Testing the Signaling Theory of Advertising: Evidence from Search Advertisements

Taylor Bentley, Tat Y. Chan and Young-Hoon Park†

January 2015

†Taylor Bentley is a PhD Candidate in Marketing at the Olin School of Business, Washington University, St. Louis, MO 63130; phone: (314) 488-9521; fax: (314) 935-6359; email: bentleyt@wustl.edu. Tat Y. Chan is Associate Professor of Marketing at the Olin School of Business, Washington University, St. Louis, MO 63130; phone: (314) 935-6096; fax: (314) 935-6359; email: chan@wustl.edu. Young-Hoon Park is Sung-Whan Suh Professor of Management and Associate Professor of Marketing at the Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853; phone: (607) 255-3217; fax: (607) 254-4590; email: yp34@cornell.edu. The authors would like to thank the company, which wishes to remain anonymous, that provided the data used in this study.
Testing the Signaling Theory of Advertising: Evidence from Search Advertisements

Abstract

Using a dataset of travel-related keywords obtained from a search engine, we test to what extent consumers are searching and advertisers are bidding in accordance with the signaling theory of advertising in literature. We find significant evidence that consumers are more likely to click on advertisers at higher positions because they infer that such advertisers are more likely to match with their needs. Further, consumers are more likely to find a match with advertisers who have paid more for higher positions. We also find strong evidence that advertisers increase their bids when there is an improvement in the likelihood that their offerings match with consumers’ needs, and the improvement cannot be readily observed by consumers prior to searching advertisers’ websites. These results are consistent with the predictions from the signaling theory. We test several alternative explanations and show that they cannot fully explain the results. Furthermore, through an extension we find that advertisers can generate more clicks when competing against advertisers with higher match value. We offer an explanation for this finding based on the signaling theory.

Keywords: Signaling, Advertising, Search advertisements, Consumer search, Position competition
1. Introduction

Since Nelson (1970, 1974), a stream of theoretical works (e.g., Klein and Leffler 1981, Grossman and Shapiro 1984, Milgrom and Roberts 1986, Bagwell and Riordan 1991, Meurer and Stahl 1994, Anderson and Renault 2006) have been developed to formalize the signaling theory of advertising. However, empirical evidence supporting the signaling function of advertising is rather limited. In this paper, we empirically test the theory using a rich micro-level dataset of search advertising obtained from a search engine. Although alternative explanations for why advertising helps increase sales, including the persuasive and informative functions of advertising, are available in the literature, the signaling theory provides unique insights into how both consumers and advertisers strategically behave in markets with differentiated products. In such markets, consumers have uncertainty regarding the attributes of product offerings, and it is costly for them to search for product information. The theory predicts that, under general conditions, firms with products that can better match with consumer needs and preferences will spend more on advertising to signal the high “match value” of their products. This is because advertisers expect that consumers, after being exposed to the advertising signal, will infer that their products are more likely to meet their needs. At equilibrium, consumers make such rational inferences based on the signals, leading to the positive correlation between advertisement spending and sales or consumer quality perception. Since the theory has major implications for firm competition and welfare analysis, testing the validity of the theory with real market data is important to firms, policy makers, and academic researchers.

We choose the search advertising market for this study for several reasons. First, search advertising has played a dominant role in the advertising industry. According to a report from the Interactive Advertising Bureau (http://www.iab.net/AdRevenueReport), internet advertising revenues have soared to $42.8 billion in the U.S. in 2013, surpassing broadcast television advertising revenues for the first time. Search advertising is the leading internet advertising
format. It has a 43% share of internet advertising revenues, reaching $18 billion, much higher than all other formats including display advertising ($13 billion) and mobile advertising ($7 billion).

Understanding the underlying mechanisms that drive the effectiveness of search advertising thus is substantively important. Second, the nature of search advertising offers an ideal environment for testing the signaling theory. In many cases, using keyword search implies that consumers have uncertainty for keyword-related products or services, a necessary condition for signaling theory to work. Another key condition is that consumers have knowledge on which firms spend more on advertising. Since consumers observe the ad positions of sponsored links on the search results page, it is reasonable to assume that they can infer advertisers at higher positions pay more for their clicks, via keyword auctions, than advertisers at lower positions. In contrast, this condition may not be satisfied in other markets such as print and TV advertising. Finally, although past research has tested the relationship between demand and advertising spending and, in the case of search advertising, the impact of ad positions on consumer clicks and purchases, there are no empirical studies testing whether firm strategies are consistent with the predictions from the signaling theory. The rich micro-level data that we use in the study enables us to test the theory not only on consumer but also on firm behaviors. Alternative explanations for why advertising works have to be examined. As we will discuss in the section of literature review, testing firm behaviors and ruling out alternative explanations are a challenging task in previous empirical studies because they use aggregate data sourced from a single advertising firm which lack data on other advertisers.

In this study, we choose several travel-related keywords that have different properties associated with consumer needs and preferences. The data provides detailed information on both consumers and advertisers, including the identity of every advertiser whose sponsored link is visible to consumers, their ad positions on the search results page, and the entire sequence of clicks for every keyword search made by consumers. We also observe for each advertiser how
much it pays as well as how much it bids for a click on its sponsored link. Therefore, we are able to test not only the demand side but also the supply side behaviors predicted by the theory. A typical advertiser (e.g., travel agency) in data carries a full line of differentiated travel services (e.g., packaged tours, flights, hotels, car rentals). For many service attributes, it is impossible for consumers to know all details by just browsing the search results page. Searching for the details at each advertiser’s website, however, is very time-consuming. Consumers therefore may rely on the signals they observe from search advertisements to determine their clicking and purchasing decisions.

We test three main hypotheses that are developed from the signaling theory of advertising. Our empirical tests show that, on the consumer side, sponsored links at high ad positions attract more clicks even after controlling for advertiser fixed effects. Furthermore, consumers are more likely to choose high-positioned sponsored links for terminal clicks, the last link clicked in the search process. We argue that terminal clicks and the probability that consumers find a match should be positively correlated. This result therefore is consistent with the prediction that the offerings from advertisers at higher (lower) positions match with the needs and preferences for a larger (smaller) proportion of consumers. We also test the advertiser behaviors as predicted by the theory using a unique feature in the data. Advertisers frequently adjust their bids for a click within a day but, because of the uncertainty regarding how much their competitors will bid, it is possible that their ad positions remain unchanged. Empirical tests using these observations show that an advertiser’s decision to increase its bids is positively correlated with terminal clicks. Yet, without changing ad positions, there is no significant correlation between bid amounts and clicks. This provides significant evidence that advertisers’ bid decisions are based on the match value that has to be signaled, and not solely the information that consumers already know prior to clicking sponsored links.
We further examine whether alternative explanations for how advertising works can account for these results. Making use of the individual-level consumer search data, we test how consumer clicks are correlated with changes in ad positions within a day and find similar results, implying that our findings are unlikely to be driven by external factors (e.g., advertising campaigns in other media) that may increase the click potential which we as researchers do not observe. We also show that the persuasive and informative functions of advertising, as well as the consumer non-strategic top-down search behavior, cannot fully explain our findings. In sum, these results provide thorough support that signaling is at work in our empirical context. Finally, we further illustrate from data an important phenomenon of informational externality, where a sponsored link with higher match value has a positive spillover effect on other sponsored links listed above or below. The signaling theory offers an explanation for this finding.

1.1 Related Literature

Nelson (1970, 1974) argues that firms in an experience goods market have incentive to advertise heavily to signal the quality of their products. Later works (e.g., Grossman and Shapiro 1984, Milgrom and Roberts 1986, Meurer and Stahl 1994, Anderson and Renault 2006) formalize the idea by developing a game-theoretic model in which high-quality firms are incentivized to spend more on advertising to signal but low-quality firms do not find it profitable to mimic. Klein and Leffler (1981) endogenize product quality in the signaling model. Bagwell and Riordan (1991) develop a model for durable goods, where price instead of advertising is used to signal product quality, and show that at equilibrium high-quality firms will charge more than the full information price.

A stream of empirical research follows the theoretical development by using aggregate data to test how the signaling theory applies to advertising. Tellis and Fornell (1988), for example, use PIMS datasets to examine how product quality (represented by the difference between the percentage of sales of products superior to competing products and the percentage of sales
inferior) is affected by advertising. Caves and Greene (1996) use *Consumer Reports* to examine the relationship between quality ratings and advertising outlays of product brands. Thomas et al. (1998) use data from the U.S. automobile industry and find a positive association between future sales and current advertising levels. The challenge with using aggregate data is that it is difficult to rule out alternative explanations. This is because the relationship between advertising and sales or quality rankings can be due to other unobserved factors (e.g., changed product quality) or other functions of advertising unrelated to quality signaling. Because of this issue, another stream of research uses experimental data to investigate how participants’ perceived product quality changes under manipulated conditions of advertising spending (e.g., Kirmani and Wright 1989, Kirmani 1990, Moorthy and Hawkins 2005).

Our research differs from the existing studies in a few important ways. Our study uses rich micro-level data that help test the theory against other alternative explanations, instead of aggregate-level data. It also differs from the experimental literature by using observational data which is important for the external validity of the study. Most importantly, because we obtain data on not only how consumers search but also how advertisers bid, we are able to study the strategic behaviors on the firm side, which has never been tested in the previous literature, simultaneously with the consumer behavior predicted by the theory.

In the search advertising context, a number of empirical studies have documented the effects of ad positions on clicks and purchase conversions (e.g., Ghose and Yang 2009, Yang and Ghose 2010, Agarwal et al. 2011, Goldfarb and Tucker 2011, Rutz and Bucklin 2011, Yao and Mela 2011, Chan and Park 2014). The common result is that the number of clicks increases as advertiser position moves up in the sponsored list. Recently, Narayanan and Kalyanam (2014) find a few moderators of the positions effects, and show that the effects are weaker for smaller firms and for keywords with less prior consumer experience. These empirical findings have served as the basis for a large stream of research in search advertising (and also served as a guideline for
practitioners), but limited research is available to examine the underlying mechanisms of the position effects in search advertising. The exceptions are the theoretical works of Athey and Ellison (2011) and Chen and He (2011), who apply the signaling theory to explain the positive position effects. They study the optimal bidding strategy of advertisers with heterogeneous products or services competing against each other for search ad positions. Their analyses show that, for consumers with heterogeneous needs and preferences who are searching for product information, when they can rationally infer the equilibrium auction outcomes, advertisers with high match value (i.e., their product offering can match with the needs and preferences of a large proportion of consumers) will bid more for high ad positions, since more relevant advertisers will benefit more from attracting consumers to their website. We therefore observe a separating equilibrium.¹ The signaling approach we use in this study is based on their models.²

To the best of our knowledge, this is the first paper that provides empirical evidence that firms use ad positions in search advertising to signal advertiser match value, which will influence consumers’ search behaviors. As the signaling theory has major implications for firm competition

¹ Jerath et al. (2011), however, show that it may also be optimal for a low-quality advertiser to outbid high-quality competitors. This “Position Paradox” occurs when some portion of the consumers are knowledgeable about the firms’ offerings and high-quality advertisers drop their bids to save cost, they do not lose too many clicks to inferior advertisers. It may be optimal for low-quality advertisers to maintain a high bid to attract potential consumers before they browse and click links from high-quality advertisers.

² Search advertising has attracted much research interest in the economics and marketing literature. On the firm side, researchers have focused on the bidding strategies and advertiser competition in position auctions for keywords. In generalized second price auctions, equilibrium bidding strategies of which bidders bid optimally by their value per click have been studied (e.g., Edelman et al. 2007, Varian 2007). Chan and Park (2014) model the decisions of advertisers and propose using the method of moment inequalities when optimal conditions are not well defined. Yao and Mela (2011) structurally model advertisers’ dynamic bidding decisions assuming asymmetric valuations of sponsored positions for advertisers. Edelman and Ostrovsky (2007) estimate advertiser valuations without uncertainty in the search environment, and Athey and Nekipelov (2012) develop a homotopy-based method for computing equilibria when there is uncertainty about the set of competitors and individual users, and competing bids. Our research is also closely related to how consumers search for product information in the optimal sequence. Past literature has studied either non-sequential search (e.g., Stigler 1961) or sequential search (e.g., McCall 1970, Weitzman 1979). Kim et al. (2009) apply the Weitzman framework to model the optimal dynamic search process of consumers when shopping for camcorders at Amazon.com. De los Santos et al. (2012) use data on browsing and prices across websites to test different search models and argue that non-sequential search is more consistent with the data than sequential search and that consumers may update their belief of the distribution of prices during the search. Honka and Chintagunta (2014) use data on consumer shopping behavior in the U.S. auto insurance industry to identify the search strategy consumers use and argue that the largest insurance companies are better off when consumers search sequentially, while smaller companies profit from consumers searching simultaneously.
and welfare analysis, our findings are important to firms, policy makers, and academic
researchers.

The remainder of the paper is organized as follows. Section 2 discusses the data. We
outline the signaling model and present the key hypotheses to be tested in section 3. Section 4
presents the results of the hypotheses testing. In section 5, we test against several alternative
explanations and conclude that our findings cannot be fully explained by these explanations.
Section 6 explores an extended test of the data. We conclude with directions for future research in
section 7.

2. Data
The data are from a leading search engine firm in Korea. In response to a keyword search, a list of
sponsored ads is placed at the top of the search results page, with a maximum of five ads displayed.
The search engine decides which sponsored links are displayed in which order, based on a second-
price position auction similar to what Google uses; however, no quality score based on a link’s
click potential is applied. Each advertiser pays each time the link is clicked, this is known as the
cost-per-click (CPC) pricing. A list of organic links is placed below the list of sponsored links.3
Our data provider noted that commercial websites, which sell products related to search queries,
are restricted from the organic listing. As such, there is negligible overlap between the links
displayed in the organic (informational) and sponsored (commercial) listings. To be viewed by the
consumer during the search, advertisers must bid to be placed in the sponsored listing. Organic
links are chosen from a proprietary database consisting of different types of related content, such
as general information pages on the topic (e.g., pages from Wikipedia), news, blogs and cafes (i.e.,

---

3 We note that this layout is similar to the layout of popular US search engines such as Google and Bing, which
display several (typically, up to three) sponsored links on the top of the results page following by the organic link. We
refer readers to Jerath, Ma, and Park (2014) for details of the data.
online communities run by the portal associated with the search engine), and a knowledge database where online users post questions and other users provide answers.

The search engine provided us the data on consumer search and advertiser bidding activities for a set of 8 keywords, all associated with travel, over the period of one month. On average, there are 16,059 search instances per keyword over the data period, with a standard deviation of 10,469. We observe which sponsored links and organic links the individual user clicks and at what time. Thus, the data provides us detailed information on the entire sequence of links clicked by a consumer (including either the sponsored or organic link), and the time of clicks. The information on the order of how consumers click sponsored links is important for this study since it helps test the hypotheses against some of the alternative explanations. Data on the entire sequence of consumer clicks across advertisers provide us a unique opportunity to test signaling effects from search advertising. Past research typically only relied on aggregate data sourced from a single advertiser. Based on the sequence of clicks made by an individual consumer, we can correctly identify which link is the last one that a consumer clicks during the search process; thus, unlike Chan and Park (2014) we do not rely on behavioral assumptions (e.g., top-down browsing and sequential search) to infer terminal clicks.

**Insert Table 1 about here**

We also observe from the data the identity of each advertiser displayed in the sponsored listing in response to a consumer’s search query. On average, as shown in Table 1, there are 15.6 advertisers per keyword competing for ad positions in the sponsored section. This is different from previous works (e.g., Ghose and Yang 2009, Yang and Ghose 2010, Agarwal et al. 2011, Rutz and Bucklin 2011) which often lack data on the complete list of sponsored links presented to consumers, as their data source comes from a single advertiser. In addition, we have data on advertisers’ bid amounts and the prices they pay. An advertiser’s position may change throughout the day as it alters its bid. There is significant variation in ad positions obtained by advertisers
within a single day in data. The third row in Table 1 shows that the average number of ad positions per advertiser per keyword is 2.4 per day, with a standard deviation of 1.2. The table also presents the descriptive statistics for the top three most searched keywords, “Clearance Sale Flight Ticket”, “Travel Agency”, and “Travel to Jeju Island”, all demonstrating similar patterns. The information on the advertiser bidding activity in sponsored search advertising is unique since it helps understand how advertisers indeed behave in sponsored search advertising. We will explain in later sections how we use this daily variation in ad positions to test the effect of ad positions on click behavior. We also observe advertisers’ selling propositions which typically highlight current promotions, special travel packages, a website’s unique attributes, and so on. Most advertisers use the same proposition throughout the data period. The average number of unique selling propositions for each specific advertiser-keyword combination is 1.2. Table 1 shows that selling propositions have never been changed for “Clearance Sale Flight Ticket” and “Travel to Jeju Island” during the data period. The table also provides the information on the average number of advertisers who obtains each position. On average, 9.4 advertisers are placed at the topmost position at some point in the data period, while 15.3 advertisers obtain the fifth position. This indicates significant changes in advertisers’ bidding activity for these keywords in the data period.

It is worth discussing that the travel-related category satisfies several criteria for empirically testing the signaling theory in the search advertising context. First, consumers typically have large uncertainty regarding the travel services that they search. Searches can be costly as it takes time to browse and process all details provided from travel websites. Second, on the advertiser side, the travel service market is highly fragmented, with many competing service providers without anyone dominating. Consumers are not likely to have full information on the service offerings from every website prior to clicking its sponsored link. Advertisers’ offerings, and thus their match values, change frequently as airline seats sell out, hotels fill up, they receive blocks of tickets on promotion, and flight dates approach.
In addition, the travel-related category has a few important data features which help statistical testing. First, we observe a large number of search instances and active bidding from advertisers to obtain ad positions, with all five ad positions being occupied for the majority of the search instances (97.2%). As such, we avoid small sample issues for statistical testing. Second, each ad position has been occupied by many different advertisers (see Table 1). This helps us to separate the advertiser fixed effects from the effects of ad positions. For each advertiser, ad positions also vary across search instances, implying intensive competition among the advertisers. Third, we observe significant variations in CPC across ad positions, ranging from an average of $0.57 for the topmost position to an average of $0.40 for the fifth position (see Table 1).\footnote{We note that all bids are in Korean currency (won), where 1000 won corresponds approximately to $1.} We also observe significant variation in CPC for the same position and the same keyword during the data period. Fourth, for the purpose of testing alternative explanations for our findings, it would be desirable to have keywords that are narrowly defined (e.g., “Flight Ticket to Jeju Island”) and some keywords that are more general (e.g., “Flight Ticket”), and keywords focusing on a specific attribute (e.g., price in “Clearance Sale Flight Ticket”) and some which are more comprehensive (e.g., “Flight Ticket”). The purpose for this criterion will become clear later in the paper. Furthermore, because our research objectives are related to firm sales, we look for keywords that are relatively more likely to indicate the purchase intention from consumers. For example, a keyword search for “Flight Ticket to Jeju Island” is likely to be related to the intention of buying tickets, while a keyword search for “Jeju Island” may only serve for informational purposes.

3. The Signaling Model and Hypotheses

This section outlines the signaling model and describes how it is applied to the search advertising context. We highlight a set of testable predictions on the behaviors of consumers and advertisers, which constitute the signaling theory. Since our modeling approach is similar to Athey and Ellison
(2011) and Chen and He (2011), we refer readers who are interested in how the equilibrium outcomes are derived to their papers for details.

We consider each search instance in the data as a search query made by an individual consumer. Athey and Ellison (2011) assume that each advertiser differs in the match value, i.e., the probability of meeting a consumer’s needs, of its product or service offering, implying that advertisers are horizontally differentiated. By focusing on horizontal, rather than vertical, differentiation, we study the most basic case in search advertising, in which search queries could be sufficiently broad that users may have many different intentions when searching the same keyword. In such context, the key differentiation among advertisers is the likelihood that their offerings fit with heterogeneous consumer needs.

Subsequent to a keyword search, consumer \( i \) observes advertiser \( j \) in the sponsored listing. To test the signaling theory from data, we use an empirical specification to measure the likelihood that the advertiser’s offering matches with the consumer’s needs. We assume that the likelihood is driven by a latent utility that is determined by the advertiser’s offering for consumer, which is specified as follows:

\[
V_{ij} = u_{ij} + v_{ij}.
\]

(1)

The first component \( u_{ij} \) is the portion of the utility which the consumer already has information about prior to clicking the advertiser’s sponsored link. In empirical testing, it is specified as

\[
 u_{ij} = \alpha_j + e_{ij},
\]

(2)

where \( \alpha_j \) is the value averaged across consumers, and \( e_{ij} \) is assumed to be iid across consumers. By definition the average value of \( e_{ij} \) across consumers is zero. The second component, \( v_{ij} \), is the unknown portion of the utility which the consumer can obtain from the advertiser’s website by
If $V_{ij}$ is larger than a threshold value (normalized to zero in empirical testing), the match is one, i.e., the consumer finds a match.

As an illustration, suppose the consumer is planning a vacation to Jeju Island, a popular vacation destination in Korea. She searches the keyword “Flight Ticket to Jeju Island” and finds a list of sponsored advertisers, including travel agents and airline carriers. The selling propositions in the ad, which typically consist of a phrase (such as “We offer flights to Jeju Island from major cities”), may convey information on what the advertiser offers to the consumer. This information, and the consumer’s prior knowledge about the advertiser, gives the consumer $u_{ij}$. Still, the consumer may have specific needs or preferences regarding the departure place, date and time of travel, etc. Detailed information of the advertiser’s offerings, captured by $u_{ij}$, is not available on the search results page. To find out what flights are available, she has to search for details from the advertiser’s website. The higher the value of $u_{ij}$, the higher the likelihood that the consumer finds the advertiser’s offering a good match. In empirical testing, we use a reduced-form way to specify this component as follows:

$$u_{ij} = \xi_j + \varepsilon_{ij}, \quad (3)$$

where $\xi_j$ represents the average match value across consumers, and $\varepsilon_{ij}$ captures the individual-specific stochastic component whose value is zero averaged across consumers. In the above example, $\xi_j$ will be higher if the advertiser has more flights to Jeju Island from different cities, or charges lower prices on average. The advertiser can thus match the needs and preferences of more consumers. Some consumers, however, may have low $\varepsilon_{ij}$ because the advertiser does not offer a direct flight from their cities to Jeju Island or, even if there is a flight, the price of that specific

\footnote{For experience goods, consumers still have uncertainty regarding the actual consumption experience which information is not revealed from the advertiser’s website. We assume that the consumer’s objective is to maximize the expected value $V_{ij}$ conditional on all of the available information. As long as the uncertainty of post-purchase experience is independent from the conditional expectation, it will have no impact on the consumer’s search and purchase decisions.}
flight is higher than the others. Both $\xi_j$ and $\varepsilon_{ij}$ are unknown to consumers by only seeing selling propositions. We also assume that $\varepsilon_{ij}$ is iid across advertisers and consumers.

Search advertising is costly for advertisers via second-price auctions. The higher the ad position the higher the cost an advertiser has to pay for each click. Based on the signaling theory, consumers use the observed ad positions to infer the unobserved match value of advertisers’ offerings. Therefore, the expected utility of the consumer, prior to clicking the sponsored link, can be expressed as:

$$E(V_{ij}|p_j) = \alpha_j + E(\xi_j|p_j) + \varepsilon_{ij}.$$  \hspace{1cm} (4)

Observing that the ad position of advertiser $j$ ($p_j$) is higher than that of advertiser $k$ ($p_k$), the consumer will rationally infer that $E(\xi_j|p_j) > E(\xi_k|p_k)$. At a cost of clicking, $c_i$, the consumer can click on the advertiser’s sponsored link. In the above example, the cost comes from the time of searching for several flights from the departure city to Jeju Island that the advertiser offers on a specific date, checking for the price information, and possibly searching for packages with hotels and car rentals. Psychological costs (e.g., processing the information on flights and their prices) may also incur. Suppose the consumer searches optimally in a sequential way, she will click on the link if its expected match value, conditional on the ad position of the link, is higher than other sponsored links that have not been clicked, and the expected benefit of clicking the link outweighs the cost of clicking.\(^6\) This explains why consumers are more likely to click on top ad positions, offering us the first hypothesis to be tested:

**Hypothesis 1:** Top ad positions generate more clicks, after controlling for the partial information ($u_{ij}$) that consumers may have prior to clicking.

After the consumer clicks into the advertiser’s website, the match value component $v_{ij} = \xi_j + \varepsilon_{ij}$ will be revealed. If the consumer finds the advertiser’s offering a match (i.e., $V_{ij}$ is

\(^6\) In the stylized models in Athey and Ellison (2011) and Chen and He (2011), the link will be the topmost one in the sponsored listing. In our setup, however, consumer $i$ may start clicking advertiser $j$ listed at a lower position if the prior knowledge $u_{ij}$ is sufficiently high.
larger than the threshold value), she will terminate the search. Otherwise, the consumer will click on the next link which, among all other links that have not been clicked, has the highest expected match value, as long as the expected benefit of clicking the link outweighs the cost of clicking. The process will continue in a similar fashion. When the market is at equilibrium, the consumer’s belief is consistent with the advertiser’s bidding strategy. That is, if \( p_j \) is higher than \( p_k \) (and thus \( E(\xi_j | p_j) > E(\xi_k | p_k) \)), \( \xi_j \) is higher than \( \xi_k \). Therefore, compared with those who click at low ad positions, more consumers who click at top ad positions will be able to find a match.

We do not directly observe the match since we do not have consumer purchase data; however, we observe the entire sequence of sponsored and organic links that a consumer clicks. Since consumers will terminate their search after finding a match, we use terminal clicks of keyword searches to test the prediction. A terminal click refers to the last sponsored or organic link clicked by a consumer during the search process. It is a precondition that the advertiser’s offering is a match to the consumer. Chan and Park (2014) argue that, since there is little reason why the probability of non-match occasions for terminal clicks may differ across ad positions, terminal clicks can be a good proxy of the match likelihood. On this logic, non-terminal clicks and terminal clicks on organic links likely indicate that consumers, after clicking sponsored links, do not find a match from advertisers’ offering. Based on this reasoning, we propose the second hypothesis that focuses on consumer terminal clicks by ad positions:

**Hypothesis 2:** Conditional on being clicked, top ad positions have a higher likelihood for being a terminal click.

Our empirical tests extend beyond the consumer behavior on the search results page. The signaling theory states that, with a higher match value (\( \xi_j \)), an advertiser is willing to bid more for terminals.

---

\(^7\) Existing studies use aggregate data sourced from a single advertising firm which includes purchase conversions at the advertiser. However, they are conducted from the perspective of one single advertiser, and thus lack data on clicks on sponsored and organic links of other entities on the search results page. In other words, they do not have sufficient data to give a comprehensive picture of user behavior on the search results page. In this study, we use data from a search engine and have information on clicks on the full lists of sponsored and organic links presented after a keyword search.
a high ad position to signal the match value. This is because, after clicking the sponsored link, a large proportion of consumers will find the offering a good match; thus, each consumer click is more valuable to the advertiser (see Athey and Ellison 2011). For the theory to work, it is important that the bidding decision is driven by $\xi_j$, and not entirely by $\alpha_j$, the part that consumers already know before clicking the sponsored link. Testing this prediction is difficult, since consumers’ information set is unobserved to researchers. In this study, we make use of the panel structure of the data: we observe how bid amounts vary not only across advertisers, but also over time within advertisers. For each advertiser, $\xi_j$ is not necessarily fixed over time. Available flight times to Jeju Island, for example, may fluctuate within the advertiser, depending on how many tickets are sold as the flight date gets closer. We observe a unique feature in the data: often an advertiser’s ad position remains unchanged even when it increases the bid (i.e. the advertiser cannot outbid any of its competitors at higher positions). This is because the advertiser has uncertainty regarding how much other advertisers will bid, so the ad position only fluctuates with bid amounts in a probabilistic way. If the bid decision is driven by the increase in $\xi_j$, we should observe an increase in the likelihood of terminal clicks. However, since $\xi_j$ is not observed by consumers before clicking, the likelihood of clicking the advertiser’s sponsored link should not increase without the change in ad position. On the other hand, if the bid decision is entirely driven by the change in $\alpha_j$, the likelihood of clicking should also change accordingly. Based on this reasoning, we test the last hypothesis as the following:

Hypothesis 3: *The likelihood that an advertiser’s link is chosen for a terminal click improves (diminishes), as the advertiser bids more (less) even though the ad position remains unchanged. The likelihood that the link is chosen for a click, however, will not change.*

The three hypotheses constitute the signaling story: Consumers believe higher positioned advertisers to be of higher match value, advertisers with higher match value are on average in higher positions, and when its match value improves, an advertiser increases its bid. This paper
uses rich micro-level data to test these hypotheses. It is important to note that we do not argue that signaling is the only explanation for observed consumer and advertiser behaviors. A number of alternative explanations might predict similar outcomes. In a later section, we control for these non-signaling explanations for the results to investigate whether the signaling effects on both consumers and advertisers are still present.

4. Empirical Analyses and Results

4.1 Testing Hypothesis 1

Table 2 presents summary statistics about the click-through rate (CTR) and terminal click-through rate (TCTR) by ad positions. It shows that higher CTR associates with higher ad positions. This is consistent with the main finding from the empirical work in search advertising that advertisements at higher positions attract more clicks from consumers (e.g., Agarwal et al. 2009, Feng et al. 2007, Ghose and Yang 2009, Chan and Park 2014). The topmost position is especially valuable as its CTR is almost twice that of the second position. We also compare the CTR across positions for the three most commonly searched keywords and find similar results (see the last three columns in Table 2). These results provide support for Hypothesis 1, which is consistent with the predictions from the signaling model.

**Insert Table 2 about here**

Comparing the average CTR across ad positions, however, does not directly support the signaling theory on how search advertising works. The theory states that high ad positions are more likely to be clicked because they signal the unobserved $\xi$’s, not because of the observed $\alpha$’s (see equation (4)). Suppose an advertiser who has a reputation of high match value bids higher to obtain top positions. If consumers have prior knowledge about the reputation, they will be more likely to click the advertiser’s link. Another possibility is that the advertiser provides information about its high match value in the selling propositions (e.g., “50% off flights”). If the advertiser
also bids higher, data will exhibit the positive correlation between ad positions and CTR; however, these cases are not consistent with the signaling explanation that we study. Therefore, a regression approach is required for testing Hypothesis 1 to control for the \( \alpha \)'s of advertisers that consumers know prior to clicking sponsored links.

We use a reduced-form regression that is based on equation (4). We assume that consumer \( i \) makes a binary choice of whether or not to click on the sponsored link of advertiser \( j \), and the stochastic term \( e_{ij} \) has an EV Type 1 distribution (for the identification in the estimation). We estimate \( \alpha_j \) as a fixed effect. If an advertiser attracts clicks because of its reputation, the fixed effect will capture such effect thus the variation in the positions of the advertiser’s sponsored link should have no impact on CTR. To test the signaling effects, we estimate \( \delta_{p_j} \equiv \mathbb{E}(\xi_j|p_j) \) as a parameter that is specific for each ad position, representing consumers’ inference of the unobserved match value, conditional on the ad position, averaged across all advertisers on that position. As a result, the probability that user \( i \) clicks advertiser \( j \)'s link for a given keyword in the regression is as follows:

\[
\Pr(\text{user } i \text{ clicks advertiser } j \text{'s link}|p_j) = \frac{\exp(\alpha_j + \delta_{p_j})}{1 + \exp(\alpha_j + \delta_{p_j})}, \tag{5}
\]

The results for the effects of ad positions, \( \delta_{p_j} \), are reported in Table 3. The first two columns summarize the results from all keywords considered in this research, with the standard deviation of the estimates across keywords, and the last three columns report the results from the three most searched keywords as examples. The effect of the topmost position (Position 1) is normalized to zero, so the estimates for lower positions reflect the change in click potential relative to the topmost position. The estimates from Position 2 to Position 5 are all significant and negative for each keyword, implying that, after controlling for the prior information \( \alpha_j \), consumers expect a higher \( \xi_j \), as the advertiser’s ad position improves, and thus are more likely to click on

\(^8\) Competition effects from other advertisers are not included here. We will examine such effects later in the paper.
the website. The largest jump in the click probability occurs when an advertiser moves from Position 2 to Position 1. This result is consistent throughout all other analyses. To conclude, our test results support Hypothesis 1.9

**Insert Table 3 about here**

4.2 Testing Hypothesis 2

For the signaling theory to work, consumers should also be more likely to find the offering from advertisers at higher ad positions a match. Under the assumption that terminal clicks are a good measure for comparing matches, it is expected to find that advertisers at higher positions are more adept at producing terminal clicks once consumers have clicked the link. To test this hypothesis, we first look at the TCTR, unconditional on clicks, from the top to the bottom position in the data. The lower panel of Table 2 shows that the higher the ad position the more likely a sponsored link will be chosen as the terminal click during keyword search. Specifically, the topmost position generates twice as many terminal clicks as the second position across the keywords in this study. These patterns are also consistent when comparing the top three searched keywords.

We further utilize the regression approach to test Hypothesis 2. From equations (1) to (3), the consumer latent utility can be written as

\[ V_{ij} = \alpha_j + \xi_j + \omega_{ij}, \]

where \( \omega_{ij} \equiv u_{ij} + e_{ij} \). After clicking into a website, the consumer will terminate search if the advertiser’s offering matches her needs; otherwise she will continue searching. Conditional on the ad position \( p_j \), the probability that the link is a match implies that \( V_{ij} \) is larger than a threshold (normalized to zero) is

\[ \Pr(\text{user } i \text{ terminates search after clicking link } j | p_j) = \Pr(V_{ij} \geq 0 | p_j). \]

---

9 There are additional issues, as \( \alpha_j \) may vary over time and lead to the change in the advertiser’s ad position. In this case our estimated \( \delta_{nj} \) may be biased. Further, consumers may simply search in a non-strategic top-down manner. Such behavior could also explain our results. We will investigate these issues in the next section when we examine alternative explanations for the results.
The signaling theory predicts that $p_j$ is monotonically increasing with $\xi_j$; therefore, we can rewrite the above probability function as

$$
\Pr(\text{user } i \text{ terminates search after clicking link } j | p_j) = \Pr(\alpha_j + \xi(p_j) + \omega_{ij} \geq 0), \quad (6)
$$

where $\xi(p_j)$ is a monotonically increasing function of the ad position.

We use a reduced-form regression to test the relationship between $\xi_j$ and $p_j$. For the identification of model estimation, we assume that $\omega_{ij}$ has EV Type 1 distribution, and treat $\xi(p_j)$ as a position-specific parameter to be estimated. Therefore, probability function (6) is as follows,

$$
\Pr(\text{user } i \text{ terminates search after clicking link } j | p_j) = \frac{\exp(\alpha_j + \xi(p_j))}{1 + \exp(\alpha_j + \xi(p_j))}.
$$

(7)

Let $S_i$ be the set of links clicked by consumer $i$, and $t_{ij}$ be an indicator function that is equal to 1 if the user terminates search after clicking link $j$, and 0 otherwise. At most one link from $S_i$ has $t_{ij} = 1$ and, if none of the clicks are terminal clicks, we have $\sum_{j \in S_i} t_{ij} = 0$. The conditional likelihood of the choice of terminating clicks $\{t_{ij}; j \in S_i\}$ is

$$
\Pr(\{t_{ij}; j \in S_i\} | p) = \prod_{j \in S_i} \left( \frac{\exp(\alpha_j + \xi(p_j))}{1 + \exp(\alpha_j + \xi(p_j))} \right)^{t_{ij}} \times \left( 1 - \frac{\exp(\alpha_j + \xi(p_j))}{1 + \exp(\alpha_j + \xi(p_j))} \right)^{1-t_{ij}},
$$

(8)

where $p$ is the vector of all advertisers’ positions. Under this specification, the likelihood is the product of binary logit probabilities, conditional on the set of links that have been clicked.\(^\text{10}\) Table 4 presents the estimation results. The negative estimates for $\xi(p_j)$ suggest that the $\xi$’s at lower positions are lower than that for the top position, which is normalized to 0. Furthermore, the lower the ad position the smaller the estimated $\xi(p_j)$, a monotonic relationship predicted by the signaling theory. These results provide significant evidence supporting Hypothesis 2 and the

\(^{10}\) A multinomial logit (MNL) model is not a correct specification. The implicit assumption of the MNL model is that a consumer first clicks on a number of links and then determines which will be her terminal click. The main issue is that if only one sponsored link is clicked, the likelihood that the link is a terminal link is 100% no matter how estimates change; thus, this observation is non-informative for estimating the position effects. In our data most of these consumers choose the top-positioned link. On the other hand, consumers who click multiple links typically terminate their search at lower positions. This is because consumers often start searching at top positions and, if the links are not a match, they will continue down the list and stop searching at lower positions. We estimate such a model. Results show that, relative to the topmost position, lower-ranked positions are more likely to generate terminal clicks, clearly inconsistent with the observations from Table 2.
signaling prediction that advertisers at higher positions are more likely to match with consumers than advertisers at lower positions.

4.3 Testing Hypothesis 3

We now test the relationship between an advertiser’s bid decision and the likelihood that its link will match with a given consumer. As we have discussed before, advertisers in our data frequently change bids for specific keywords but their ad positions could remain unchanged (i.e., they cannot outbid advertisers on top). The signaling theory predicts that, when an advertiser increases its bid, it will generate more terminal clicks, while remaining at the same ad position, as the bid increase reflects an increase in the match value unobserved to consumers ($\xi_j$). The consumer will be more likely to terminate her search with the advertiser because, after clicking into the advertiser’s website, the consumer will be more likely to find a match.

We test the correlation between the TCTR of an advertiser and its bids. We construct indicator variables for every advertiser-position combination in the data set, and then model terminal click as a MNL decision out of all listed sponsored links for each search occasion. The probability that user $i$ chooses advertiser $j$ at position $l$ for the terminal click is as follows:

$$\Pr(\text{user } i \text{ terminal-clicks advertiser } j \text{'s link at position } l) = \frac{\exp(BID_{ij} \cdot \beta + \alpha_{jl})}{1 + \sum_{j'=1}^{J} \exp(BID_{ij'} \cdot \beta + \alpha_{j'l})},$$

where $\alpha_{jl}$ is the fixed effect for advertiser $j$ listed at position $l$, and $\alpha_{j'l}$ is the fixed effect for competing advertiser $j'$ at position $l'$. $BID_{ij}$ is the bid amount of advertiser $j$ for the keyword. Terminal clicks for organic links, or no clicks at all, are the outside option, whose value is captured by “1” in the denominator. We estimate 503 advertiser-position fixed effects in the regression. If changing bid amounts leads to changing positions, the effect on terminal clicks is captured by such fixed effects. The parameter $\beta$, therefore, is estimated from the variation in advertisers’ bids which does not alter their positions. The reason that we choose the MNL specification is because it controls for competitor effects, where other advertisers may also
increase bids and have an effect on the TCTR of advertiser $j$. Our estimation results show that $\beta$ is positive (0.56) and is significant at the 1% level. This indicates that an advertiser’s bid is positively correlated with the TCTR, even when its ad position remains unchanged, providing support for the first part of Hypothesis 3.

To test whether this positive correlation is due to some “macro shocks”, under which the advertisers simultaneously increase (decrease) bid amounts, due to increased (decreased) profitability from search advertising shared by everyone (e.g., advertisers may drop their bids at night if fewer clicks at night convert to purchases), but remain in the same ad positions, we further estimate another MNL model, allowing indicators for both the day of the month and the time of the day as additional explanatory variables in the regression. In this robustness check, $\beta$ remains positive (0.63) and significant at the 1% level. The results are strong evidence supporting the firm behaviors predicted by the signaling theory.

The second part of Hypothesis 3 states that the change in the advertiser’s bid decision is not driven by the change in $\alpha_j$, the part in the latent utility that consumers already know before clicking into the advertiser’s website. If only $\xi_j$ changes, the CTR should not correlate with the advertiser’s bid decision, since consumers do not observe how much each advertiser bids. To test this hypothesis, we further run a “placebo” test by using a logit model to regress clicks, rather than terminal clicks, against advertiser-position fixed effects and bid amounts. Our estimation results show that the estimated $\beta$ is small in magnitude and statistically insignificant ($\beta = 0.09$, $p$-value = 0.48). Therefore, when an advertiser increases its bid, but remains in the same position, it is more likely to generate terminal clicks, but not more likely to generate clicks. Combining all of these test results, it suggests that when advertisers increase their bids, it is not entirely due to the change in their service offering that is known by consumers prior to clicking sponsored links. They do so because of the improvement in the unknown match value, which they can signal through their ad positions.
In total, these tests have provided evidence for the three main predictions of signaling theory: consumers are more likely to click on advertisers in higher positions, consumers are more likely to match with advertisers in higher positions, and when advertisers will increase their bids, their offerings are more likely to match with consumers’ needs. In the next section we will discuss a number of possible non-signaling explanations for these results.

5. Alternative Explanations

In section 4 we have established the key results to support that signaling theory is at work in our empirical context. However, there may be other, non-signaling, explanations that lead to such results. In this section we explore several key alternative explanations and show that they cannot fully explain our findings. We do not argue that these factors do not affect consumer click behaviors, or that signaling is the only explanation; instead, our goal is to test whether, despite the existence of other factors, signaling match value plays a significant role in the search advertising market.

5.1 Endogeneity due to External Factors

We have shown a positive relationship between the positions of advertisers and their CTR and TCTR. Sponsored positions, however, are endogenously determined by the bidding decisions of advertisers. Suppose advertisers engage in marketing activities (e.g., an ad campaign in radio or on TV may increase consumers’ interest in the advertiser) that consumers know prior to clicking into their websites but as researchers we do not observe. If advertisers also bid for higher positions, this creates an endogeneity issue what can lead to biased results in the tests for Hypotheses 1 and 2. Notice that if an advertiser participates in such activities that increase its match value, and at the same time improves its ad position to send signals to consumers, this is consistent with the signaling model. We differentiate between the activities of which consumers are informed only via
signaling from ad positions, and activities that do not rely on signaling. To establish the signaling effect from search advertising, we have to control for the latter explanation.

Results from testing Hypothesis 3 partly address such a concern, since they suggest that advertisers’ bid decisions are not entirely driven by the activities that consumers already know. This strategy of testing, however, relies on the observations in which the advertiser’s ad position does not change with bids. What about the test results that are conditional on changes in ad positions? To rule out that they are driven by the external factors that are observed by consumers prior to clicking, we turn to the binary logit model in equation (5) that we use to test Hypothesis 1, where we assume $\alpha_j$ to be an advertiser-specific fixed effect. If the advertiser’s external activities (as well as any other factors that impact consumer clicks) remain unchanged in the data period, then the effect has been captured by $\alpha_j$. The concern is that the advertiser effect may change within the data period because of time-varying external factors. We use a unique feature in our data to control for this possible explanation: on average, an advertiser in data is placed in 2.4 (or 2 as the median) positions per day. We estimate the consumer click decision with a binary logit model, but allow for unique advertiser-day fixed effects. This estimation will separate out the effects from unobserved external factors that remain constant within a day from the position effects. The latter effects can only be identified from those advertisers whose ad positions have changed within a day. The results are presented in Table 5. We again find significant position effects that are of roughly the same magnitude to those estimates presented in Table 3. The results provide significant evidence that unobserved outside factors are not key drivers for the position effects, since their effects should have a more gradual effect on the consumer search behavior.

**Insert Table 5 about here**

After controlling for the advertiser-day fixed effects, there may still be a reason why $\alpha_j$ varies even within the same day. Sponsored search ads convey messages, selling propositions, to consumers regarding advertisers’ offerings or promotions in their websites. Consumers observe
the ads but as researchers we don’t. If, by creating more attractive ad message, advertisers also increase bids and thus obtain top positions, our results are generated by the change in the ad message, and not by the signaling effect. In our data, however, advertisers rarely change ad messages (108 of 116 advertiser-keyword combinations use the same selling propositions throughout the data period) even when positions change frequently. To ensure that the eight remaining advertisers who have used multiple selling propositions are not driving the position effects, we estimate the click model for three keywords, “Clearance Sale Flight Tickets”, “Travel Agency”, and “Travel to Jeju Island”. For these keywords, no advertisers change selling propositions within the sample period. We find that the position effects are still significant and of similar magnitude compared to the other keywords (see results from Table 3).

5.2 Top-Down Browsing

Another potential explanation is that consumers simply employ non-strategic top-down browsing. If the search cost is high, consumers are more likely to click top-positioned links and also terminate their search process there. There is an important distinction to be made here. If users are searching from top to bottom on the search results page because they infer that advertisers with higher match value are placed at the top, this is consistent with the signaling model we are testing. We will show that our findings are not entirely driven by non-strategic behaviors due to habit or preference for the simplicity of browsing.

To investigate this explanation, we classify top-down browsing as the case when a consumer clicks on sponsored links from top to down (they can skip links), then follow with clicking organic links, on the search results page. In 37% of the search occasions, in which consumers click on multiple links and at least one being a sponsored link, such top-down behavior is exhibited. Since we do not have data on the positions of the organic links clicked by a consumer, and it is possible that these consumers followed a strategic decision process that simply led to a top-down browsing pattern, this is the maximum proportion of consumer clicking non-
strategically from top down on the search results page, implying that at least 63% of the searches do not follow such behavior. We further investigate the position effects among those 63% search instances. We first study consumers whose first clicks are on an organic link, and employ the binary logit regression as in equation (5). This segment of consumers self-selects out of being a top-down searcher if they then click on a sponsored link. Note that this behavior does not violate the signaling model in a general sense, because consumers may collect information from organic links first, then go back to decide from which advertiser they will buy. Bentley et al. (2015) study why, when there are uncertainties for the keyword and advertisers’ offerings, consumers may search organic results before clicking sponsored links, and how the information from organic results impacts the probability of clicking sponsored links. Table 6 presents the regression results. The position effects are still significant, similar to our previous findings. These cannot be explained by the top-down browsing behavior.

**Insert Table 6 about here**

To further investigate the explanation of top-down browsing behavior, we estimate two additional models in which consumers select out of top-down browsing. First, we estimate the model conditional on observations in which consumers first click on the fifth (last) sponsored link. Very few consumers do so (1,124 out of 128,473 search instances), and even fewer follow up by clicking more sponsored links. Because of the lack of observations, we pool all keywords in the estimation and only focus on the effect from the top position. The estimated coefficient is 0.34 (the effects of lower-positions are normalized to zero), significant at the 1% level, indicating that the topmost position attracts more clicks than all other positions that are above the fifth link. We also estimate the model conditional on the first click being at any sponsored link lower than the topmost position. Again, we only estimate the effect from the topmost position, while normalizing the effects from all lower positions to be zero. Once again we find a significant positive top rank
effect (0.19 at the 5% level). Both results provide further evidence that the non-strategic top down behavior is not the sole driver of the results in our empirical context.

5.3 Persuasive Function

Another potential driver of our results on the signaling theory is persuasion in advertising. The signaling model suggests that consumers believe that advertisers with high match value are more likely to occupy high ad positions. In a broad sense, the persuasion explanation is consistent with such a model. The alternative explanation we investigate here is that the consumer belief is not rational, as advertisers with lower match value outbid advertisers with higher match value to gain clicks and sales from consumers who follow the belief. Jerath et al. (2011) offer a possible explanation why such bidding strategy is an equilibrium outcome and rationalize the so-called “Position Paradox”. They assume that uninformed consumers follow a top-down search strategy which, in some occasions, may not be optimal given advertisers’ bidding strategy, while informed consumers begin their search with the superior firm regardless of ad position in sponsored listing. Our goal is to test whether our findings are only driven by this pure persuasive advertising explanation.

Suppose consumers’ expectation is not consistent with advertisers’ bidding strategy. In such case, it is expected to find TCTR not to be positively correlated with the ad position. This is inconsistent with our findings when testing Hypothesis 2. A potential counter-argument is that, suppose search is costly for consumers. Consumers are more likely to stop searching after clicking top positions. However, this argument cannot explain our findings when testing Hypothesis 3. There should be no reasons that the TCTR increases with increasing bids from advertisers, when their ad positions remain constant. Also, since the CTR does not change (as we have shown in the “placebo” test) in this scenario, the persuasion function of ad positions does not seem to have improved. Therefore, the tests for Hypotheses 2 and 3 show that our findings are not solely driven by the pure persuasive function of ad positions.
We follow up with another test comparing the position effects across keywords. There are some “general” keywords that consist of multiple product or service offerings (e.g., “Flight Ticket” consists of flights to different destinations), and multiple attributes for the offerings (e.g., “Flight Ticket” consists of not only prices but also departure and arrival places and flight schedules), that consumers care about. For these keywords, the offerings from an advertiser may be a good match for some consumers but not for the others. The markets are therefore more horizontally differentiated. In contrast, there are “focused” keywords that are restricted to specific product or service offerings (e.g., “Flight Ticket to Jeju Island” focuses on a single destination) or attributes (e.g., “Clearance Sales Flight Ticket” targets consumers looking for low prices). For such keywords, the needs and preferences of consumers are well defined and less heterogeneous. The difference can be illustrated from equation (3). The magnitude of the variance of $\xi$’s across advertisers, relative to the magnitude of the variance of $\epsilon$’s across individual consumers and advertisers, should be larger for focused keyword than for general keywords. For focused keywords, advertisers with high $\xi$’s have a larger likelihood that their offerings match the needs and preferences of individual consumers who click at their sponsored links, so they are more likely to bid higher than advertisers with low $\xi$’s. Knowing this, rational consumers are also more likely to click on high-positioned sponsored links. The signaling function thus is stronger for focused keywords than general keywords. If our findings are only driven by persuasive advertising, however, there should not be any difference between “general” keywords and “focused” keywords regarding the advertiser and consumer behaviors, as consumers would not be rationally inferring advertisers’ bidding incentives.

We compare four pairs of keywords in which the first keyword in the pair is more general than the second, which is more focused, as follows:

1. “Flight Ticket” vs. “Flight Ticket to Jeju Island”
2. “Travel to Jeju Island” vs. “Flight Ticket to Jeju Island”
3. “Travel Agencies” and “Price Comparison of Travel Agencies”
4. “Flight Ticket” and “Clearance Sale Flight Ticket”

“Flight Ticket to Jeju Island” is similar to “Flight Ticket” without the variety of destinations, “Flight Ticket to Jeju Island” is a specific travel mode of "Travel to Jeju Island,” “Price Comparison of Travel Agencies” narrows down the attributes of interest for “Travel Agencies” to price, as does “Clearance Sale Flight Ticket” when compared to “Flight Ticket.”

We run a binary logit model of consumer clicks to estimate the position effects for general keywords, and the differences in the effects between the two types of keywords for each ad position. The coefficient for the top position is normalized to zero, so a negative coefficient for the differences will mean that the signaling effect (of top position, relative to lower positions) is stronger for focused keywords. Table 7 shows the results. All coefficients are negative, with the majority being significant. These results offer further evidence that the persuasive effect cannot fully explain our findings. Yet the signaling theory can rationalize why there are differences between keywords.

**Insert Table 7 about here**

5.4 Informative Function

Another important role of advertising identified in the literature is the informative function. Broadly speaking, the signaling function of advertising is also informative to consumers regarding the match value of advertisers. We examine here the narrow definition of the informative function, i.e., the selling proposition in an advertisement informs consumers about the existence of the advertiser’s website, prices, and product or service offerings that affects consumers’ click and purchase decisions.

If an ad in sponsored listing is to inform consumers of the advertiser’s website, there is no reason why top positions will get more clicks than lower positions, assuming that consumers browse all sponsored links before clicking. If consumers browse top-down and decide whether they will click each sponsored link sequentially, the advertiser at the top position may generate
more clicks. However, our previous tests, using observations in which consumers click organic links first and click lower positions first, do not support this argument. Therefore, we can rule out that our findings of the position effects are driven by the informative function of search advertising on the consumer awareness. Furthermore, since we have included advertiser fixed effects when testing Hypothesis 1, the position effects we have identified could be driven by the changes in selling propositions within advertisers. As we have shown in Section 5.1 when investigating the alternative explanation that the advertiser effects may be time-varying, advertisers rarely change ad messages even when positions change frequently. We also find that the position effects are still significant for the three keywords of which selling propositions have never changed. Therefore, the position effects we have documented are not driven by the informative function.

6. **Spillover Effects in Search Advertising**

We study in this section how the CTR of a sponsored link is influenced by competing sponsored links, and show the existence of an information externality that is consistent with the signaling framework. In search advertising, sponsored ads appear so that consumers can navigate through listings to search for information. The advertisement from an advertiser with good offerings may send a positive signal about the match value of other sponsored links. Jeziorski and Segal (2014) suggest that “such updating would generate positive informational externalities across ads, i.e., an ad would benefit from having better ads in the same impression.” On the other hand, such an advertiser will also attract consumers searching its website first and thus may reduce the likelihood of clicking other sponsored links, creating a competitive effect. Which effect dominates the other one is an important empirical question for online advertisers as well as search engines.

The phenomenon of the “informational externality” can be rationalized via the signaling model. In the model of Athey and Ellison (2011), consumers update the expected value for the $\xi$’s
for the links that have not been clicked, conditional on the revealed \( \xi \)'s from links that have been clicked, as well as the positions of all links, in a Bayesian way. Consider searching the keyword “Travel to Jeju Island” returns two sponsored links. Let the match value \( \xi \)'s be distributed uniformly from 0 to 1, with 1 being the highest possible value. Suppose a consumer finds that the average match value of the advertiser at the top position is high at 0.8 (e.g., the advertiser offers low prices for packaged tours to Jeju Island) but nothing matches with her specific needs (e.g., there are no tours on the dates she wants to travel). Based on the information of the match value of the advertiser, the consumer will infer the match value of the advertiser below is between 0 and 0.8, with the updated expectation equal to 0.4. She may continue her search by clicking the sponsored link below. If, instead, the average match value of the top advertiser is 0.2, the consumer will infer that the average match value of the advertiser below is between 0 and 0.2 with the updated expectation at 0.1. The consumer may be more likely to terminate her search after clicking the sponsored link at the top. This suggests that the existence of an advertiser with high match value may increase the number of clicks that a competing advertiser obtains, creating a “market expansion” effect, i.e., increasing the number of consumers who clicks sponsored links. Competition of course is also important. If most consumers, after first searching the website of the advertiser at the top, find a match and thus terminate the search, the advertiser below will not get additional clicks. In this case the negative competitive effect dominates the positive informational externality.

We use two ways to test the magnitude of the spillover effects. We first use the binary logit model from equation (5) but modify as follows:

\[
\Pr(\text{user } i \text{ clicks advertiser } j's \text{ link}) = \frac{\exp(\alpha_j + \delta p_j + \text{SPILLOVER}_{ij} \cdot \gamma)}{1 + \exp(\alpha_j + \delta p_j + \text{SPILLOVER}_{ij} \cdot \gamma)},
\]

where \( \text{SPILLOVER}_{ij} \) is a measure of the match value of all other advertisers above and below advertiser \( j \)'s ad position for a given keyword, and \( \gamma \) is a set of parameters that captures the net
spillover effects. The regression controls for the advertiser fixed effects, $\alpha_j$, and the position effect, $\delta_{pij}$: SPILLOVER$_{ij}$ is specified as a function of $\{\alpha_k, \text{for all } k \neq j\}$, the set of all advertisers competing with advertiser $j$ on the search results page. We estimate the above model with three specifications for SPILLOVER$_{ij}$. The first specification (“Mean”) takes the average of the estimated fixed effects from competing advertisers, from advertisers positioned above (zero for the topmost position) and positioned below (zero for the lowest position). The second specification (“Max”) takes the advertiser, positioned above and below, with the highest fixed effect. For the third specification (“Percent”), we calculate for a specific advertiser the percentage of advertisers who have a higher fixed effect, positioned above and below.

Since the click probability of each advertiser depends on not only its own fixed effect but also the fixed effects from advertisers at other ad positions, we estimate all advertiser fixed effects simultaneously through equation (9) for all of the advertisers. Results are reported in Table 8. For each specification, an increase in the match value of higher positioned advertisers will increase the click potential for a given advertiser, significant at the 1% level. In the Mean and Max specifications, we also find a positive spillover effect from lower-positioned advertisers, significant at the 10% and 5% level, respectively. Since the estimates represent the combined spillover effects and competitive effects, and because the latter is a negative effect, these results suggest that the positive information externalities from high-match value competitors are strong in the search advertising context.

**Insert Table 8 about here**

Our results suggest that high-match value advertisers increase consumer clicks for competing firms. It is possible that, due to their high match value, they may end up stealing terminal clicks. We next examine the impact of competition on consumers’ terminal click decisions using a MNL model. The dependent variable is, given the full set of sponsored links and an organic option, which link is chosen to be the terminal click. We employ the same three
specifications utilized above to generate the measure for competition (SPILLOVER$_{ij}$). Results are reported in the lower panel in Table 8. Again there is a significant positive impact from increased match value from advertisers at higher positions in all three specifications. In the Max specification we also see a significant positive impact from advertisers at lower positions.

The findings of positive spillover effects from high-match value competitors on consumer clicks and terminal clicks have an important managerial implication. In the advertising literature, previous studies have shown how competitors’ advertisements can “crowd-out” the effectiveness of advertising. In the search advertising context, past research has typically focused on the competitive effects (e.g., Mela and Yao 2011, Chan and Park 2014). These studies suggest that a high-match value advertiser will steal clicks and sales from other advertisers. Our results instead suggest the spillover effects from high-match value advertisers may increase consumer clicks and terminal clicks for competing firms.

7. Conclusions

In this paper, we test the empirical validity of the signaling theory of advertising in the search advertising context. By using detailed data of travel-related keywords which are obtained from a search engine, we have tested a series of predictions both on consumer and advertiser behaviors in accordance to the signaling theory of advertising. We have shown that consumers are more likely to not only click on an advertiser listed at higher positions but also terminate their search at such link. On the advertiser side, we find that the increase in terminal clicks is positively correlated with the increase in advertisers’ bid amounts, even when ad positions remain unchanged. However, there is no increase in CTR. In sum, our empirical results support predictions from the signaling theory of advertising in the literature. Additionally, we have controlled for several key alternative explanations for how search advertising works, and find that our results still hold. Finally, we find that advertisers can generate more clicks and terminal clicks when competing against advertisers
with higher match value, due to an information externality. This finding can be explained based on the signaling theory.

Our application has two important limitations that should be addressed by further research. First, we focus on a dataset of travel-related keywords to empirically test the signaling theory in the search advertising context. We chose the travel category because consumers have uncertainty regarding the attributes of products or services they search and their searches are costly in such experience goods markets with differentiated products or services. This means our conclusions may not generalize to other settings. Future work can build on our approaches and findings by analyzing other types of keywords and/or obtained from other sources, which can possibly lead to the empirical generalization regarding the relevance of the signaling theory in advertising. Second, we used terminal clicks to proxy whether an advertiser’s offering is a match for consumer needs, because our data did not include post-conversion behavior. The availability of data on post-click conversion rates can further enhance our understanding of consumer behavior. By the same token, the availability of data on advertisers’ other types of marketing activities can help study the signaling theory in a broader advertising context. We hope that this study demonstrates the efficacy of the signaling theory and can motivate further research in the contexts where both consumers and advertisers strategically behave when uncertainty exists.
References


36


### Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Across all keywords</th>
<th>Clearance Sale</th>
<th>Travel Agency</th>
<th>Travel to Jeju Island</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search volume</td>
<td>16,059</td>
<td>34,801</td>
<td>26,163</td>
<td>19,479</td>
</tr>
<tr>
<td>Advertisers</td>
<td>15.6</td>
<td>11</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Positions held daily per advertiser</td>
<td>2.4</td>
<td>2.1</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Selling propositions per advertiser</td>
<td>1.2</td>
<td>1.0</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>Advertisers per ad position</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1</td>
<td>9.4</td>
<td>8</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Position 2</td>
<td>10.5</td>
<td>8</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Position 3</td>
<td>13.0</td>
<td>10</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Position 4</td>
<td>13.9</td>
<td>11</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>Position 5</td>
<td>15.3</td>
<td>11</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>CPC ($)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1</td>
<td>0.57</td>
<td>0.30</td>
<td>0.55</td>
<td>0.71</td>
</tr>
<tr>
<td>Position 2</td>
<td>0.53</td>
<td>0.28</td>
<td>0.53</td>
<td>0.65</td>
</tr>
<tr>
<td>Position 3</td>
<td>0.49</td>
<td>0.25</td>
<td>0.50</td>
<td>0.63</td>
</tr>
<tr>
<td>Position 4</td>
<td>0.45</td>
<td>0.21</td>
<td>0.43</td>
<td>0.58</td>
</tr>
<tr>
<td>Position 5</td>
<td>0.40</td>
<td>0.18</td>
<td>0.34</td>
<td>0.50</td>
</tr>
</tbody>
</table>

### Table 2: CTR and TCTR

<table>
<thead>
<tr>
<th></th>
<th>Across all keywords</th>
<th>Clearance Sale</th>
<th>Travel Agency</th>
<th>Travel to Jeju Island</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTR (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1</td>
<td>5.1</td>
<td>6.2</td>
<td>3.3</td>
<td>7.7</td>
</tr>
<tr>
<td>Position 2</td>
<td>2.7</td>
<td>2.9</td>
<td>1.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Position 3</td>
<td>2.0</td>
<td>2.1</td>
<td>1.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Position 4</td>
<td>1.6</td>
<td>1.5</td>
<td>1.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Position 5</td>
<td>1.3</td>
<td>1.2</td>
<td>1.1</td>
<td>1.7</td>
</tr>
<tr>
<td>TCTR (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1</td>
<td>2.4</td>
<td>2.8</td>
<td>1.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Position 2</td>
<td>1.2</td>
<td>1.2</td>
<td>0.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Position 3</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Position 4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Position 5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

### Table 3: Results of Click Behavior

<table>
<thead>
<tr>
<th></th>
<th>Across all keywords</th>
<th>Clearance Sale</th>
<th>Travel Agency</th>
<th>Travel to Jeju Island</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 1: Base</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Position 2</td>
<td>-0.72</td>
<td>-0.82 a</td>
<td>-0.62 a</td>
<td>-0.59 a</td>
</tr>
<tr>
<td>Position 3</td>
<td>-0.97</td>
<td>-1.13 a</td>
<td>-0.73 a</td>
<td>-0.95 a</td>
</tr>
<tr>
<td>Position 4</td>
<td>-1.09</td>
<td>-1.12 a</td>
<td>-0.99 a</td>
<td>-0.99 a</td>
</tr>
<tr>
<td>Position 5</td>
<td>-1.13</td>
<td>-1.28 a</td>
<td>-0.97 a</td>
<td>-1.13 a</td>
</tr>
</tbody>
</table>

*a Significant at the 1 percent level*
Table 4: Results of Terminal Click Behavior

<table>
<thead>
<tr>
<th>Across all keywords</th>
<th>Clearance Sale</th>
<th>Travel Agency</th>
<th>Travel to Jeju Island</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 1: Base</td>
<td>0.00</td>
<td>--</td>
<td>0.00</td>
</tr>
<tr>
<td>Position 2</td>
<td>-0.18</td>
<td>0.22</td>
<td>-0.21&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 3</td>
<td>-0.25</td>
<td>0.25</td>
<td>-0.22&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 4</td>
<td>-0.20</td>
<td>0.24</td>
<td>-0.24&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 5</td>
<td>-0.34</td>
<td>0.49</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

<sup>a</sup> Significant at the 1 percent level  
<sup>b</sup> Significant at the 5 percent level  
<sup>c</sup> Significant at the 10 percent level

Table 5: Results of Click Behavior with Advertiser-Day Fixed Effects

<table>
<thead>
<tr>
<th>Across all keywords</th>
<th>Clearance Sale</th>
<th>Travel Agency</th>
<th>Travel to Jeju Island</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 1: Base</td>
<td>0.00</td>
<td>--</td>
<td>0.00</td>
</tr>
<tr>
<td>Position 2</td>
<td>-0.71</td>
<td>0.22</td>
<td>-0.77&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 3</td>
<td>-0.90</td>
<td>0.28</td>
<td>-1.16&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 4</td>
<td>-1.02</td>
<td>0.36</td>
<td>-1.25&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 5</td>
<td>-1.09</td>
<td>0.37</td>
<td>-1.25&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Significant at the 1 percent level

Table 6: Results of Click Behavior with First Clicks Being on Organic Links

<table>
<thead>
<tr>
<th>Across all keywords</th>
<th>Clearance Sale</th>
<th>Travel Agency</th>
<th>Travel to Jeju Island</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 1: Base</td>
<td>0.00</td>
<td>--</td>
<td>0.00</td>
</tr>
<tr>
<td>Position 2</td>
<td>-0.55</td>
<td>0.30</td>
<td>-0.73&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 3</td>
<td>-0.71</td>
<td>0.20</td>
<td>-1.00&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 4</td>
<td>-0.86</td>
<td>0.21</td>
<td>-1.07&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Position 5</td>
<td>-0.87</td>
<td>0.34</td>
<td>-1.05&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Significant at the 1 percent level
Table 7: Results of Click Behavior with Differences in Position Effects between General and Focused Keywords

<table>
<thead>
<tr>
<th>Keyword Pairs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>General keywords</td>
<td>Flight Ticket</td>
<td>Travel to Jeju Island</td>
<td>Travel Agencies</td>
<td>Flight Ticket</td>
</tr>
<tr>
<td>Focused keywords</td>
<td>Flight Ticket to Jeju Island</td>
<td>Flight Ticket to Jeju Island</td>
<td>Price Comparison of Travel Agencies</td>
<td>Clearance Sale Flight Ticket</td>
</tr>
<tr>
<td>Position 1: Base</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Position 2</td>
<td>-0.12</td>
<td>-0.25(^b)</td>
<td>-0.46(^a)</td>
<td>-0.11</td>
</tr>
<tr>
<td>Position 3</td>
<td>-0.31(^b)</td>
<td>-0.19</td>
<td>-0.72(^a)</td>
<td>-0.32(^a)</td>
</tr>
<tr>
<td>Position 4</td>
<td>-0.43(^a)</td>
<td>-0.30(^b)</td>
<td>-0.58(^a)</td>
<td>-0.34(^a)</td>
</tr>
<tr>
<td>Position 5</td>
<td>-0.31(^c)</td>
<td>-0.13</td>
<td>-0.77(^a)</td>
<td>-0.29(^b)</td>
</tr>
</tbody>
</table>

\(^a\) Significant at the 1 percent level  
\(^b\) Significant at the 5 percent level  
\(^c\) Significant at the 10 percent level

Table 8: Results of Click and Terminal-Click Behavior with Spillover Effects

<table>
<thead>
<tr>
<th>Specification</th>
<th>Mean</th>
<th>Max</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary logit model with clicks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With effects from advertisers listed above</td>
<td>0.14(^a)</td>
<td>0.12(^a)</td>
<td>0.28(^a)</td>
</tr>
<tr>
<td>With effects from advertisers listed below</td>
<td>0.03(^c)</td>
<td>0.04(^b)</td>
<td>0.05</td>
</tr>
<tr>
<td>Multinomial logit model with terminal clicks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With effects from advertisers listed above</td>
<td>0.08(^a)</td>
<td>0.11(^a)</td>
<td>0.28(^a)</td>
</tr>
<tr>
<td>With effects from advertisers listed below</td>
<td>-0.01</td>
<td>0.06(^c)</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

\(^a\) Significant at the 1 percent level  
\(^b\) Significant at the 5 percent level  
\(^c\) Significant at the 10 percent level