer queries. Among others, the two control variables, *Topbrand* and *OTA*, are positive and significant, showing that the search engine prefers to locate paid ads – which either fall into top ranked brand group or belong to online travel aggregator category – to more prominent ranks.

Table 5.7 Coefficient Estimates on Paid Ads Ranking

	Mean (s.e.)	*P-value*
Bid	**3.630** **(0.031)**	<0.001
Ads Quality	**2.573** **(0.025)**	<0.001
Accurate Keyword	**0.278** **(0.053)**	<0.001
Phrase Keyword	**0.349** **(0.052)**	<0.001
Topbrand	**0.451** **(0.031)**	<0.001
OTA	**0.176** **(0.020)**	<0.001

5.3.1.3 Consumer Click-Through Rate

Finally, consider the empirical results on consumer click-through rate. As shown in Table 5.8, there is a negative significant relationship between *Size* and consumer click-through rate, which indicates that the probability of a consumer clicking on a single paid ad is decreased as the size of displayed paid ads set becomes larger. The quadratic term capturing the non-linearity effect of *Size* ($Size^2$) is positive and significant, showing that there is a concave relationship (as a U-shape) between *Size* and click-through rate. Given that the coefficient of *Size* and $Size^2$ is -0.499 and 0.020 respectively, we can further compute the inflection point where the impact of *Size* on consumer click-through rate alters from negative to positive, and it turns out that the alternation happens when displayed paid ads size equals to 13. However, since the boundary of *Size* in our study is in between 1 and 10 displayed ads, which does not surpass the inflection point (*Size* = 13), the impact of *Size* on the rate of change in click-through rate is decreasing as *Size* becomes larger (from 1 to 10 in our case). Consistent with previous literature (Feng et al. 2007; Ghose and Yang 2009), *Rank* has a negative impact on consumer click-through rate. That is, consumer click-through rate decreases from top rank to the bottom. Interestingly however, the interaction effect of *Rank* on the relationship of *Size* with click-through rate is positive and significant, suggesting that the rate of decrease in click-through rate influenced by ranking positions is attenuated when *Size* becomes larger. Similarly, another interaction effect of *Topbrand* on the relationship of *Size* is positive and significant, indicating that the probability of a consumer clicking on a top-ranked brand paid ad will be increasing when there is larger displayed paid ads set shown in the search results. In other words, the power of top-ranked brands is manifested as the search engine returns more alternative ads to the search results.

Next, we show empirical results regarding the impact of the attractiveness of paid ads on consumer click-through rate. Our analysis reveals that, both covariates, *Topbrand_cum* and *Accurate_cum* have negative and significant impact on consumer click-through rate, suggesting that the more top-ranked brand ads or ads with "accurate match" keywords displayed in front of ad *j*, the less likely that a consumer will click on ad *j*. On the other hand, the coefficient of *OTA_cum* is not significant, indicating that the number of online travel aggregators displayed in front of ad *j* may not necessarily lead to a reduced click-through rate on ad *j*. This finding suggests that even though online travel aggregators invest large amount of money on daily budgets and sometimes occupy prominent ranking positions in front of hotel brands paid ads, the number of travel aggregators' ads display will not significantly decrease the consumer click-through rate on hotel paid ads which is located below the clustered online travel aggregators' paid ads.

Lastly, we show the relationships between keyword characteristics – in terms of keyword match type and keyword branded content – and consumer click-through rate on individual paid ads. For keyword match type, both covariates, *Accurate Keyword* and *Phrase Keyword*, are positive and significant, suggesting that paid ads with "accurate keyword match" or "phrase keyword match" get higher consumer click-through rate, compare with the baseline where paid ads are set to be "broad keyword match". We further conduct a Wald test to differentiate the impact on consumer click-through rate generated by these two types of keyword match. The result from the statistical test shows that the estimate of *Accurate Keyword* is significantly greater than the estimate of *Phrase Keyword* ($10.549 > Z_{0.001}=3.08$), which indicates that, consumers prefer to click on "accurate match" paid ads than "phrase match" paid ads. For keyword branded content, our empirical analysis exhibits that, consumer click-through rate on ads with keyword that embedded advertiser's own brand information is significantly greater than ads with generic keywords, whereas there is no significant difference in terms of consumer click-through rate between ads with keywords that include advertiser competitor's brand name and ads with generic keywords.

Table 5.8 Coefficient Estimates on Consumer Click-Through Rate

	Mean (s.e.)	*P-value*
Size	**-0.499** **(0.027)**	<0.001
Size^2	**0.020** **(0.003)**	<0.001
Rank	**-0.938**	<0.001

	(0.046)	
*Rank*Size*	**0.078** **(0.004)**	<0.001
Topbrand	-0.073 (0.068)	0.287
*Topbrand*Size*	**0.067** **(0.011)**	<0.001
Topbrand_Cum	**-0.066** **(0.022)**	0.003
Accurate_Cum	**-0.276** **(0.015)**	<0.001
OTA_Cum	0.026 (0.034)	0.445
Accurate Keyword	**0.520** **(0.053)**	<0.001
Phrase Keyword	**0.163** **(0.054)**	0.003
Brandself Keyword	**0.945** **(0.104)**	<0.001
Brandother Keyword	0.071 (0.062)	0.260
Brand Query	**-0.268** **(0.057)**	<0.001
OTA	**0.158** **(0.049)**	0.012

5.3.2 Robustness Check

We conducted several additional robustness tests to examine the consistency of the main empirical results that we derived from our three-stage joint distribution model (See Table 5.9 and 5.10). We first applied the same modeling approach to a separate randomly generated dataset with the same configurations as we used in the previous empirical analyses. The main results from the robustness test are consistent with the original results. Consistent with the literature, we then added a quadratic form of *Rank* ($Rank^2$) to the consumer click-through equation to capture the non-linearity effect of

Rank. Our main empirical results still hold. Finally, taking into account the heterogeneity issue of consumer search (e.g. informational search versus navigational search), we included consumer query character (e.g. branded versus generic) as a proxy in capturing consumer different search intentions – seeking information or searching for a specific brand (Broder 2002). By adding one more interaction term (*BrandQuery*×*Size*) to the consumer click-through model segment, we examined whether the assortment size impact on consumer click-through rate on individual paid ads will be significantly different given consumers' diverse search intentions. The coefficient of *BrandQuery*×*Size* is positive and significant, indicating that when a consumer entered a branded search query, her average click-through rate will be less affected by the displayed ads assortment size, compare with the situation where search queries are generic. Meanwhile, the results from the robustness test are still consistent with the main results we get from the original joint distribution model.

Table 5.9 Summary of Robustness Test Results -- Part I

	Coefficient Estimates		
	Test 1	*Test 2*	*Test 3*
Size Model:			
Query Popularity	0.003	0.003	0.003
	(0.004)	(0.004)	(0.004)
Keyword Competition	**0.199*** **	**0.199*** **	**0.199*** **
	(0.004)	**(0.004)**	**(0.004)**
Rank Model:			
Bid	**3.608*** **	**3.608*** **	**3.608*** **
	(0.031)	**(0.031)**	**(0.031)**
Ad Quality	**2.546*** **	**2.546*** **	**2.546*** **
	(0.024)	**(0.024)**	**(0.024)**
Accurate Keyword	**0.399*** **	**0.399*** **	**0.399*** **
	(0.051)	**(0.051)**	**(0.051)**
Phrase Keyword	**0.468*** **	**0.467*** **	**0.468*** **
	(0.051)	**(0.051)**	**(0.051)**
OTA	**0.450*** **	**0.450*** **	**0.450*** **
	(0.031)	**(0.031)**	**(0.031)**
Topbrand	**0.199*** **	**0.199*** **	**0.199*** **
	(0.020)	**(0.020)**	**(0.020)**

Table 5.10 Summary of Robustness Tests Results -- Part II

	Coefficient Estimates		
	Test 1	*Test 2*	*Test 3*
Click Model:			
Size	**-0.488***** **(0.028)**	**-0.470***** **(0.028)**	**-0.482***** **(0.028)**
Size^2	**0.020***** **(0.003)**	**0.023***** **(0.003)**	**0.024***** **(0.003)**
Rank	**-0.878***** **(0.045)**	**-0.934***** **(0.045)**	**-0.934***** **(0.045)**
*Rank*Size*	**0.074***** **(0.004)**	**0.048***** **(0.006)**	**0.048***** **(0.006)**
Rank^2	--	**0.030***** **(0.004)**	**0.030***** **(0.004)**
Topbrand	-0.006 (0.069)	-0.027 (0.069)	-0.019 (0.069)
*Topbrand*Size*	**0.053***** **(0.011)**	**0.059***** **(0.012)**	**0.058***** **(0.012)**
Topbrand_Cum	**-0.119***** **(0.022)**	**-0.145***** **(0.022)**	**-0.146***** **(0.022)**
Accurate_Cum	**-0.220***** **(0.015)**	**-0.209***** **(0.015)**	**-0.208***** **(0.015)**
OTA_Cum	-0.002 (0.033)	0.022 (0.033)	0.021 (0.033)
Accurate Keyword	**0.387***** **(0.052)**	**0.389***** **(0.052)**	**0.392***** **(0.052)**
Phrase Keyword	0.092 (0.052)	0.099 (0.052)	0.101 (0.052)
Brandself Keyword	**0.826***** **(0.099)**	**0.799***** **(0.099)**	**0.805***** **(0.099)**
Brandother Keyword	0.084 (0.061)	0.082 (0.062)	0.075 (0.061)
Brand Query	**-0.227***** **(0.055)**	**-0.219***** **(0.055)**	**-0.408***** **(0.097)**
*Brand Query*Size*	--	--	**0.038*** **(0.016)**
OTA	**0.102*** **(0.048)**	0.092 (0.048)	0.093 (0.048)

Next, we show the empirical evidence in supporting our sample selection process used in this study, where each of the searches included in the sample should contain at least one click made by consumers. To do that, we randomly selected another sample set (which contains 10,000 searches). It includes searches with at least one click, searches with zero clicks generated by consumers. We then fit the data to the model and compare the results generated from these two different sets. As illustrated in Table 5.11 below, the key parameters, such as *Size*, *Size*^{2} and *Brand Query* generated from the test sample are significantly biased.

Table 5.11 Comparison between Two Different Sample Sets

Parameters	*Original sample*	*Test Sample*
	Mean	*Mean*
Size	-0.499	0.252
Size^2	0.020	-0.023
Accurate Keyword	0.520	0.454
Phrase Keyword	0.163	0.434
Brandself Keyword	0.945	0.546
Brandother Keyword	0.071	-0.053
Brand Query	-0.268	0.062

Finally, we calculate the values of the pseudo R square of the joint probability model. As summarized in Table 5.12, four methodologies have been applied in the calculation. $\ln(\hat{L}(M_{Full}))$ is the estimated log likelihood of the full model, where $\ln(\hat{L}(M_{Full}))$ = -64,726.49. $\ln(\hat{L}(M_{Intercept}))$ is the estimated log likelihood of the reduced model with only intercept parameters included, where $\ln(\hat{L}(M_{Intercept}))$ = -71446.67. *K* is the number of parameters that are excluded from the full model. *N* is the number of observations included in the dataset, where *N* = 49,510. Likelihood Ratio Test generates a result which is 13440.36 > $\chi^2(17) =$ 33.41, indicating that we can reject the null hypothesis, where the reduced form model – with only intercept parameters included – is sufficient.

Table 5.12 Pseudo R Square of Joint Probability Model – Study II

Approach	Function	Value
McFadden	$R^2 = 1 - \frac{\ln(\hat{L}(M_{Full}))}{\ln(\hat{L}(M_{Intercept}))}$	0.09406
McFadden (adjusted)	$R^2 = 1 - \frac{\ln(\hat{L}(M_{Full})) - K}{\ln(\hat{L}(M_{Intercept}))}$	0.09382
Cox & Snell	$R^2 = 1 - \left(\frac{\ln(\hat{L}(M_{Intercept}))}{\ln(\hat{L}(M_{Full}))} \right)^{2/N}$	0
Likelihood Ratio	$LR = -2\left[\ln(\hat{L}(M_{Intercept})) - \ln(\hat{L}(M_{Full})) \right]$	13440.36

5.4 Managerial Implications

The empirical insights generated from our three-stage joint distribution model provide several implications to advertisers in the paid search advertising area. First, our empirical results regarding search engines giving priority to more relevant paid ads suggest that, in order to be placed in more prominent rank positions in the search results, advertisers should not only focus on optimizing paid ads bid strategies, but also need to develop effective keyword campaigns in a way that could both highlight advertisers' products (or service) features and cater to their potential targeted consumers' preference.

Second, our empirical findings on displayed ads assortment size impact on consumer click-through rate show that it is important for online advertisers to take into consideration the impact of displayed ads assortment size while they are setting up their keyword schemes in the pre-auction stage. Conventional wisdom suggests that advertisers should choose popular keywords in order to increase the frequency of impressions in the search results. Such keywords are usually defined as words or short phrases that get high volume of consumer search and a large number of advertisers' participation. The rationale behind this statement is that, by repetitively displaying inside of consumers' search results webpage, ads with popular keywords could deepen consumer's brand awareness, lead to consumer's click-through and eventually convert into consumer purchases. However, our empirical results show that simply selecting popular keywords may not improve advertiser's ads click-through rate performance; instead, it may even lead to the opposite results. Since when a keyword becomes more competitive, meaning that there are a larger number of advertisers bidding for such keyword, the probability of the search engine returning a larger number of alternative ads in the search results will be higher. Traditional behavioral research has shown that when consumers are facing bigger alternative sets, they need to spend extra cognitive efforts to choose between items (Chernev et al. 2003; Schwartz 2004). And if consumers feel overwhelmed comparing alternatives, instead of continuing searching for the "perfect match" (Lancaster 1990), they tend to simplify their decision process by directly choosing the alternative(s) that are perceivably attractive or well-known in the marketplace to reduce cognitive efforts as well as lower perceived risks (Botti and Iyengar 2004; Gourville and Soman 2005; Huffman and Kahn 1998). Similar to our research, when the displayed paid ads set becomes larger, there is a manifestation of increase in click-through rate on top-ranked brand ads while consumer's average click-through rate on individual paid ads decreases. Therefore, from practical point of view, it might be beneficial (or at least not detrimental) for top-ranked brand advertisers to remain in larger paid ads set, given the higher brand recognition that top-ranked brands could receive under this circumstance. Moreover, our empirical results show that the rate of decrease in click-through rate generated by paid ad ranking positions is attenuated as the size of displayed ads set grows larger. Thus, without losing too much consumer click-through, top-ranked brand ads might even be willing to place ads at comparably lower positions (e.g 3 or 4), rather than the topmost location (e.g. rank 1), which is consistent with the conclusion drawn from the research conducted by Jerath et al. (2011).

On the other side, from regular (or non top-ranked) brand advertiser's standpoint, it would be better off if they could land their sponsored ads in relatively smaller displayed paid ads set. Even though it is up to search engines to decide the actual number of ads to be shown in the search results, advertisers can still achieve the purpose of displaying in smaller ads set by choosing more "specific" keywords (namely "long tail" keyword[11]) when they tailor keywords to carter to their targeted customers. For keywords to be more "specific", as suggested in the literature (Rutz and Trusov 2011), advertisers could design keywords or even build upon existing popular keywords by including unique product or service features, or specifying geographical or demographical information when they customize keywords. Additionally, major search engines usually provide keyword analytical tools (e.g. Google Keyword Planner) to assist advertisers make modification of their keyword campaigns.

Second, our estimates regarding the impact of the attractiveness of paid ads on consumer click-through rates illustrate that, for a specific paid ad, even under the exact same configurations of keyword settings, bid and ranking performance, the average click-through rate made by consumers may still be significantly different. In fact, our empirical results show that a paid ad will receive lower click-through rate when there are more top-ranked brand sponsored ads displayed in front of it in the search results. Same as when there are many sponsored ads with "accurate match" keywords displayed in front

[11] "long tail" keyword is defined as keywords that are less popular but more specific to targeted consumer segments, which can also lead to higher consumer conversions, according to the explanation provided by Google.

of a certain paid ads. These findings are similar to the results uncovered from behavioral marketing literature (Chernev and Hamilton 2009) in which consumers prefer to choose more attractive alternatives – alternative can be perceived more attractive if the alternative itself is a high quality product or it matches consumer's preference. Therefore, to avoid being "less attractive", one approach that regular online advertisers could take is to increase the bids dramatically in hope that the corresponding ads could rank in more prominent positions (e.g. rank 1 rather than 4) to compensate its lack of high quality or "perfect" match to consumer preference. However, the downside of this method is that ads displaying in a top position do not necessarily lead to higher consumer conversions (Agarwal et al. 2011; Ghose and Yang 2009). Moreover, given the fact that many regular online advertisers are under daily budget constraints, it might be less cost-efficient by executing such aggressive maneuver (Abrams et al. 2007; Ganchev et al. 2007). Another approach that online advertisers could use is to create their own "niche" at their keyword designing level. That is, advertisers may adjust keywords or keyword match types in a way that could reduce the chance of displaying below too many top-ranked brand ads or ads with "accurate match" keywords. To do that, advertisers could consider customizing keywords to be more specific or setting more "specific" keyword to be "accurate match", rather than choosing general keywords to be "accurate match". However, the prerequisite of employing this approach is that advertisers should be able to identify their interested customer segments and well understand their targeted consumers' search behavior, given the strict restriction embedded in "accurate keyword match".

5.5 Conclusion

With enormous number of advertisers participating in paid search advertising activities across all types of industries, the sponsored search advertising becomes increasingly competitive. This situation places an imperative requirement on advertisers to wisely and efficiently plan their search engine campaigns in order to connect to genuine interested consumers and attract the targeted consumers to visit advertiser sites – through consumer click-through. In this research, using a unique dataset collected from one of the leading search engines in China, we empirically examining the impact of displayed ads assortment size and ads composition on consumer click-through rate, which shows several descriptive empirical insights to practitioners regarding the way consumers react to individual paid ads – in terms of consumer click-through rate – under various types of displayed assortment size and ads composition situations. To enhance search engine performance, especially to increase (or remain a certain level of) consumer click-through rate, our empirical findings suggest that advertisers should not only focus on their own paid ads attributes (e.g. bids, keyword characters), but also need to pay attention to other competitors' ads characteristics and the potential displayed ads size in which paid ads might be placed.

From methodological perspective, we employ a joint distribution modeling approach to examine the relationships between displayed ads assortment size, ads composition and consumer's click-through rate. In the model, we partition each individual paid ads advertising activity into three-stage decision processes starting from search engine's displayed ads size decision, rank order decision to consumer click-through decision in order to take into account the endogeneity issue generated by displayed ads size and rankings. The purpose of us using this specific modeling approach is to integrate displayed ads size and composition impact in consumer click-through rate estimation process and to show descriptive insights with respect to these relationships, rather than provide advertisers with a mathematical tool to precisely predict paid ads click-through rate, given the fact that most advertisers are unable to discover other competitors' exact movement. However, even though it is difficult for individual advertiser to precisely estimate consumer's click-through rate for each of its paid ads, advertisers can still use this model as a simulation approach to roughly estimate their paid ads performance.

Our model has several limitations, which are also the limitations of this research. First, since we do not have the empirical data regarding the different timings of consumer's sequential clicks within a consumer search, we assume that consumer's each time clicks are independent when estimating consumer's click-through rate. Under this situation, our model does not examine how much a consumer learned from a prior click might affect her post click decision. If researchers can get access to this part of the empirical data, it is worthwhile to empirically investigate the relationships between consumer learning and the corresponding sequential click behaviors. Second, given the fact that our dataset is collected from search engine side, we are unable to track consumers' conversion information after they left the search engine and landed in the advertisers' websites. Therefore, unlike the classic modeling path (Ghose et al. 2014; Ghose and Yang 2009; Rutz and Trusov 2011), we end our joint distribution model at the consumer click-through rate stage, rather than consumer conversion rate stage. It would be interesting to put together two pieces – consumer click-through and conversion – to examine how the displayed ads assortment size might influence consumer's conversion rate. Third, we only examine consumer click-through within the paid search domain, since our dataset does not include the empirical information of consumer behaviors in the organic search results section. Future research could be conducted by combining both sides consumer click behavior and examine how assortment size in the organic section might influence consumer click through on paid ads section.

CHAPTER 6 CONCLUSION AND FUTURE RESEARCH

In the previous chapters (Chapter 4 and 5), we have developed and discussed empirical findings of two studies, both of which have been focused on providing appropriate advertiser's keyword strategies to improve advertiser paid ads' click-through performance. However, as we mentioned at the beginning of this dissertation, search engine advertising (or online search advertising) is at its early age and there are still many interesting phenomenon left unexplored which might generate great insights to facilitate practitioners' decision-making in the real world.

In this chapter, we divide our discussion into two sections. In Section 6.1, we make conclusions of the two studies we described in the previous two chapters. In Section 6.2, we discuss some potential research that can be developed in the future, which is also along the line of the dissertation theme that we keep reinforcing throughout this entire work – modeling and analyzing the interactions among consumers, online advertisers and search engine providers. We categorize the potential studies into four subsections. Each of the sections includes discussion from different aspects. In Section 6.2.1, we discuss the possibility of search engine providing an optimal number of paid ads for fulfillment of its revenue maximization. In Section 6.2.2, we talk about examinations of unstructured consumer search queries to infer consumer search intentions and the following navigation and click behavior. In Section 6.2.3, we discuss the attribution modeling approach to improve advertiser's sponsored advertising budget allocation strategies. Finally, in Section 6.2.4, we briefly go over some of the potential working projects which might need additional data sources to support the analysis.

6.1 Conclusion

In this dissertation, we develop two studies that focus on examining the interactions between consumers and advertisers as it intermediated by search engine providers. We specifically explore two major issues from advertiser's perspective and provide empirical insights to assist advertiser reinforce its search advertising strategies and improve its paid ads campaigns.

In the first study, we specifically address different types of advertisers' keyword branding issue, where we found that service retailers (online travel aggregators) experience superior performance than the service providers (hotels). The superior performance experienced by OTAs is not only a reflection of the more prominent rankings given by the search engine, but also the higher click-through rates generated by consumers, where consumers subjectively differentiate sponsored ads from OTAs to Hotels. Based upon the differences found among different advertisers, we provider several suggestions with regard to advertisers' keyword selection – more specifically, keyword branding strategies. First of all, for both types of advertisers (OTAs and hotels), choosing own brands will increase the chance of consumers' click-through. Second, when considering add other brands information into advertisers' keyword portfolios, OTAs would be better off including other hotel brand names in its keyword sets ("cross-category" selection), whereas avoid adding other OTA brand names in its keywords ("within-category" selection). As for hotels, it would be better off making "within-category" branded keyword selections, rather than "across-category" selections. Doing so could also increase the probability of consumers' click-through. However, the level of increase while choosing other brand names in advertiser's keyword sets will be significantly lower than the increase brought by selecting advertiser's own brand name. Finally, we find that generic keywords (no brand information included) generate the lowest consumer click-through rates for both types of advertisers, which indicates that if an advertiser's goal is to attract consumer visits at its own website through the search engine, it would be better off not devoting too much energy and money resources in the development of generic keywords, which is applicable for both OTAs and Hotels.

In the second study, we take the displayed paid ads set into consideration and examine the impact of ads assortment size on consumer click-through rates and its corresponding influence on advertisers' keywords strategies. The fundamental purpose of conducting this study is to address a popular debate which is about whether or not advertisers should prefer more popular keywords rather than specific keywords. Conventional wisdom used to suggest that advertisers should always think of popular keywords and they would benefit from allocating more money into this type of keywords. The rationale behind this suggestion is that popular keywords will increase the chance of ads displayed in front of consumers, which will leave some sort of impressions in consumers' mind and might eventually lead to a consumer's purchase sometime in the future. However, recent research developed in this area has advocated that more specific keywords should be preferred by advertisers because when consumers search for popular keywords, they tend to click on results shown in the organic section rather than paid ads section, which significantly drove down the click-through rates received in the paid section.

In this study, we investigate from a different aspect – by taking into account the number of ads displayed in the paid search results. We find that when advertisers choose more popular keywords, the chance of the search engine giving out a longer list of displayed ads will be increased. When the displayed ads list becomes longer, consumer's click-through rates on an ads will be significantly decreased. Interestingly however, as the ads list growing larger, consumer click-through rates on well-known brands are significantly increased, indicating that at some point consumers might give up on looking for the option that is "perfect matching" his or her search demand, while turning to choose an alternative with brands they are already familiar with. Thus, comparing what has been suggested from the conventional wisdom, we suggest that whether or not advertisers should choose more popular keywords depends upon the advertiser's brand popularity. For advertisers which are "well-known", they can choose and might benefit from choosing popular keywords. Although there is a much higher chance of their ads being placed in a longer ads list, consumers can easily identify them due to the brand awareness towards "well-known" brands. In fact, we have also found empirical evidence where "well-known" brand advertisers might be willing to stay in the middle of the paid search results as it suggested by Jerath et al. (2011).

As for regular brands, the suggestion we made is different from what we've discussed for "well-known" brands. Given that the click-through rates received by regular brands will be significantly decreased as the size of displayed paid ads grows, it would be better off for regular brands to look for and select more specific keywords while managing its paid ads campaigns. Doing so would increase the chance of positioning ads in a smaller displayed ads set, where the probability of consumer click-through rates might be kept in a certain level. One thing we would like to highlight though is that we make our suggestions solely from helping advertisers get higher consumer click-through rates, if a regular advertiser's goal is to increase the chance of paid ads display, then they should definitely adopt the popular keywords. However, from getting higher consumer click-through rates perspective, it might not be a wise decision for regular brand to choose popular keywords, because doing so would increase the chance of ads appearing in a bigger ads set, which will drive down the regular brand's consumer click-through rates.

6.2 Future Research

In this section, we discuss some of the potential research that can be developed in the future, which is consistent along the line of analyzing interactions among consumers, advertisers and search engine providers. Four subsections are developed accordingly and focusing on different aspects.

6.2.1 Estimating Search Engine's "Optimal Size"

As we described above, we conduct two studies which mainly focus on providing advertisers with keyword strategies to improve their sponsored search advertising performance. Following the similar path, which is specified as "sponsored ads displayed in the search results – consumer click-through", a potential study can be developed to explore whether or not there is an optimal strategy that search engines could carry out, when deciding how many sponsored ads shall be provided in response to consumers' search queries. After all, the missions that search engines should accomplish not only include satisfying consumer's search demand, connecting advertisers to their preferred customer segments, but also include maximizing its own revenue. To that end, researchers can drill down into examinations of two key components: 1) the variation of number of consumer clicks and 2) the level of competition created by advertisers as the number of ads displayed changes, given that search engine's revenue are mostly determined by these two factors. In addition, research focus in this branch can be examined both empirically and analytically.

6.2.2 Inferring Consumer's Intention using Query Information

When characterizing consumer search queries, we only consider "branded" and "generic" and use them in our empirical analysis. However, we have found that, aside from brand information mentioned in consumer search queries, queries generally contain information such as "location-specific" (e.g. hotel near the Newark airport), "price-specific" (e.g. cheap hotel or hotel under $100 per night), "star-specific" (e.g. 5-star hotel), and "amenity-specific" (e.g. hotel with free wifi). Oftentimes, the information described above may appear simultaneously in a single search query. And, based upon how the words or phrases being articulated, different types of combinations included in search queries may reveal different search intentions. For instance, a consumer may search "3-star Hotel near Newark airport with free parking", which includes "location-", "amenity-" and "star-specific" information. Compare with someone whose search query only contains "3-star hotel", the intentions from the two consumers might be completely different.

For the former consumer, s/he might be looking for a hotel with reasonable quality ("3-star") and is nearby the airport because s/he might have to travel from that airport. In the meantime, this customer might also be slightly price sensitive given the "free parking" s/he included in the query. For the latter consumer who searched for "3-star hotel", s/he might be simply gathering some information about 3-star hotel (e.g. finding out what are the hotel brands included under the 3-star category, or the common features provided by this level of hotels). According to Broder (2002)'s classification regarding search intentions, the former consumer might conduct a transactional search, whereas the latter consumer might experience a informational search.

Given the different search intentions, their following browsing and click behavior might be different, which further would affect search engine's revenue as well as advertisers' click-through rates. How can search engines or advertisers compartmentalize consumer unstructured search queries in a way that could effectively infer consumer's search intention and predict their following behavior in the search engine? What kind of screen layout strategies that search engines could develop to maintain high customer engagement (or increase customer "stickiness") within search engines' site by studying their search queries? What kind of keyword strategies that advertisers could apply to connect to customers who are prone to purchase? All of these questions are unsolved and may generate significant contributions into the search engine marketing field.

6.2.3 Developing Attribution Modeling Approach

When estimating click-through rates in the previous two studies, we assume that consumer click decisions are independent. That said, a consumer's prior click decision will not affect his or her posterior click decision. This assumption is made based upon the research questions we were examining and the characteristics of the dataset we employed. However, as we drill down into consumer sequential search issue and begin constructing some of the consumers' search paths, we found that it generally takes several searches for a consumer to complete a "search – purchase – new search" cycle. For instance, we create a sample set by randomly selecting 1,000 consumers (who converted at Day 50 – which is the last day included in our dataset) and construct their entire search paths (starting from Day *M* when they initiated a search to their purchase at Day 50), as illustrated by Figure 6.1, the average number of searches – starting from a consumer initiated a search to convert at advertiser's website – is about 10 searches. And as shown in Figure 6.2, the average amount of time spent by a consumer during the search process is about 15 days.

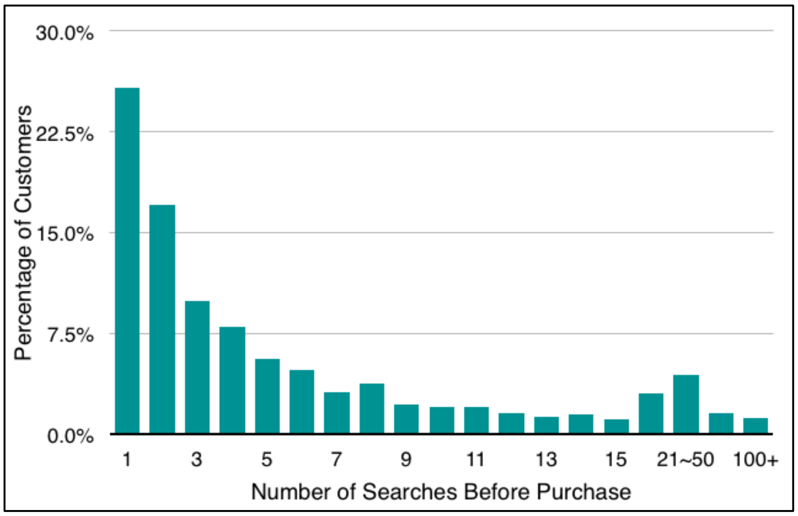

Figure 6.1 Distribution of Number of Searches Generated by Consumers before Purchase

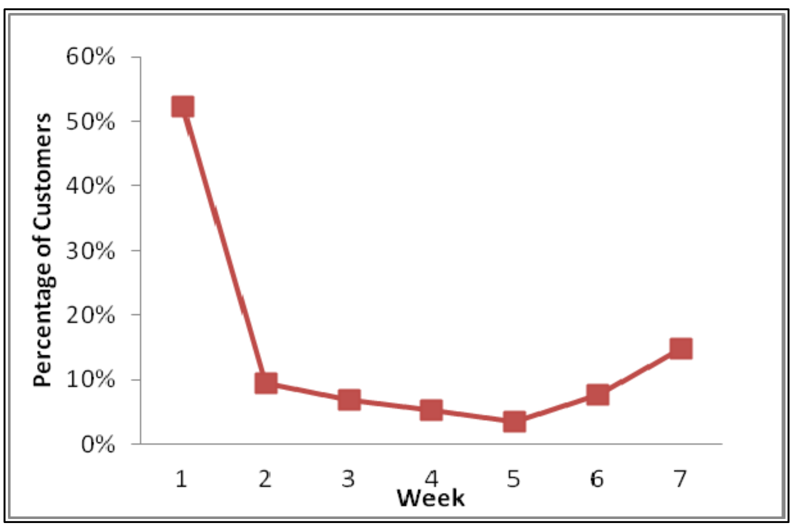

Figure 6.2 Distribution of Time Duration before Purchasing

As consumers progressively approach to the final purchase points, a series of keywords have also been triggered, as can be seen from Figure 6.3. To advertisers, being able to efficiently allocate budget to individual keyword appeared along the search paths becomes an important issue for advertisers to consider. Research focus on solving this specific type of issue can be developed using attribution modeling approach – which contains unique advantages of modeling sequential search behavior than traditional budget allocation tactics.

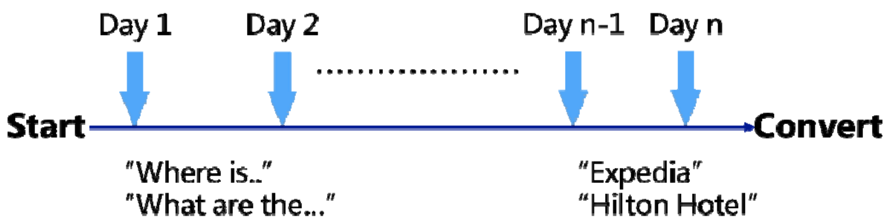

Figure 6.3 Example of Consumer Search Path

6.2.4 Other Potential Projects

In this Section, we briefly go over some of the potential working projects which might need additional data sources to support the analysis.

6.2.4.1 Integrating Advertising with Consumer Web Browsing

In online marketing, a purchase decision made by a consumer is generally the product of several influential factors such as the effectiveness of advertising (e.g. through sponsored search advertising, email or social media), the quality of the advertiser or seller's website (e.g. its readability and convenience), and the appropriateness of the products or services – whether or not the products provided from the website could satisfy consumer's demand. Traditionally, researchers usually follow a path where the discussion of the advertising effectiveness on purchase (Danaher and Dagger 2013) and the discussion of the impact of consumer web browsing behavior on conversion (Sismeiro and Bucklin 2004) are separated.

In the dissertation, we analyze the impact of sponsored search advertising on consumer click-through rates and provide advertisers with tactics that could bring more customer visits to advertisers' websites throughout search engine platforms. Whether or not advertisers could convert customer's visits into a purchase and what the conversion probability might be is out of the scope and the discussion of this dissertation, based upon the fact that our dataset is collected from the search engine, not from the advertiser's end. However, given that data can be collected from a variety of sources, researchers can combine different sources of data whereby figure out ways to integrate advertising and consumer web browsing data when predicting conversions. Research that will be developed along this path might bring significantly improvement in the accuracy of the conversion prediction in the online marketing field.

6.2.4.2 Comparing "Big Data" approach with Traditional Econometrics

In this dissertation, we follow the traditional econometric methodologies (e.g. building joint probability models, using maximum likelihood estimation) while conducting this two essays. However, given the size of the dataset we obtained and given the situation that many corporations or research institutes nowadays are willing to share some of their resources (e.g. large-scale datasets) to researchers, there is a great potential for studies to be deveoped in a way that embraces information system techniques (e.g. text-mining, Big Data analytics) into marketing analysis – especially under the circumstance where so many people are interested in knowing what kind of impacts that the "Big Data" analytics would bring to both academia and industrial areas.

6.2.4.3 Entering into Mobile Marketing Arena

In the dissertation, the empirical analysis and discussion regarding sponsored search advertising are generated based upon the dataset where consumers conducted their searches solely from PC, desktop or laptop – therefore, no mobile devices (e.g. smartphone) or tablets are involved. However, as more and more companies start jumping into the mobile marketing field, striving to grab their "piece" in the marketplace, the demand of developing research in the mobile marketing field will be increasing dramatically.

With regard to the sponsored search advertising's application on mobile devices, there are several directions that future research can head into. First, study can be done to analyze consumer search behavior on mobile devices or to compare and contract consumer's diverse search behavior through different platforms (e.g. mobile versus desktop). Second, research can be done to figure out what types of advertisers might benefit from using mobile advertising format (e.g. mom and pop businesses or national retailers). Third, research can also make investigations with respect to the integration of traditional retailing (e.g. department stores) and mobile applications.

REFERENCES

Bibliography

Abhishek, Vibhanshu, Kartik Hosanagar, and Peter S. Fader (2012), "On Aggregation Bias in Sponsored Search Data: Existence and Implications." *EC: Electronic Commerce*, 1.

Abrams, Zoe, Ofer Mendelevitch, and John Tomlin (2007), Optimal delivery of sponsored search advertisements subject to budget constraints, in Proceedings of the 8th ACM conference on Electronic commerce. San Diego, California, USA: ACM.

Agarwal, Hosanagar, and Smith (2011), "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets." *Journal of Marketing Research* 48.6, 1057-73.

Aggarwal, Gagan , Jon Feldman, and S. Muthukrishnan (2006), "Bidding to the Top: Vcg and Equilibria of Position-Based Auctions." *Springer-Verlag Berlin Heidelberg*, 15-28.

Athey, Susan and Glenn Ellison (2008), Position Auctions with Consumer Search, *Working Paper*.

Anderson, Christopher K. (2011), "Search, Otas and Online Booking: An Expanded Analysis of the Billboard Effect." in *Center for Hospitality Research Report*: Cornell University.

Athey, S. and G. Ellison (2011), "Position Auctions with Consumer Search." *Quarterly Journal of Economics* 126.3, 1213-70.

Ariely, Dan (2009), "Predictably Irrational : The Hidden Forces That Shape Our Decisions." *HarperCollins, New York*

Beggs, S., S. Cardell, and J. Hausman (1981), "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics* 17.1, 1-19.

Botti, S. and S. S. Iyengar (2004), The psychological pleasure and pain of choosing: When people prefer choosing at the cost of subsequent outcome satisfaction, *Journal of Personality and Social Psychology*, 87 (3), 312-326.

Broder, Andrei (2002), A taxonomy of web search, *IBM research*, 36.

Brucks, Merrie (1985), "The Effects of Product Class Knowledge on Information Search Behavior." in: *Journal of Consumer Research*, 1.

Chapman, Randall G. and Richard Staelin (1982), Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model, *Journal of Marketing Research*, 19 (3), 288-301.

Chen, Yongmin and Chuan He (2011), "Paid Placement: Advertising and Search on the Internet." *Economic Journal* 121.556, F309-F28.

Chernev, Alexander and Ryan Hamilton (2009), Assortment Size and Option Attractiveness in Consumer Choice Among Retailers, *Journal of Marketing Research*, 46 (3), 410-420.

Chernev, Alexander, David Glen Mick, and Michael D. Johnson (2003), When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice, *Journal of Consumer Research*, 30 (2), 170-183.

Danaher, Peter J. and Tracey S. Dagger (2013), "Comparing the Relative Effectiveness of Advertising Channels: A Case Study of a Multimedia Blitz Campaign." Journal of Marketing Research 50.4, 517-34.

Dhar, V. and A. Ghose (2010), sponsored search and market efficiency, *Information Systems Research*, 21 (4), 760-772.

Desai, P. S., R. Staelin, and W. Shin (2014), "The Company That You Keep: When to Buy a Competitor's Keyword." *Marketing Science* 33.4, 485-508.

Edelman, Benjamin , Michael Ostrovsky, and Michael Schwarz (2007), "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords." *The American Economic Review* 97.1, 242-59.

Feng, Juan, Hemant K. Bhargava, and David M. Pennock (2007), Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms, *INFORMS Journal on Computing*, 19 (1), 137-148.

Ganchev, K., A. Kulesza, J. Tan, R. Gabbard, Q. Liu, and M. Kearns (2007), Empirical price modeling for sponsored search, *Working Paper*.

Ghose, A., P. G. Ipeirotis, and B. Li (2012), "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content." *Marketing Science* 31.3, 493-520.

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li (2014), "Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue." *Management Science* 60.7, 1632-54.

Ghose, Anindya and Sha Yang (2009), "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets." *Management Science* 55.No.10, 1605-22.

Goodman, Andrew (2006), "Bid Fight." in *Target Marketing*, [available at online: http://www.targetmarketingmag.com/article/understand-search-environment-then-plan-your-keyword-strategy-optimize-your-search-budget-33406/1].

Gourville, John T. and Dilip Soman (2005), Overchoice and Assortment Type: When and Why Variety Backfires, *Marketing Science*, 24 (3), 382-395.
Greenspan, R. (2004), Searching for balance, [available at online: /http://www.clickz.com/stats/sectors/software/article.php/3348071S].

Gretzel, Ulrike and KyungHyan Yoo (2008), "Use and Impact of Online Travel Reviews." in *Information and Communication Technologies in Tourism 2008*, Peter O'Connor, Wolfram Höpken and Ulrike Gretzel eds.: Springer Vienna, 35-46.

Huffman, C. and B. E. Kahn (1998), Variety for sale: Mass customization or mass confusion?, *Journal of Retailing*, 74 (4), 491-513.

Hung, K. H. and S. Y. Li (2007), "The Influence of Ewom on Virtual Consumer Communities: Social Capital, Consumer Learning, and Behavioral Outcomes." *Journal of Advertising Research* 47.4.

Jerath, Ma, and Park (2014), "Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity." *Journal of Marketing Research* 51.4, 480-86.

Jerath, K., L. Ma, Y. H. Park, and K. Srinivasan (2011), "A "Position Paradox" in Sponsored Search Auctions." *Marketing Science* 30.4, 612-27.

Johnson, Eric J. and J. Edward Russo (1984), "Product Familiarity and Learning New Information." *Journal of Consumer Research* 11.1, 542-50.

Katona, Zsolt and Miklos Sarvary (2010), "The Race for Sponsored Links: Bidding Patterns for Search Advertising." *Marketing Science* 29.2, 199-215.

Lancaster, Kelvin (1990), The Economics of Product Variety: A Survey, Institute of Management Sciences.

Liaw, Shu-Sheng and Hsiu-Mei Huang (2003), "An Investigation of User Attitudes toward Search Engines as an Information Retrieval Tool." *Computers in Human Behavior* 19.6, 751-65.

Lorigo, Lori, Maya Haridasan, Hrönn Brynjarsdóttir, Ling Xia, Thorsten Joachims, Geri Gay, Laura Granka, Fabio Pellacini, and Bing Pan (2008), "Eye Tracking and Online Search: Lessons Learned and Challenges Ahead." *Journal of the American Society for Information Science and Technology* 59.7, 1041-52.

McCarthy, Daniel (2010), Social media marketing drives search traffic 61% for one small business: a case study, [available at http://www.themediatransformation.com/2010/06/07/social-media-marketing-drives-search-traffic-61-for-one-small-business-a-case-study/].

Muramatsu, Jack and Wanda Pratt (2001), Transparent Queries: investigation users' mental models of search engines, in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. New Orleans, Louisiana, USA: ACM.

Raman, Niranjan V. (1997), "A Qualitative Investigation of Web-Browsing Behavior." *Advances in Consumer Research* 24.1, 511-16.

Ratchford, Brian T., Debabrata Talukdar, and Myung-Soo Lee (2001), "A Model of Consumer Choice of the Internet as an Information Source." *International Journal of Electronic Commerce* 5.3, 7-21.

Rutz, Oliver J. and Randolph E. Bucklin (2011), "From Generic to Branded: A Model of Spillover in Paid Search Advertising." *Journal of Marketing Research* 48.1, 87-102.

Rutz, Oliver J., Randolph E. Bucklin, and Garrett P. Sonnier (2012), "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising." *Journal of Marketing Research* 49.3, 306-19.

Rutz, Oliver J. and Michael Trusov (2011), "Zooming in on Paid Search Ads—a Consumer-Level Model Calibrated on Aggregated Data." *Marketing Science* 30.5, 789-800.

Schwartz, Barry (2004), the paradox of choice why more is less. New York: Harper Perennial.

Sismeiro, Catarina and Randolph E. Bucklin (2004), "Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach." *Journal of Marketing Research* 41.3, 306-23.

Song, Yao and Carl F. Mela (2011), "A Dynamic Model of Sponsored Search Advertising." *Marketing Science* 30.3, 447-68.

Swait, Joffre and Jordan Louviere (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *Journal of Marketing Research* 30.3, 305-14.

Varian, Hal R. (2007), "Position Auctions." *International Journal of Industrial Organization* 25, 1163-78.

Vuylsteke, A., B. Baesens, J. Poelmans, and Z. Wen (2010), "Consumers' Search for

Information on the Internet: How and Why China Differs from Western Europe." *Journal of Interactive Marketing* 24.4, 309-31.

Weber, T. A. and Z. Zheng (2007), "A Model of Search Intermediaries and Paid Referrals." *Information Systems Research* 18.4, 414-36.

Wilbur, Kenneth C. and Zhu Yi (2009), Click Fraud, *Marketing Science*, 28 (2), 293-308.

Winkelmann, Rainer (2008), Econometric Analysis of Count Data (5 ed.): Springer-Verlag Berlin Heidelberg.

Xiang, Zheng, Karl Wöber, and Daniel R. Fesenmaier (2008), "Representation of the Online Tourism Domain in Search Engines." *Journal of Travel Research* 47.2, 137-50.

Xu, Lizhen, Jianqing Chen, and Andrew Whinston (2011), "Price Competition and Endogenous Valuation in Search Advertising." *Journal of Marketing Research* 48.3, 566-86.

Yalcin, Taylan and Elie Ofek (2011), Competing for Consumers Online: The Advertising Strategies of Vertically Di¤erentiated Firms, *Working Paper*

Yang, Sha and Anindya Ghose (2010), Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?, *Marketing Science*, 29 (No.4), 602-623.

Appendices

Appendix A: The Derivation of the Functions in Essay One

a) The derivation of the ordered logit model (Rank Model)

$$P(R_{ij}=1)=P(R_{ij}^{*}\geq\mu_1)=P(V_{ij}^{R}+\varepsilon_{ij}^{R}\geq\mu_1)=P(\varepsilon_{ij}^{R}\geq\mu_1-V_{ij}^{R})=1-\Lambda(\mu_1-V_{ij}^{R})$$

$$P(R_{ij}=2)=P(\mu_2<R_{ij}^{*}\leq\mu_1)=P(\mu_2-V_{ij}^{R}<\varepsilon_{ij}^{R}\leq\mu_1-V_{ij}^{R})=\Lambda(\mu_1-V_{ij}^{R})-\Lambda(\mu_2-V_{ij}^{R})$$

$$\vdots$$

$$P(R_{ij}=k)=P(\mu_k<R_{ij}^{*}\leq\mu_{k-1})=P(\mu_k-V_{ij}^{R}<\varepsilon_{ij}^{R}\leq\mu_{k-1}-V_{ij}^{R})=\Lambda(\mu_{k-1}-V_{ij}^{R})-\Lambda(\mu_k-V_{ij}^{R})$$

$$P(R_{ij}=10)=P(R_{ij}^{*}\leq\mu_9)=P(\varepsilon_{ij}^{R}\leq\mu_9-V_{ij}^{R})=\Lambda(\mu_9-V_{ij}^{R})$$

where: $\Lambda(\mu_k-V_{ij}^{R})=\dfrac{\exp(\mu_k-V_{ij}^{R})}{(1+\exp(\mu_k-V_{ij}^{R}))}$

Let $l=V_{ij}^{R}-\mu_k \Rightarrow \Lambda(\mu_k-V_{ij}^{R})=\dfrac{e^{-l}}{1+e^{-l}}=\dfrac{\frac{1}{e^{l}}}{\frac{e^{l}+1}{e^{l}}}=\dfrac{1}{1+e^{l}}=\dfrac{1}{1+\exp(V_{ij}^{R}-\mu_k)}$

$$\therefore \Lambda(\mu_k-V_{ij}^{R})=\frac{1}{1+\exp(V_{ij}^{R}-\mu_k)}$$

$$\therefore P(R_{ij}=k)=\Lambda(\mu_{k-1}-V_{ij}^{R})-\Lambda(\mu_k-V_{ij}^{R})=\frac{1}{1+\exp(V_{ij}^{R}-\mu_{k-1})}-\frac{1}{1+\exp(V_{ij}^{R}-\mu_k)}$$

b) The derivation of the log likelihood function

$$LH=\prod_{j}\prod_{i}P(Y_{ij}=1\,|\,R_{ij}=k)^{Y_{ij}}\,P(Y_{ij}=0\,|\,R_{ij}=k)^{1-Y_{ij}}\cdot P(R_{ij}=k)$$

$$\log(LH)=\sum_{j}\sum_{i}\log[P(Y_{ij}=1\,|\,R_{ij}=k)^{Y_{ij}}\cdot P(Y_{ij}=0\,|\,R_{ij}=k)^{1-Y_{ij}}\cdot P(R_{ij}=k)]$$

$$=\sum_{j}\sum_{i}\left\{\log[P(Y_{ij}=1\,|\,R_{ij}=k)^{Y_{ij}}]+\log[P(Y_{ij}=0\,|\,R_{ij}=k)^{1-Y_{ij}}]+\log[P(R_{ij}=k)]\right\}$$

$$=\sum_{j}\sum_{i}\left\{Y_{ij}\log[P(Y_{ij}=1\,|\,R_{ij}=k)]+(1-Y_{ij})\log[P(Y_{ij}=0\,|\,R_{ij}=k)]+\log[P(R_{ij}=k)]\right\}$$

Appendix B: Wald test for comparison between variables in pairs

The detailed procedure of Wald test is summarized as below:

Step 1: Setting hypothesis

H_o (Null): $h(\beta) = \beta_1 - \beta_2 \leq 0$

In our case, it can be written as $Accurate\_Keyword_{ij} - Phrase\_Keyword_{ij} \leq 0$

H_a (Alternative): $h(\beta) = \beta_1 > \beta_2$, thus, $Accurate\_Keyword_{ij} > Phrase\_Keyword_{ij}$

Step 2: Calculating W, where $\widehat{\beta_1}$ and $\widehat{\beta_2}$ are coefficients generated from MLE:

$$W = \left[\widehat{\beta_2}(\widehat{\beta_1} - \widehat{\beta_2}) \right]^2 \times \left[\widehat{\beta_2}^2 Var(\widehat{\beta_1}) - 2\widehat{\beta_1}\widehat{\beta_2} Cov(\widehat{\beta_1}, \widehat{\beta_2}) + \widehat{\beta_1}^2 Var(\widehat{\beta_2}) \right]^{-1}$$

Since W is asymptotically $\chi^2(1)$ distributed, therefore, $\sqrt{W}$ is asymptotically standard normal distributed.

Step 3: Calculating $\sqrt{W}$ and comparing it with Z_α

If $\sqrt{W} > Z_\alpha$, we can reject the null hypothesis. Otherwise, we cannot reject the null hypothesis.

Appendix C: Swait & Louviere Test for Comparison between Different Advertisers

The advantage of using Swait & Louviere test is that the methodology can overcome the shortcomings of the Chow test. It also can be applied in the non-linear world (e.g. logit model).

The detail steps of conducting this test is listed below:

First, we need to create two subsets (one subset is all OTA ads impressions, another set is all Hotel ads impressions), and run the Click Model for both OTA set and Hotel set.

Let β_1 represents a set of coefficients generated from OTA set and L_1 represents the log-likelihood value.

Let β_2 represents a set of coefficients generated from Hotel set and L_2 represents the log-likelihood value.

Next, we need to conduct two sets of testing:

1) $\beta_1 = \beta_2 = \beta$ and;

2) $\mu_1 = \mu_2 = \mu$,where μ_1, μ_2 are the scale parameters in the logit model

We start with the test of $\beta_1 = \beta_2 = \beta$, to do that, we plot β_1 & β_2, and fit a line $\beta_1 = \overline{\mu_2}\beta_2$, where $\overline{\mu_2}$ is the slope of the line.

Then, we stack the data from two subsets which can be expressed as:

$$W = \begin{bmatrix} X_1 \\ \overline{\mu_2} X_2 \end{bmatrix}$$

Using MLE, we can estimate a set of coefficients, denoted as β, and we also get a log-likelihood value L_μ

Using a formula $\lambda_A = -2\left[L_\mu - (L_1 + L_2)\right] \sim \chi^2(K+1)$ where K is the number of parameter estimates in each β, we can either reject or not reject the hypothesis of $\beta_1 = \beta_2 = \beta$. If we reject the null hypothesis, then the test is completed and we stop it at this stage. If not, we move into the second test, where we test the following hypothesis: $\mu_1 = \mu_2 = \mu$

To do that, we assume that $\mu_1 = \mu_2 = \mu$ and pool the two subsets together where

$$W = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Using maximum likelihood estimation, we can get another set of estimates, denote as $\beta^{'}$ and the log-likelihood value generated from the MLE, denote as L_p.

Using a similar formula as we did previously: $\lambda_B = -2\left[L_p - L_\mu\right] \sim \chi^2(1)$

Compare the value of λ_B and $\chi^2(1)$, we decide whether or not to reject the null hypothesis $\mu_1 = \mu_2 = \mu$.

Appendix D: The derivation of functions in Essay Two

1) The derivation of the reduced form of Rank Ordered Logit Model

The derivation of the reduced form of the rank ordered logit model

First, the function is specified in the random utility model format, where: $U_{ij} = V_{ij} + \varepsilon_{ij}$

V_{ij} is the deterministic component, and ε_{ij} follow some distribution function

Therefore, for an observed ranking order, the general function can be specified as:

$$P(U_{i1} > U_{i2} > \cdots > U_{iJ}) = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{i1}} \int_{-\infty}^{U_{i2}} \cdots \int_{\infty}^{U_{i,J-1}} dG(U_{i1}, U_{i2}, \cdots, U_{iJ})$$

Assuming that ε_{ij} follows the extreme value distribution, thus:

$P(\varepsilon_{ij} < t) = F(t) = e^{-e^{-t}}$, which forms the basis of logit model specification.

Meanwhile, since ε_{ij} follows the identical and independent distributed (IID) assumption, the function $P(U_{ij} > U_{ik}, j \neq k)$ can be expressed as:

$P(U_{ij} > U_{ik}, j \neq k) = \dfrac{e^{V_{ij}}}{e^{V_{ij}} + e^{V_{ik}}}$, which is the basic logit model specification.

To extend the logit model into rank ordered logit model, we need to use the independent property of the conditional distribution. It means that once the most favorite option (alternative) is chosen, it will be remove from the list and is independent from the rest ordering process. This is equivalent to the IIA property of the logit specification. Therefore, the original function can be written as:

$$P(U_{i1}>U_{i2}>\cdots>U_{iJ})=\int_{-\infty}^{\infty}\int_{-\infty}^{U_{i1}}\int_{-\infty}^{U_{i2}}\cdots\int_{\infty}^{U_{i,J-1}} dG(U_{i1},U_{i2},\cdots,U_{iJ})$$

$$=P(U_{i1}>U_{i2},\cdots,U_{iJ})\cdot P(U_{i2}>U_{i3},\cdots,U_{iJ})\cdots\cdot P(U_{i,J-1}>U_{iJ})$$

$$=P(U_{i1}>U_{ij},\text{for } j=2,J)\cdot P(U_{i2}>U_{ij},\text{for } j=3,J)\cdots P(U_{i,J-1}>U_{iJ})$$

$$=\frac{e^{V_{i1}}}{e^{V_{i1}}+\cdots+e^{V_{iJ}}}\cdot\frac{e^{V_{i2}}}{e^{V_{i2}}+\cdots+e^{V_{iJ}}}\cdots\frac{e^{V_{i,J-1}}}{e^{V_{i,J-1}}+e^{V_{iJ}}}\cdot\frac{e^{V_{iJ}}}{e^{V_{iJ}}}=\prod_{j=1}^{J}\left[\frac{\exp(V_{ij})}{\sum_{k=1}^{J}\delta_{ijk}\exp(V_{ij})}\right]$$

2) The derivation of the log-likelihood function of the three-stage joint probability model

$$L=\prod_{i=1}^{n}\prod_{j=1}^{m}P(Y_{ij}=1,C_i,S_i)^{Y_{ij}}\,P(Y_{ij}=0,C_i,S_i)^{1-Y_{ij}}$$

$$=\prod_{i=1}^{n}\left\{\begin{array}{l}\left\{\left[\prod_{j=1}^{m}[P(Y_{ij}=1\,|\,C_i,S_i)\cdot P(C_i\,|\,S_i)]\right]\cdot P(S_i=m)\right\}^{Y_{ij}}\\ \times\left\{\left[\prod_{j=1}^{m}[P(Y_{ij}=0\,|\,C_i,S_i)\cdot P(C_i\,|\,S_i)]\right]\cdot P(S_i=m)\right\}^{1-Y_{ij}}\end{array}\right\}$$

$$=\prod_{i=1}^{n}\left\{\begin{array}{l}\left\{\left\{\prod_{j=1}^{m}\frac{e^{V_{ij}}}{1+e^{V_{ij}}}\cdot\left[\frac{\exp(V_{ij})}{\sum_{k=1}^{m}\delta_{ijk}\exp(V_{ik})}\right]\right\}\cdot P(S_i=m)\right\}^{Y_{ij}}\\ \times\left\{\left\{\prod_{j=1}^{m}\frac{1}{1+e^{V_{ij}}}\cdot\left[\frac{\exp(V_{ij})}{\sum_{k=1}^{m}\delta_{ijk}\exp(V_{ik})}\right]\right\}\cdot P(S_i=m)\right\}^{1-Y_{ij}}\end{array}\right\}$$

$$L=\prod_{i=1}^{n}\left\{\left\{\prod_{j=1}^{m}\frac{e^{V_{ij}}}{1+e^{V_{ij}}}\cdot\left[\frac{\exp(V_{ij})}{\sum_{k=1}^{m}\delta_{ijk}\exp(V_{ik})}\right]\right\}^{Y_{ij}}\left\{\prod_{j=1}^{m}\frac{1}{1+e^{V_{ij}}}\cdot\left[\frac{\exp(V_{ij})}{\sum_{k=1}^{m}\delta_{ijk}\exp(V_{ik})}\right]\right\}^{1-Y_{ij}}\cdot P(S_i=m)\right\}$$

$$\Rightarrow Log-Likelihood:$$

$$\ln L=\sum_{i=1}^{n}\left(\begin{array}{l}\sum_{j=1}^{m}\left(Y_{ij}\cdot\log(\frac{e^{V_{ij}}}{1+e^{V_{ij}}})+(1-Y_{ij})\cdot\log(\frac{1}{1+e^{V_{ij}}})\right)\\ +\sum_{j=1}^{m}\left(V_{ij}-\log(\sum_{k=1}^{m}\delta_{ijk}\exp(V_{ik})\right)+\log\left(P(S_i=m)\right)\end{array}\right)$$