An Empirical Analysis of Sponsored Search Performance in Search Engine Advertising

Anindya Ghose
Stern School of Business
New York University
New York, NY-10012
aghose@stern.nyu.edu

Sha Yang
Stern School of Business
New York University
New York, NY-10012
syang0@stern.nyu.edu

ABSTRACT
The phenomenon of sponsored search advertising – where advertisers pay a fee to Internet search engines to be displayed alongside organic (non-sponsored) web search results – is gaining ground as the largest source of revenues for search engines. Despite the growth of search advertising, we have little understanding of how consumers respond to contextual and sponsored search advertising on the Internet. Using a unique panel dataset of several hundred keywords collected from a large nationwide retailer that advertises on Google, we empirically model the relationship between different metrics such as click-through rates, conversion rates, bid prices and keyword ranks. Our paper proposes a novel framework and data to better understand what drives these differences. We use a Hierarchical Bayesian modeling framework and estimate the model using Markov Chain Monte Carlo (MCMC) methods. We empirically estimate the impact of keyword attributes on consumer search and purchase behavior as well as on firms' decision-making behavior on bid prices and ranks. We find that the presence of retailer-specific information in the keyword increases click-through rates, and the presence of brand-specific information in the keyword increases conversion rates. We also demonstrate that as suggested by anecdotal evidence, search engines like Google factor in both the auction bid price as well as prior click-through rates before allotting a final rank to an advertisement. To the best of our knowledge, this is the first study that uses real world data from an advertiser and jointly estimates the effect of sponsored search advertising at a keyword level on consumer search, click and purchase behavior in electronic markets.

Categories and Subject Descriptors
J.4 [Social and Behavioral Sciences]: Economics

General Terms
Performance, Measurement, Economics.

Keywords
Online advertising, Search engines, Hierarchical Bayesian modeling, Paid search advertising, Electronic commerce, Internet markets.

1. INTRODUCTION
The Internet has brought about a fundamental change in the way consumers obtain information. It is well documented now that the online retailing revolution has established a new distribution channel that represents a fundamental paradigm shift in consumer search and purchase patterns. Hence, firms are realizing that increasing the number of online customers by attracting them to their websites is the key to bolstering sales. In this regard, search engines are able to leverage the value as information location tools by selling advertising linked to search terms entered by online users and referring them to the advertisers. Indeed, the phenomenon of sponsored search advertising – where advertisers pay a fee to Internet search engines to be displayed alongside organic (non-sponsored) web search results – is gaining ground as the largest source of revenues for search engines. The global paid search advertising market is predicted to have a 37 percent compound annual growth rate (CAGR), to more than $33 billion in 2010 and has become a critical component of firm's marketing campaigns. This is not surprising given that 94% of consumers use search engines to find information on the Web, and 81% who use search engines find the information they are looking for every time they search (Nielson-Net Ratings).

Search engines like Google, Yahoo and MSN have discovered that as intermediaries between users and firms, they are in a unique position to try new forms of advertisements without annoying consumers. In this regard, the advent of sponsored search advertisements – the delivery of relevant, targeted text advertisements as part of the search experience, makes it increasingly possible for firms to attract consumers to their websites. How does this mechanism work? In sponsored search, advertisers who wish to market their product or services on the Internet submit their website information in the form of keyword listings to search engines. Bid values are assigned to each individual keyword to determine the placement of each listing among search results when a user performs a search. Basically, search engines pit advertisers against each other in auction-style bidding for the highest ad placement positions on search result pages. When a consumer searches for that term on a search engine, the advertisers' web page appears as a sponsored link next to the organic search results that would otherwise be returned using the neutral criteria employed by the search engine. Different search engines had different advertising models but most of them are now moving towards the Google auction system where both bid price and previous click-through rates are factored in before the final ranks are allotted to different advertisers.

Sponsored search has gradually evolved to satisfy consumers' penchant for relevant search results and advertisers' desire for inviting high quality traffic to their websites. These keyword advertisements are based on customers' own queries and are thus considered far less intrusive than online banner advertisements or pop-ups. In many ways, one could imagine that this enabled a shift in advertising from 'mass' advertising to more 'targeted' advertising. By allotting a specific value to each keyword, an advertiser only pays the assigned price for the people who click on their listing to visit its website. Because listings appear when a keyword is searched for, an advertiser can reach a more targeted audience on a much lower budget. Hence, it is now considered to be among the most effective marketing vehicles available in the online world.[1]

Despite the growth of search advertising, we have little understanding of how consumers respond to contextual and sponsored search advertising on the Internet. In this paper, we focus on previously unexplored questions: How does sponsored search advertising affect search and purchasing patterns on the Internet? What features of a sponsored keyword advertisement do consumers respond to the most during web search? Do advertising firms exhibit any learning behavior? While an emerging stream of theoretical literature in sponsored search has looked at issues such as mechanism design in keyword auctions, no prior work has empirically analyzed these questions. Given the shift in advertising from traditional banner advertising to search engine advertising, an understanding of the determinants of conversion rates and click-through rates in search advertising can be useful for both traditional and Internet retailers.

Using a unique panel dataset of several hundred keywords collected from a large nationwide retailer that advertises on Google, we study the effect of sponsored search advertising at a keyword level on consumer search, click and purchase behavior in electronic markets. We propose a Hierarchical Bayesian modeling framework in which we model consumers' behavior jointly with the advertiser's decisions. To the best of our knowledge, our paper is the first empirical study that models and documents the impact of search advertising on consumer's click-through, conversion and purchase behavior in electronic markets.

We empirically estimate the impact of keyword attributes (such as the presence of *retailer information*, *brand information* and the *length of the keyword*) on consumer click-through, and purchase propensities. This classification is motivated by prior work on the goals for users' web search such as [3, 19]. We find that the presence of retailer-specific information in the keyword increases click-through rates, and the presence of brand-specific information in the keyword increases conversion rates. Further, we provide some evidence that firms exhibit learning behavior over time. This learning is based on naïve measures such as rank of the same advertisement in the previous time period or more sophisticated measures such as profit accruing from the same advertisement in the previous period.

[1] Search engines relied on banner advertising before the adoption of sponsored search, so they faced a dilemma - : keep users on the site as long as possible to view more banners or send the users promptly to the sites appearing in the search results. Paid search reconciled this dilemma by tying the search engine's revenue to the act of transferring the user to an advertiser's site. (http://www.asis.org/Bulletin/Dec-05/pedersen.html)

2. ECONOMETRIC MODEL: A SIMULTANEOUS MODEL OF CLICK-THROUGH, PURCHASE, AND RANKING

We cast our model in a hierarchical Bayesian framework and estimate it using Markov Chain Monte Carlo (MCMC) methods (see [24] for a detailed review of such models). We postulate that the decision of whether to click and purchase in a given week will be affected by the probability of advertising exposure (for example, through the rank of the keyword) and individual differences (both observed and unobserved heterogeneity). We use the Metropolis-Hastings algorithm with a random walk chain to generate draws in the MCMC methods ([6]).

Assume for search keyword i at week j, there are n_{ij} click-throughs among N_{ij} impressions (the number of times an advertisement is displayed by the retailer), where $n_{ij} \leq N_{ij}$. Suppose that among the n_{ij} click-throughs, there are m_{ij} click-throughs that lead to purchases, where $m_{ij} \leq n_{ij}$. Let us further assume that the probability of having a click-through is p_{ij} and the probability of having a purchase is q_{ij}. In our model, a consumer faces decisions at two levels – one, when she sees a keyword advertisement, she makes decision whether or not to click it; two, if she clicks on the advertisement, she can take any one of the following two actions – make a purchase or not make a purchase.

Thus, there are three types of observations. First, a person clicked through and made a purchase. The probability of such an event is $p_{ij}q_{ij}$. Second, a person clicked through but did not make a purchase. The probability of such an event is $p_{ij}(1- q_{ij})$. Third, an impression did not lead to a click-through or purchase. The probability of such an event is $1- p_{ij}$. Then, the probability of observing (n_{ij}, m_{ij}) is given by:

$$f(n_{ij}, m_{ij}, p_{ij}, q_{ij}) = \frac{N_{ij}!}{m_{ij}!(n_{ij}-m_{ij})!(N_{ij}-n_{ij})!} \times \{p_{ij}q_{ij}\}^{m_{ij}}\{p_{ij}(1-q_{ij})\}^{n_{ij}-m_{ij}}\{1-p_{ij}\}^{N_{ij}-n_{ij}} \qquad (3.1)$$

2.1 Modeling Click-throughs

The click-through probability is likely to be influenced by the position of the ad (*Rank*), how specific or broad the keyword is (*Length*), and whether is contains any retailer-specific (*Retailer*) or brand-specific information (*Brand*). Hence, in equation (3.1), p_{ij}, the click-through probability is modeled as:

$$p_{ij} = \frac{\begin{pmatrix} \exp(\beta_{i0} + \beta_{i1}\text{Rank}_{ij} + \alpha_1 \text{Retailer}_i + \\ \alpha_2 \text{Brand}_i + \alpha_3 \text{Length}_i + \varepsilon_{ij}) \end{pmatrix}}{1 + \begin{pmatrix} \exp(\beta_{i0} + \beta_{i1}\text{Rank}_{ij} + \alpha_1 \text{Retailer}_i + \\ \alpha_2 \text{Brand}_i + \alpha_3 \text{Length}_i + \varepsilon_{ij}) \end{pmatrix}} \tag{3.2}$$

We capture the unobserved heterogeneity with the distribution of β_{i0} (where β_{i0} is the intercept of i^{th} keyword) by allowing it to be varying along its population mean $\overline{\beta_0}$ as follows:

$$\beta_{i0} = \overline{\beta_0} + \varsigma_{i0}^{\beta} \tag{3.3}$$

We allow the rank coefficient of the i^{th} keyword to vary along the population mean $\overline{\beta_1}$ and the keywords' characteristics as follows:

$$\beta_{i1} = \overline{\beta_1} + \gamma_1 Retailer_i + \gamma_2 Brand_i + \gamma_3 Length_i + \varsigma_{i1}^{\beta} \tag{3.4}$$

$$\begin{bmatrix} \varsigma_{i0}^{\beta} \\ \varsigma_{i1}^{\beta} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11}^{\beta} & \Sigma_{12}^{\beta} \\ \Sigma_{21}^{\beta} & \Sigma_{22}^{\beta} \end{bmatrix}\right) \tag{3.5}$$

2.2 Modeling Conversion Rates

The conversion probability is likely to be influenced by the position of the ad (*Rank*), how specific or broad the keyword is (*Length*), and whether it contains any retailer-specific (*Retailer*) or brand-specific information (*Brand*). In addition the click-through rate (CTR) will also have an impact on conversion rates. Hence, in equation (3.1), q_{ij} , the conversion probability is modeled as follows:

$$q_{ij} = \frac{\begin{pmatrix} \exp(\theta_{i0} + \theta_{i1}\text{Rank}_{ij} + \overline{\theta_2}\text{CTR}_{ij} + \\ \delta_1 \text{Retailer}_i + \delta_2 \text{Brand}_i + \delta_3 \text{Length}_i + \eta_{ij}) \end{pmatrix}}{1 + \begin{pmatrix} \exp(\theta_{i0} + \theta_{i1}\text{Rank}_{ij} + \overline{\theta_2}\text{CTR}_{ij} + \\ \delta_1 \text{Retailer}_i + \delta_2 \text{Brand}_i + \delta_3 \text{Length}_i + \eta_{ij}) \end{pmatrix}} \tag{3.6}$$

$$\theta_{i0} = \overline{\theta_0} + \varsigma_{i0}^{\theta} \tag{3.7}$$

$$\theta_{i1} = \overline{\theta_1} + \kappa_1 Retailer_i + \kappa_2 Brand_i + \kappa_3 Length_i + \varsigma_{i1}^{\theta} \tag{3.8}$$

$$\begin{bmatrix} \varsigma_{i0}^{\theta} \\ \varsigma_{i1}^{\theta} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11}^{\theta} & \Sigma_{12}^{\theta} \\ \Sigma_{21}^{\theta} & \Sigma_{22}^{\theta} \end{bmatrix}\right) \tag{3.9}$$

Thus, equations (3.1) - (3.9) model the demand for a keyword.

2.3 Modeling the Advertiser's Decision – Bid Price

Next, we model the advertiser's (i.e., the firm's) strategic behavior. The advertiser decides its bidding strategy in terms of how much to bid for each keyword at week j. Since the firm optimizes its advertising strategies based on learning from past performances, we take into account two types of learning. The *first* is the most naïve learning that involves bidding sufficiently high so as to secure a good rank. This kind of learning is based on the outcome from the keyword's rank in the previous period. The *second* kind is the more sophisticated kind of learning that will be based on the keyword's profit in the previous time period where profit is defined as revenues from sponsored search advertising minus the costs of placing that advertisement for the firm (the cost is equal to the total number of clicks times cost per click).

These learning mechanisms can be expressed as follows:

$$\begin{aligned} ln(BidPrice_{ij}) &= \omega_{i0} + \omega_{i1} Rank_{i,j-1} + \omega_{i2} Profit_{i,j-1} + \\ &\lambda_1 Retailer_i + \lambda_2 Brand_i + \lambda_3 Length_i + \mu_{ij} \end{aligned} \tag{3.10}$$

$$\omega_{i0} = \overline{\omega_0} + \varsigma_{i0}^{\omega} \tag{3.11}$$

$$\omega_{i1} = \overline{\omega_1} + \rho_{11} Retailer_i + \rho_{12} Brand_i + \rho_{13} Length_i + \varsigma_{i1}^{\omega} \tag{3.12}$$

$$\omega_{i2} = \overline{\omega_2} + \rho_{21} Retailer_i + \rho_{22} Brand_i + \rho_{23} Length_i + \varsigma_{i2}^{\omega} \tag{3.13}$$

The error terms in equations (4.11) – (4.13) are distributed as follows:

$$\begin{bmatrix} \varsigma_{i0}^{\omega} \\ \varsigma_{i1}^{\omega} \\ \varsigma_{i2}^{\omega} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11}^{\omega} & \Sigma_{12}^{\omega} & \Sigma_{13}^{\omega} \\ \Sigma_{21}^{\omega} & \Sigma_{22}^{\omega} & \Sigma_{23}^{\omega} \\ \Sigma_{31}^{\omega} & \Sigma_{32}^{\omega} & \Sigma_{33}^{\omega} \end{bmatrix}\right) \tag{3.14}$$

2.3 Modeling the Search Engine's Decision – Rank

Next, we model the search engine's strategic behavior. The search engine decides on the ranking of each search keyword base on the submitted bid price from the advertiser and its previous click-through rate.

$$\begin{aligned} ln(Rank_{ij}) &= \phi_{i0} + \phi_{i1} BidPrice_{i,j} + \overline{\phi_2} CTR_{i,j-1} + \\ &\tau_1 Retailer_i + \tau_2 Brand_i + \tau_3 Length_i + \nu_{ij} \end{aligned} \tag{3.15}$$

$$\phi_{i0} = \overline{\phi_0} + \varsigma_{i0}^{\phi} \tag{3.16}$$

$$\phi_{i1} = \overline{\phi_1} + \pi_1 Retailer_i + \pi_2 Brand_i + \pi_3 Length_i + \varsigma_{i1}^{\pi} \tag{3.17}$$

The error terms in equations (4.16) and (4.17) are distributed as follows:

$$\begin{bmatrix} \varsigma_{i0}^{\phi} \\ \varsigma_{i1}^{\phi} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11}^{\phi} & \Sigma_{12}^{\phi} \\ \Sigma_{21}^{\phi} & \Sigma_{22}^{\phi} \end{bmatrix}\right) \tag{3.18}$$

Finally, to model the unobserved co-variation among click-through, conversions, bid price and the keyword ranking, we let the four error terms to be correlated in the following manner:

$$\begin{bmatrix} \varepsilon_{ij} \\ \eta_{ij} \\ \mu_{ij} \\ \nu_{ij} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \Omega_{14} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \Omega_{34} \\ \Omega_{41} & \Omega_{42} & \Omega_{43} & \Omega_{44} \end{bmatrix}\right) \tag{3.19}$$

A couple of clarifications are useful to note here. First, the three characteristics of a keyword (*Retailer*, *Brand*, *Length*) are all mean centered. This means that $\overline{\beta_1}$ is the average effect of $\overline{\beta_{i1}}$ in equation (3.4). A similar interpretation applies to the parameters θ_{i1}, ω_{i1}, and ω_{i2}. Second, in equations (3.6) and (3.10), the coefficient of click-through rate (*CTR*) is modeled as a fixed effect rather as a random coefficient in order to facilitate empirical identification. Due to the fact that we have a large number of observations with zero click-through rates, empirical identification is difficult if we were to model *CTR* as a random coefficient.

To ensure that the model is fully identified even with sparse data (data in which a large proportion of observations are zero), we conduct the following simulation. We picked a set of parameter values, and generated the number of click-throughs, the number of purchases, and ranking for each keyword, which mimicked their actual observed values in the data according to the model and the actual independent variables observed in our data. We then estimated the proposed model with the simulated dataset and found that we were able to recover the true parameter values. This relieves a potential concern on empirical identification of the model with due to the sparseness of the data.

3. DATA

3.1 Data Description

We first describe the data generation process for paid search advertisement since it differs on many dimensions from traditional offline advertisement. Once the advertiser gets a rank allotted (based on the bid price) to display its textual ad, these sponsored ads show up on the top left, right and bottom of the computer screen in response to a query that a consumer types on the search engine. The textual ad typically consists of headline, a word or a limited number of words describing the product or service and a hyperlink that refers the consumer to the advertiser's website after a click. The serving of a text ad in response to a query for a certain keyword is denoted as an impression. If the consumer clicks on the ad, he is led to the landing page of the advertiser's website. This is recorded as a click, and advertisers usually pay on a per click basis. In the event that the consumer ends up purchasing a product from the advertiser, this is recorded as a conversion. The time between a click and an actual purchase is known as latency. This is usually measured in days. In the majority of cases the value of this variable is 0, denoting that the consumer placed an order at the same time as when they landed on a firm's website.

Our data contains weekly information on paid search advertising from a large nationwide retail chain, which advertises on Google.[2] The data span *all keyword advertisements* by the company during a period of three months in the first quarter of 2007, specifically for the 13 calendar weeks from January 1 to March 31. Unlike most datasets used to investigate on-line environments which usually comprise of browsing behavior only, our data are unique in that we have individual level stimulus (advertising) and response (purchase incidence).

Each keyword in our data has a unique advertisement ID. The data consists of the number of impressions, number of clicks, the average cost per click (CPC) which represents the bid price in the case of successful bid, the rank of the keyword, the number of conversions, the total revenues from a click (revenues from conversion) and the average order value for a given keyword for a given week. While a search can lead to an impression, and often to a click, it may not lead to an actual purchase (defined as a conversion). The product of CPC and number of clicks gives the total costs to the firm for sponsoring a particular advertisement. Thus the difference in total revenues and total costs gives the total profits accruing to the retailer from advertising a given keyword in a given week. Our dataset includes 5147 observations from a total of 1799 unique keywords that had at least one positive impression.

3.2 Keyword Characteristics

There are three important keyword specific characteristics for a firm (the advertiser) when it advertises on a search engine. This includes whether the keyword should have (i) retailer-specific information, (ii) brand-specific information, (iii) and the length (number of words) of the keyword. A consumer seeking to purchase a digital camera is as likely to search for a popular brand name such as NIKON, CANON or KODAK on a search engine as searching for the generic phrase "digital camera" on the same search engine. Similarly, the same consumer may search directly for a retailer such as "BEST BUY" or "CIRCUIT CITY" on the search engine. In recognition of these electronic marketplace realities, search engines do not merely sell generic identifiers such as "digital cameras" as keywords, but also well-known brand names that can be purchased by any third-party advertiser in order to attract consumers to its Web site.

The length of the keyword is also an important determinant of search and purchase behavior but anecdotal evidence on this varies across trade press reports. Some studies have shown that the percentage of searchers who use a combination of keywords is 1.6 times the percentage of those who use single-keyword queries [19]. In contrast, in 2005 Oneupweb conducted a study to determine if the number of keywords in a search query was related to conversion rates. They focused their study on data generated by natural or organic search engine results listings and

[2] The firm is a Fortune-500 firm but due to the nature of the data sharing agreement between the firm and us, we are unable to reveal the name of the firm.

found that single-keywords have on average the highest number of unique visitors. To investigate the impact of the length of a keyword, we constructed a variable that indicates the number of words in a keyword that a user queried for on the search engine (in response to which the paid advertisement was displayed to the user).

We enhanced the dataset by introducing some keyword-specific characteristics such as Brand, Retailer and Length. For each keyword, we constructed two dummy variables, based on whether they were (i) branded or unbranded keywords and (ii) retailer-specific or non-retailer specific keywords. To be precise, for creating the variable in (i) we looked for the presence of a brand name (either a product-specific or a company specific) in the keyword, and labeled the dummy as 1 or 0, with 1 indicating the presence of a brand name. For (ii), we looked for the presence of the advertising retailer's name in the keyword, and then labeled the dummy as 1 or 0, with 1 indicating the presence of the retailer's name. There were no keywords that contained both retailer name and brand name information. This enabled a clean classification in our data. This classification is similar in notion to [3, 19] who classify user queries in search engines as *navigational* (searching for a specific firm or retailer), *transactional* (searching for a specific product) or *informational* (longer keywords).

Table 1: Summary Statistics (Keyword level)

Variable	*Mean*	*Std. Dev.*	*Min*	*Max*
Impressions	383.37	2082.08	1	97424
Clicks	32.915	519.555	0	33330
Orders	0.483	8.212	0	527
Click-through Rate (CTR)	0.008	0.059	0	1
Conversion Rate	0.013	0.073	0	1
Bid Price	0.294	0.173	0.005	1.410
Lag Rank	4.851	6.394	1	64
Log (Lag Profit)	0.106	1.748	-5.160	10.710
Rank	5.179	7.112	1	64
Lag CTR	0.007	0.053	0	1
Retailer	0.057	0.232	0	1
Brand	0.398	0.490	0	1

4. RESULTS

Next, we discuss our empirical findings. We first discuss the effects of various keyword characteristics and keyword ranking on click-through rates of the sponsored search advertisements. The coefficient of *Retailer*, α_1, is positive and significant indicating that keyword advertisements that contain retailer-specific information lead to a significant increase in click-through rates. Specifically, the magnitude of the various estimates suggests that the presence of retailer information in the keyword increases click-through rates by 28.31%. This result is useful for managers because it confirms that keyword advertisements that explicitly contain information identifying the advertiser lead to higher click-through rates than other kinds of keywords which lack such information.

Perhaps surprisingly, we find that the presence of a brand name in the search keyword (either a product-specific brand or a manufacturer-specific brand) has no statistically significant effect on click-through rates although it does affect the conversion rates.

Table 2a: Coefficient Estimates on Click-through Rates[3]

	Intercept	*Retailer*	*Brand*	*Length*
	$\overline{\beta_0}$	α_1	α_2	α_3
Intercept	**-2.062**	**2.031**	-0.105	**-0.109**
	(0.050)	**(0.155)**	(0.090)	**(0.049)**
	$\overline{\beta_1}$	γ_1	γ_2	γ_3
Rank	**-0.251**	**-0.251**	**-0.056**	-0.002
	(0.013)	**(0.061)**	**(0.022)**	(0.014)

Table 2b: Estimated Unobserved Heterogeneity Matrix in the Click-through Rate Model

	β_{i0} (Intercept)	β_{i1} (Rank)
β_{i0} (Intercept)	**0.905**	**-0.085**
	(0.077)	**(0.013)**
β_{i1} (Rank)	**-0.085**	**0.031**
	(0.013)	**(0.003)**

On the other hand, the coefficient of *Length* is negative suggesting that longer keywords typically tend to experience lower click-through rates. Specifically, we find that all else equal an increase in the length of the keyword by one word decreases the click-through rates by 6.6%. Intuitively, this result has an interesting implication if one were to tie this result with those in the literature on consideration sets in marketing. A longer keyword is less frequent and typically tends to suggest a more 'directed' or 'specific' search whereas a shorter keyword is more frequent and typically suggests a more generic search. That is, the more frequent a text word is, the less information it likely carries and the larger context should be supplied to focus the search [12]. This implies that the consideration set for the consumer is likely to shrink as the search term becomes 'narrower' in scope. It is well known from the behavioral marketing literature that changes in the level of coarseness with which people categorize products affect their decision-making processes, which in turn influences their purchasing behavior. In our context, one could imagine that as the search term (product) becomes more specialized, the

[3]Posterior means and posterior standard deviations (in the parenthesis) are reported, and estimates that are significant at 95% are bolded in all the tables.

probability of this particular retailer carrying that specific product in its assortment decreases. [9] shows that user involvement (goal directed mode versus surf mode) plays a crucial role in understanding the effectiveness of online banner ads. Since the consumers in our data get to see the advertisements displayed by all the retailers who are bidding for that keyword at the time of the search, the probability of a goal-directed consumer clicking on the retailer's advertisement decreases unless the retailer carries the specific product that the consumer is searching for. In contrast, a consumer who does not have a directed search (has a wide consideration set) and is in the surfing mode, is likely to click on several links before (s)he find a product that induces a purchase. Besides this explanation it is also possible that people use more words, or bigger words when they know less about what they are searching for, which would mean that longer queries and longer query terms are more common for less focused searches.

Rank has an overall negative relationship with *CTR* in Table 2a. This implies that lower the rank of the advertisement (i.e., higher the location of the sponsored ad on the computer screen), higher is the click-through rate. The position of the advertisement link on the search engine page clearly plays an important role in influencing click-through rates. This kind of primacy effect has also been seen in other empirical studies of the online world. [2] suggested a positive relationship between the serial position of a link in an email and recipients' clicks on that link. Similarly, [10] implied a positive relationship between a link's serial position and site visibility. Thus, *ceteris paribus*, website designers and online advertising managers would place their most desirable links toward the top of a web page or email and their least desirable links toward the bottom of the web page or email. [3] showed that the higher the link's placement in the results listing, the more likely a searcher is to select it. The study reports similar results with non-sponsored listings.

When we consider the interaction effect of these variables on the impact that *Rank* has on click-through rates, we find that keywords that contain retailer-specific information lead to an increase in the negative relationship between Rank and click-through rates. That is, for keywords that contain retailer-specific information, a lower rank (better placement) leads to even higher click-through rates. On the other hand, we find that the coefficient of *Length* is positive suggesting that longer keywords typically tend to moderate the negative relationship between *click-through rates* and *Rank*. In other words, as the number of words in a sponsored search advertisement increases, a given rank leads to even lower click-through rates than an advertisement with a fewer number of words.

As shown in Table 2b, the estimated unobserved heterogeneity covariance is significant including all of its elements. This suggests that the baseline click-through rates and the way that keyword ranking predicts the click-through rates are different across keywords, driven by factors beyond the three observed keyword characteristics.

Next consider Tables 3a and 3b with findings on conversion rates. Our analysis reveals that the coefficient of *Brand*, δ_2, is positive and significant indicating that keywords that contain information specific to a brand (either product-specific or manufacturer-specific) experience higher conversion rates on an average. Specifically, the presence of brand information in the keyword increases conversion rates by 21.35%. This suggests that 'branded' keywords are indeed more valuable to an advertiser than 'non-branded' ones.

In contrast neither *Length*, nor *Retailer* is statistically significant in their overall effect on conversion rates. As expected, *Rank* has a negative relationship with conversion rates. Lower the *Rank* (i.e., higher the sponsored keyword on the screen), higher is the *Conversion Rate*. Also as expected, CTR has a positive relationship with conversation rates. Higher the *CTR*, higher the *conversion rate*. To be precise, an increase in click through rate from 0 (min) to 1 (max) increases conversion by 63.31% while a decrease in the rank from the maximum possible position or worst case scenario (which is 64 in our data) to the minimum position or best case scenario (which is 1 in our data) increases conversion by 99.97%. These analyses suggest that in general, the rank of a keyword on the search engine has a much more significant impact on conversion rates than CTR.

Table 3a: Coefficient Estimates on Conversion Rates

	Intercept	*Retailer*	*Brand*	*Length*
	$\overline{\theta_0}$	δ_1	δ_2	δ_3
Intercept	**-4.812**	-0.481	**0.469**	-0.130
	(0.213)	(0.339)	**(0.138)**	(0.074)
	$\overline{\theta_1}$	κ_1	κ_2	κ_3
Rank	**-0.099**	**0.293**	0.049	0.037
	(0.031)	**(0.106)**	(0.035)	(0.031)
	$\overline{\theta_2}$			
CTR	**0.822**			
	(0.368)			

Table 3b: Estimated Unobserved Heterogeneity Matrix in the Conversion Rate Model

	θ_{i0} (Intercept)	θ_{i1} (Rank)
θ_{i0} (Intercept)	**0.503**	**-0.051**
	(0.116)	**(0.022)**
θ_{i1} (Rank)	**-0.051**	**0.067**
	(0.022)	**(0.007)**

Table 4a: Coefficient Estimates on Bid Price

	Intercept	Retailer	Brand	Length
	$\overline{\omega}_0$	λ_1	λ_2	λ_3
Intercept	**-1.285**	**-1.036**	**-0.171**	**0.095**
	(0.020)	**(0.089)**	**(0.043)**	**(0.027)**
	$\overline{\omega}_1$	ρ_{11}	ρ_{12}	ρ_{13}
LagRank	**-0.027**	**0.110**	0.013	-0.003
	(0.006)	**(0.039)**	(0.013)	(0.008)
	$\overline{\omega}_2$	ρ_{21}	ρ_{22}	ρ_{23}
LagProfit	**-0.020**	-0.049	-0.005	0.003
	(0.008)	(0.033)	(0.022)	(0.013)

Table 4b: Unobserved Heterogeneity Estimates in the Bid Price Model (Σ^{ω})

	ω_{i0} *(Intercept)*	ω_{i1} *(LagRank)*	ω_{i1} *(LagProfit)*
ω_{i0} (Intercept)	**0.255** **(0.017)**	**-0.027** **(0.004)**	0.009 (0.005)
ω_{i1} (Lag Rank)	**-0.027** **(0.004)**	**0.015** **(0.001)**	0.0005 (0.001)
ω_{i1} (Lag Profit)	0.009 (0.005)	0.0005 (0.001)	**0.029** **(0.003)**

When we consider the effect of these keyword characteristics on the impact of *Rank* on the conversion rate, we find that keywords that are specific to the retailer or to a brand do not have a statistically significant effect on the relationship between rank and conversion rates. Similarly the length of a keyword typically has no significant effect on the relationship between *Conversion Rate* and *Rank. CTR* is positively associated with *Conversion Rate*. This is in accordance with what one would expect: higher the click-through rate, higher the conversion. Recall that because we model the coefficient of *CTR,* $\overline{\theta}_2$, as a fixed effect for the empirical identification purpose, there are no coefficients for *Retailer, Brand* and *Length* in its case.

As shown in Table 3b, the estimated unobserved heterogeneity covariance is significant including all of its elements. This suggests that the baseline conversion rates and the way that keyword ranking predicts the click-through rates are different across keywords, driven by unobserved factors.

Next, we turn to advertiser's behavior. Interestingly, the analysis of bid prices in Table 4a reveals that there is a negative relationship between *Bid Price* and *Retailer* as well as between *Bid Price* and *Brand*, whereas there is a positive relationship between *Bid Price* and *Length*. This implies that the firm places lower bids for advertisements that contain retailer or brand information and higher bids for those advertisements that are narrow in scope. Further, there is a negative relationship between *Bid Price* and *Lag Rank* as well *Lag Profit*. These results are indicative of the fact that while there is some naïve learning behavior exhibited by the firm, it is certainly not bidding optimally. Note that the coefficient of *Lag Rank* is negative and significant while the coefficient of *Lag Profit* is negative and statistically significant. Also the coefficient of *Lag Rank* is more than the coefficient of *Lag Profit*. Moreover, note from Table 4b that the coefficient of *Lag Rank* is positively co-varying with the coefficient of *Lag Profit*. This implies that all else equal, if the firm exhibits a higher degree of *naive learning* then it exhibits a smaller degree of *sophisticated learning* and vice-versa.

Finally, on the analysis of *Rank*, we find that all three covariates- *Retailer, Brand* and *Length* have a statistically significant and negative relationship with *Rank*, suggesting that the search keywords that have retailer-specific information or brand-specific information or are more specific in their scope generally tend to have lower ranks (i.e., they are listed higher up on the screen).

How do search engines decide on the final rank? Anecdotal evidence and public disclosures by Google suggest that it incorporates a performance criterion along with bid price when determining the ranking of the advertisers. The advertiser in the top position might pay more per click than the advertiser in the second position, but there is no guarantee that it will be displayed in the first slot. This is because past performance such as click-through rates are factored in by Google before the final ranks are published. The coefficients of *Bid Price* and *Lag CTR* are negative and statistically significant in our data. Thus, our results from the estimation of the Rank equation confirms that the search engine is indeed incorporating both bid prices and previous click-through rates in determining the final rank of a keyword. Note from Table 5a that the coefficient of *Bid Price* is more than twice the coefficient of *Lag CTR,* suggesting that bid price has a much larger role to play in determining the final rank.

Table 5a: Coefficient Estimates on Keyword Ranks

	Intercept	Retailer	Brand	Length
	$\overline{\phi}_0$	τ_1	τ_2	τ_3
Intercept	**2.119**	**-0.636**	**-0.434**	**-0.109**
	(0.123)	**(0.152)**	**(0.076)**	**(0.044)**
	$\overline{\phi}_1$	π_1	π_2	π_3
Bid Price	**-3.025**	**1.787**	0.307	**0.455**
	(0.353)	**(0.390)**	(0.179)	**(0.124)**
	$\overline{\phi}_2$			
Lag(CTR)	**-1.328**			
	(0.080)			

Table 5b: Unobserved Heterogeneity Estimates in the Keyword Rank Model (Σ^{θ})

	$\overline{\phi_0}$ (Intercept)	$\overline{\phi_1}$ (Rank)
$\overline{\phi_0}$ (Intercept)	**1.289**	**-2.007**
	(0.072)	**(0.146)**
$\overline{\phi_1}$ (Bid Price)	**-2.007**	**3.886**
	(0.146)	**(0.334)**

Finally, it is worth noting in Table 6 that most of the unobserved covariance turns out to be statistically significant. This suggests that keyword ranking is endogenous and a firm's bids are likely to be based on the same keyword's past performance. Ignoring the endogenous relationship will lead to biased estimates on the impact of ranking on click-through and conversion rates.

Table 6: Estimated Covariance across Click-through, Conversion, Bid Price and Rank (Ω)

	Click-through	Conversion	Bid Price	Rank
Click-through	**0.461**	-0.077	**0.015**	**0.279**
	(0.038)	(0.062)	**(0.007)**	**(0.020)**
Conversion	-0.077	**0.254**	**-0.043**	-0.054
	(0.062)	**(0.045)**	**(0.019)**	(0.043)
Bid Price	**0.015**	**-0.043**	**0.170**	**-0.012**
	(0.007)	**(0.019)**	**(0.004)**	**(0.006)**
Rank	**0.279**	-0.054	**-0.012**	**0.250**
	(0.020)	(0.043)	**(0.006)**	**(0.008)**

5. RELATED WORK

Our paper is related to several streams of research. First, it contributes to recent research in online advertising in economics and marketing by providing the first known empirical analysis of sponsored search keyword advertising. Much of the existing academic (e.g., [7], [8], and [13]) on advertising in online world has focused on measuring changes in brand awareness, brand attitudes, and purchase intentions as a function of exposure. This is usually done via field surveys or laboratory experiments using individual (or cookie) level data. In contrast to other studies which measure (individual) exposure to advertising via aggregate advertising dollars ([17]), we use data on individual search keyword advertising exposure. [21] looks at online banner advertising. Because banner ads have been perceived by many consumers as being annoying, traditionally they have had a negative connotation associated with it. Moreover, it was argued that since there is considerably evidence that only a small proportion of visits translate into final purchase ([5], [8], [23]), click-through rates may be too imprecise for measuring the effectiveness of banners served to the mass market. Interestingly however, [21] found that banner advertising actually increases purchasing behavior, in contrast to conventional wisdom. These studies therefore highlight the importance of investigating the impact of other kinds of online advertising such as search keyword advertising on actual purchase behavior, since the success of keyword advertising is also based on consumer click-through rates.

A large literature in economics sees advertising as necessary to signal some form of quality ([15], [22]). There is also an emerging theoretical stream of literature exemplified by [11] that examines auction price and mechanism design in keyword auctions. Despite the emerging theory work, very little empirical work exists in online search advertising. The handful of empirical studies that exist in search engine advertising have mainly analyzed publicly available data from search engines. [1] looks at the presence of quality uncertainty and adverse selection in paid search advertising. [19] classifies queries as *informational*, *navigational*, and *transactional* based on the expected type of content destination desired and analyze click through patterns of each. They find that about 80% of Web queries are *informational* in nature, approximately 10% each being *transactional*, and *navigational*. In a paper related to our work, [26] studied the conversion rates of hotel marketing keywords to analyze the profitability of different campaign management strategies.

To summarize, our research is distinct from extant online advertising research as it has largely been limited to the influence of banner advertisements on attitudes and behavior. We contribute to the literature by empirically comparing the impact of different keyword characteristics on the performance of online search advertising in paid search listings.

6. CONCLUSIONS AND FUTURE WORK

The phenomenon of sponsored search advertising is gaining ground as the largest source of revenues for search engines. However, we have little understanding of how consumers respond to sponsored search advertising on the Internet, and how what factors drive firms' decision on bid prices and ranks. In this research, we focus on understanding how sponsored search advertising affects consumer search and purchasing patterns on the Internet. Specifically, we focus on analyzing the impact of different keyword level covariates on different metrics of sponsored search advertisement performance taking both consumer and firm behavior into account. Finally, we analyze the cross-selling potential from sponsored search advertising.

Using a unique panel dataset of several hundred keywords collected from a nationwide retailer that advertises on Google, we empirically model the relationship between different metrics such as click-through rates, conversion rates and keyword ranks. We use a Hierarchical Bayesian modeling framework and estimate the model using Markov Chain Monte Carlo (MCMC) methods. We began our research with an investigation of how keyword specific characteristics affect click-through rates, conversion rates and ranks, and found considerable differences across keywords. Since the ultimate aim of sponsored search advertisement is to increase demand, we also aim to analyze the profitability of such ads using different metrics of performance. Our data reveals that there is a considerable amount of heterogeneity in terms of the revenues that accrue from different keywords as well as significant differences in the performance metrics.

Arguably, the mix of retailer-specific and brand-specific keywords in an online advertiser's portfolio has some analogies to other kinds of marketing mix decisions faced by firms in many markets. For instance, typically it is the retailer who engages in 'retail store' advertising that has a relatively 'monopolistic' market. In contrast, typically it is the manufacturer who engages in advertising 'national-brands'. From the retailer's perspective, these advertisements are likely to be relatively more 'competitive' since national brands are likely to be stocked by its competitors too. Retailer-name searches are navigational searches, and are analogous to a customer finding the retailer's phone number or address in the White Pages. These searches are driven by brand awareness generated by catalog mailings, TV ads, etc, and are likely to have come from more 'loyal' consumers. Even though the referral to the retailer's website came through a search engine, the search engine had very little to do with generating the demand in the first place. On the other hand, searches on product or manufacturer specific brand names are analogous to consumers going to the Yellow Pages—they know they need a product or service, but don't yet know where to buy it (Kaufman 2007). These are likely to be "competitive" searches. Even for loyal buyers, a "branded" search means the searcher is surveying the market and is vulnerable to competition. If the advertiser wins the click and the order, that implies they have taken market share away from a competitor. Thus, retailer-specific keywords are likely to be searched and clicked by 'loyal' consumers who are inclined towards buying from that retailer whereas brand-specific keywords are likely to be searched and clicked by the 'shoppers or searchers' who can easily switch to competition.

Most firms who sponsor online keyword advertisements set a daily budget, select a set of keywords, determine a bid price for each keyword, and designate an ad associated with each selected keyword. If the company's spending has exceeded its daily budget, however, its ads will not be displayed. With millions of available keywords and a highly uncertain click-through rate associated with the ad for each keyword, identifying the most profitable set of keywords given the daily budget constraint becomes challenging for companies wishing to promote their goods and services via search-based advertising [25]. In this regard, our analysis reveals that while retailer-specific information is more important than brand-specific information in predicting click-through rates, the opposite holds true in predicting conversion rates. Sponsored advertisements that contain retailer or brand information, or are more specific in their scope generally tend to have lower ranks (i.e., they are listed higher up on the screen). Since the search engine accounts for both bid price and previous click-through rates in deciding on the final rank, these results can have useful implications for a firm's Internet paid search advertising strategy by shedding light on what the most "attractive" keywords from a firm's perspective are, and how it should optimally bid in search engine advertising campaigns. The analysis of these keyword attributes on conversion rates also provide insights into what kind of keyword advertisers should bid on in the event that search engines migrate from a pay-per-click model to a pay-per-action model as Google has recently claimed it will do.

We are cognizant of the limitations of our paper. These limitations arise primarily from the lack of information in our data. For example, we do not have data on competition. That is, we do not know the keyword auction ranks or other performance metrics such as click-through rates and conversion rates of the keyword advertisements of the competitors of the firm whose data we have used in this paper. Future research can use data on competition and highlight some more insights on how firms should manage a paid search campaign. Further, we do not have any knowledge of the other marketing variables such as any promotions during consumers' search and purchase visits. Future work can investigate if the rank preference linear across all of the advertising properties on the page to see if people are likely to click any position at the top than any position at the side. We hope that this study will generate further interest in exploring this important emerging area in web search.

7. ACKNOWLEDGMENTS

The authors would like to thank the anonymous company that provided data for this study. This work was partially supported by a grant from the NET Institute. Anindya Ghose acknowledges the generous financial support of the National Science Foundation through CAREER Award IIS-0643847. The usual disclaimer applies.

8. REFERENCES

[1] Animesh A., Ramachandran, V., and Viswanathan, S. Quality uncertainty and adverse selection in sponsored search markets. *Proceedings of the 2006 NET Institute Conference*.

[2] Ansari, A., and Mela, C. E-customization. *Journal of Marketing Research, 40* (2), 2003, 131-145.

[3] Broder, A. Taxonomy of web search, *SIGIR Forum,* vol. 36, pp. 3-10, 2002.

[4] Brooks, N. 2004. The Atlas Rank Report II: How search engine rank impacts conversions. Atlas Institute, 2004.

[5] Chatterjee, P., Hoffman, D., and Novak, T. Modeling the clickstream: implications for web-based advertising efforts. *Marketing Science*: 22(4), 2003, 520-541.

[6] Chib, S., and Greenberg. E. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49, 1995, 327-335.

[7] Cho, C., Lee, J., and Tharp, M. Different forced-exposure levels to banner advertisements, *Journal of Advertising Research*, 41(4), 2001, 45-56.

[8] Dahlen, M. Banner advertisements through a new lens, *Journal of Advertising Research*, 41(4), 2001, 23-30.

[9] Danaher, P. and Mullarkey, G. 2003. Factors affecting online advertising recall: A Study of Students, *Journal of Advertising Research*, September, 2003, 252-267.

[10] Drèze, X., and Zufryden, F. 2004. The measurement of online visibility and its impact on Internet traffic. *Journal of Interactive Marketing, 18* (1), 20-37.

[11] Edelman, B., Ostrovsky, M., and Schwarz, M. Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords. *American Economic Review*, 97(1), 2007, 242-259.

[12] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman G., and Ruppin, E. Placing search in context: the concept revisited. *Proceedings of the WWW Conference,* 2001.

[13] Gallagher, K., Foster, K., and Parsons, J. The medium is not the message: advertising effectiveness and content evaluation in print and on the web. *Journal of Advertising Research*, 41(4), 2001, 57-70.

[14] Greenspan, R. Searching for balance. vol. 2004: *ClickZ stats.*

[15] Grossman, G., and Shapiro, C. Informative advertising with differentiated products. *Review of Economic Studies.* 51(1), 1984, 63-81.

[16] Hotchkiss, G., Garrison, M., and Jensen, S. Search engine usage in North America. *Enquiro,* 2005.

[17] Ilfeld, J., and Winer, R. Generating website traffic. *Journal of Advertising Research*, 42, 2002, 49-61.

[18] iProspect Inc. Search engine user attitudes, 2005.

[19] Jansen, B., and Spink, A. The effect on click-through of combining sponsored and non-sponsored search engine results in a single listing, *Proceedings of the 2007 Workshop on Sponsored Search Auctions, WWW Conference,* 2007.

[20] Kilpatrick, D. Keyword advertising on a cost-per-click model, www.sitepronews.com/archives/2003/feb/14prt.html.

[21] Manchanda, P., Dubé, J., Goh, K. and Chintagunta, P. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, *43*(1), 2006, 98-108.

[22] Milgrom, P., and Roberts, J. Price and advertising signals of product quality, *Journal of Political Economy*, 94, 1986, 796-821.

[23] Moe, W., and Fader, P. Dynamic conversion behavior at e-commerce sites. *Management Science*, 50(3), 2003, 326-335.

[24] Rossi, P. E., and Allenby, G. Bayesian statistics and marketing. *Marketing Science*, 22, 2003, 304-329.

[25] Rusmevichientong, P., and Williamson, D. An adaptive algorithm for selecting profitable keywords for search-based advertising services, Working Paper, 2006.

[26] Rutz, O., R. E. Bucklin 2007. A model of individual keyword performance in paid search advertising, Unpublished Mimeo, UCLA.