Comparing Performance Metrics in Organic Search with Sponsored Search 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

With the rapid growth of search advertising, there has been an increased interest amongst both practitioners and academics in enhancing our understanding of how consumers respond to contextual and sponsored search advertising on the Internet. An emerging stream of work has begun to explore these issues. In this paper, we focus on a previously unexplored question: How does sponsored search advertising compare to organic listings with respect to predicting conversion rates, order values and profits from a keyword ad? We use a Hierarchical Bayesian modeling framework and estimate the model using Markov Chain Monte Carlo (MCMC) methods. Our analysis suggests that on an average while the conversion rates, order values and profits from paid search advertisements were much higher than those from natural search, most of the keyword-level characteristics have a statistically significant and stronger impact on these three performance metrics for organic search than paid search. This could shed light on understanding what the most "attractive" keywords are from advertisers' perspective, and how advertisers should invest in search engine advertising campaigns relative to search engine optimization.

Categories and Subject Descriptors

J.4 [Social and Behavioral Sciences]: Economics

General Terms

Performance, Measurement, Economics.

Keywords

Organic Search, Hierarchical Bayesian modeling, Paid search advertising, Electronic commerce, Internet Economics

ADKDD 08, August 24, 2008, Las Vegas, Nevada, USA.

1. INTRODUCTION

The advertising world has changed dramatically in the past decade. In the pre-Internet era, firms relied heavily upon traditional media advertising like television, magazines, direct mail, and even radio. But today, marketers have embraced the Internet with search engine marketing, social media networks, interactive websites, etc. In fact, in the past year alone, online advertising expenditures grew 26% to total $21.4 billion. Though there are many innovative ways firms can advertise online, the bulk of online advertising consists of two main forms: display ad (banner) advertising and paid search advertising (sponsored ads that appear on the search results pages of search engines). Since consumers perceive display ads as annoying and obtrusive, they represent a small proportion of online advertising. Conversely, paid search advertising represents 40% of online advertising expenditures, and has grown 32% in the past year alone. What has fueled this growth? 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..

As companies are showing more willingness to advertise on the internet, a recent survey conducted by McKinsey indicates that marketing executives still worry over the lack of metrics.[1] In the past, marketers sought to increase the number of page views for their website. Now, these executives want more concrete metrics which relate more directly to profitability. Currently, search engines offer the most measurable form of advertising metrics; they can provide estimates on click-through rates and average bid prices for every possible keyword. While managers recognize the importance of paid advertising, many companies

[1] Green, H. "Stumbling Blocks for Online Advertising." BusinessWeekOnline, September 2007.

have also begun investing heavily in search engine optimization (SEO) to improve their organic search results, either in addition or in lieu of search engine marketing (SEM). SEO refers to the process of tailoring a web site to optimize its unpaid (or "organic") ranking for a given set of keywords or phrases. SEM refers to investments in paid (or "sponsored") rankings. In 2007, search engine optimization accounted for 18% of all search engine marketing expenditures and is expected to grow as SEO is generally less expensive than paid search.[2] A survey conducted by the eMarketer revealed that 46% of online retailers found that SEO performed best, compared to 37% of retailers who preferred paid-per-click advertising.

Nevertheless, a question that interests many firms is which keywords will give them the best return-on-investment (ROI). For paid search, managers seek to find keywords that will result in high click-through rates and more importantly, higher conversion rates. In organic search, this is the same, but since search engine optimization depends on keyword type, firms' marketing dollars could also be tailored to focus on searches with a high rate of conversion.

Despite the growth of search advertising, our understanding of how consumers respond to sponsored search advertising on the Internet is still nascent. In this paper, we focus on a previously unexplored question: *How does the content of a keyword impact sponsored search versus natural search listings with respect to predicting conversion rates, order values and profits?* While an emerging stream of literature in sponsored search has looked at issues such as the impact of keyword attributes on sponsored search and spillover effects from keyword campaigns, no prior work has empirically analyzed this question. 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. This is particularly true for companies trying to decide between making investments in SEO versus investments in SEM. There is a growing debate on which of these two search mechanisms is more effective. On the one hand, because an advertiser can control the message of a paid search, one would expect higher conversions. On the other side, because people value the 'editorial integrity' of organic searches, one would expect higher conversions from them. Since firms are now trying to grapple with the trade-offs in each of these two forms of referrals, empirical research based on actual data from an advertiser can shed some light on these issues.

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 decision. 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 purchase propensities. This classification is motivated by prior work on the goals for users' web search such as [5, 19].

[2]Hallerman, D. "Search Engine Marketing: User and Spending Trends." eMarketer. January 2008.

We find that while the mean conversion rate, mean order value and mean profit from paid search advertisements was much higher than that from a corresponding set of natural search listings available during the same time period, the various keyword level covariates have a stronger impact on natural search than on paid search.[3] In particular, the presence of *retailer-specific* information increases the *Conversion* rate, the *Order Value* and the *Profit* in both forms of search advertising – paid and natural. In contrast, while the presence of a *brand name* increases *Conversion* rates, *Order Value* and *Profit* in natural search, it does not affect any of these performance metrics in paid search. Finally, the *length* of a keyword negatively impacts the performance on all three metrics for natural search listings but only affects the *Order Value* in paid search.

2. DATA

2.1 Data Description

We first describe the data generation process for paid keyword advertisement since it differs on many dimensions from traditional offline advertisement. 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. ***A keyword may consist of one or more 'words'.*** 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. Once the advertiser gets a rank allotted for its keyword ad, these sponsored ads are displayed on the top left, and right of the computer screen in response to a query that a consumer types on the search engine. The 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. This sponsored ad shows up next to the organic search results that would otherwise be returned using a separate criteria employed by the search engine. The serving of the ad in response to a query for a certain keyword is 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 data used in this study is similar to that used in ([11]). It contains weekly information on paid search advertising from a large nationwide retail chain, which advertises on Google. 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).

[3]Order Value refers to the price of the product that was sold during the transaction.

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), 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 data is aggregated at a weekly level.

Similar to the data on paid search results, our dataset has information that consists of conversions, order value and total revenues accruing from natural searches for the same retailer during the same time period. We compare the set of keyword advertisements across the 13-week period that appears in both the paid and natural listings. There are 776 unique keyword listings in the dataset given to us by the advertiser. However, not all keywords are sponsored by this advertiser in all the weeks in our sample. Similarly, there are certain weeks where the advertiser's link did not show up in the natural listings of Google in response to the user-generated query. Hence, we have a different number of observations for clicks and conversions from paid ads in comparison to the number of observations for clicks and conversions from the natural listings for the same product sold by the advertiser. Our mapping yielded a total of 2065 observations from the paid searches, and a total of 12382 observations from the natural searches. Table 1 reports the summary statistics. Interestingly, we note that the mean conversion rate was 5.4% and 2.76% from paid and natural searches, respectively. Similarly, the mean order value and profit from paid search advertisements was much higher than that from natural search listings.

There are three important keyword specific characteristics for a firm (the advertiser) when it advertises on a search engine ([11]). This includes whether the keyword should have (i) firm-specific information, (ii) brand-specific information, (iii) and the length (in 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 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 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 (and 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.

Table 1: Summary Statistics of the Paid and Natural Matched Data (N_Paid=2065; N_Natural=12382)

Variable	Mean	Std. Dev.	Min	Max
Paid_Retailer	0.131	0.337	0	1
Paid_Brand	0.599	0.49	0	1
Paid_Length	2.42	0.81	1	5
Paid_Impressions	919.91	3342.23	1	97424
Paid_Clicks	79.1	818.15	0	33330
Paid_Conversion Rate	0.054	0.213	0	1
Log(Paid_Order Value)	1.176	1.945	0	7.13
Log(Paid_Revenue)	1.37	2.32	0	10.73
Log(Paid_Profit)	0.667	2.77	-4.92	10.71
Natural_Retailer	0.394	0.49	0	1
Natural_Brand	0.603	0.465	0	1
Natural_Length	2.16	1.03	1	5
Natural_Clicks	51.58	776.04	1	36308
Log(Natural_Order Value)	0.37	1.23.	0	0.675
Natural_Conversion Rate	0.0276	0.15	0	1
Log(Natural_Profit)	0.433	1.48	0	10.35

3. EMPIRICAL MODEL: COMPARING PERFORMANCE METRICS IN PAID AND ORGANIC SEARCH

An important determinant of the effectiveness of sponsored search advertising is the extent to which the same keyword also appears in the natural or organic listings of the search engine. Organic rankings are determined by the content of the website and the website's "relative importance". In organic search there is no guarantee as to specific ranking positions or the timing for rankings to appear/change. In order to improve rankings a firm almost always requires changes to website content and/or structure. [17] conducted a survey with 425 respondents, wherein more than 77% of participants favored non-sponsored links more than the sponsored links, as offering sources of trusted, unbiased information. Based on a survey of 1,649 Web users, [15] found that 60% of Google users reported non-sponsored results to be more relevant than sponsored. This was even higher for predominantly Google users (70%). [21] investigated the relevance of sponsored and non-sponsored links for e-commerce queries on the major search engines, and found that average relevance ratings for sponsored and non-sponsored links are practically the same, although the relevance ratings for sponsored links are statistically higher. These studies then beget the question that if natural searches lead to more purchases than sponsored ads, then to what extent should firms invest in sponsored search advertisements.

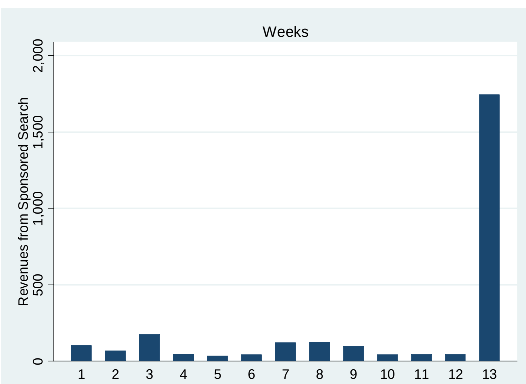

Figure 1a: Distribution of Revenues from Paid Search Across Weeks

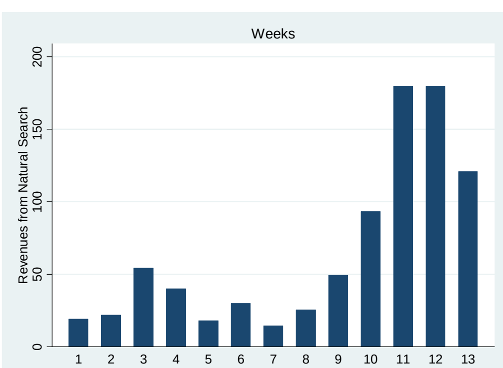

Figure 1b: Distribution of Revenues from Natural Search Across Weeks

Figures 1a and 1b show that there are considerable differences in the revenues accruing from paid and natural search over time. In this section, we intend to compare the impact of the three keyword covariates on the performance of paid vs. natural searches. More specifically, we compare the impact of the three covariates on conversion rates, order value and profit accruing from paid (sponsored) search to those from natural (organic) searches. The study of these three metrics enables us to get better insights into the factors that drive product sales and profitability for retailers in the search engine advertising industry.

3.1 Modeling Conversion

We cast our model in a Hierarchical Bayesian (HB) framework and estimate it using Markov chain Monte Carlo methods (see [28] for a detailed review of such models). In HB models, probability distributions are used to quantify prior beliefs about the parameters which are updated with the information from the data to yield a posterior distribution. The HB model is called "hierarchical" because it has two levels. At the higher level, we assume that individuals' parameters (betas) are described by a multivariate normal distribution. Such a distribution is characterized by a vector of means and a matrix of covariances. At the lower level we assume that, given an individual's betas, his/her probabilities of achieving some outcome (choosing products, or rating brands in a certain way) is governed by a particular model, such as multinomial logit or linear regression [28].

Recent advances in computational techniques such as MCMC methods have proven to be very useful in estimating such models. Rather than deriving the analytic form of the posterior distribution, MCMC methods substitute a set of repetitive calculations that, in effect, simulate draws from this distribution. These Monte Carlo draws are then used to calculate statistics of interest such as parameter estimates and confidence intervals. The idea behind the MCMC engine that drives the HB revolution is to set up a Markov chain that generates draws from posterior distribution of the model parameters [28]. An advantage of estimating hierarchical Bayes (HB) models with Markov chain Monte Carlo (MCMC) methods is that it yields estimates of all model parameters, including estimates of model parameters associated with specific respondents (which in our case translates into keywords). We use the Metropolis-Hastings algorithm with a random walk chain to generate such draws ([6]).

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). Among the n_{ij} click-throughs of paid searches, there are m_{ij} click-throughs that lead to purchases for keywords i at week j. Let us further assume that the probability of making a purchase is q_{ij}. Then, the likelihood of the number of purchases is specified as:

$$f(m_{ij}^P) = \frac{n_{ij}^P!}{m_{ij}^P!(n_{ij}^P - m_{ij}^P)!} q_{ij}^{P\, m_{ij}^P} (1 - q_{ij}^P)^{n_{ij}^P - m_{ij}^P} \tag{1.1}$$

Note that the superscript P stands for paid searches, and the superscript N stands for natural searches. Similarly, for natural searches, the likelihood of the number of purchases is specified as:

$$f(m_{ij}^N) = \frac{n_{ij}^N!}{m_{ij}^N!(n_{ij}^N - m_{ij}^N)!} q_{ij}^{N\, m_{ij}^N} (1-q_{ij}^P)^{n_{ij}^N - m_{ij}^N} \tag{1.2}$$

Next we derive the conversion probabilities in paid and organic searches. Different keywords are associated with different products. Since product-specific characteristics can influence consumer conversion rates, it is important to control for unobserved product characteristics that may influence conversion rates once the consumer is on the website of the advertiser. Hence, we include the three keyword characteristics to proxy for the unobserved keyword heterogeneity stemming from the different products sold by the advertiser. This leads us to model the conversion probabilities as follows:

$$q_{ij}^P = \frac{\exp(\alpha_i + \beta_1 Retailer_i + \beta_2 Brand_i + \beta_3 Length_i + \varepsilon_{ij})}{1+\exp(\alpha_i + \beta_1 Retailer_i + \beta_2 Brand_i + \beta_3 Length_i + \varepsilon_{ij})} \tag{1.3}$$

$$q_{ij}^N = \frac{\exp(\gamma_i + \delta_1 Retailer_i + \delta_2 Brand_i + \delta_3 Length_i + \eta_{ij})}{1+\exp(\gamma_i + \delta_1 Retailer_i + \delta_2 Brand_i + \delta_3 Length_i + \eta_{ij})} \tag{1.4}$$

To complete the specification, we have

$$\varepsilon_{ij} \sim N(0, \sigma_{P,C}^2) \tag{1.5}$$

$$\eta_{ij} \sim N(0, \sigma_{N,C}^2) \tag{1.6}$$

$$\theta_i = (\alpha_i, \gamma_i)' \sim MVN(\overline{\theta^C}, \Sigma^C) \tag{1.7}$$

When specifying the distribution of the intercept and the error terms, we use *N* to denote a normal distribution and *MVN* to denote a multivariate normal distribution.

3.2 Modeling Order Value

Note that the order value (the price of the product) is always positive. This implies that the data will be left censored. In censored data, it is well known that the use of simple OLS regressions leads to inconsistent estimates [33]. Hence, we use a Tobit specification to model the order values.[4]

We assume there is a latent spending intention ($z_{ij}^{P,Order}$) of a consumer that determines how much to spend for keyword i at order j through a paid advertisement. Hence, we have

$$y_{ij}^{P,Order} = z_{ij}^{P,Order} \quad \text{if } z_{ij}^{P,Order} > 0 \tag{1.8}$$

$$y_{ij}^{P,Order} = 0 \quad \text{if } z_{ij}^{P,Order} \leq 0 \tag{1.9}$$

Similarly, for natural searches, we have

$$y_{ij}^{N,Order} = z_{ij}^{N,Order} \quad \text{if } z_{ij}^{N,Order} > 0 \tag{1.10}$$

$$y_{ij}^{N,Order} = 0 \quad \text{if } z_{ij}^{N,Order} \leq 0 \tag{1.11}$$

We model the latent buying intention of consumers from paid advertisements and natural listings, respectively, as follows:

$$z_{ij}^{P,Order} = \alpha_i^{P,Order} + \beta_1^{P,Order}\mathrm{R}etailer_i + \beta_2^{P,Order} Brand_i + \beta_3^{P,Order} Length_i + \varepsilon_{ij}^{P,Order} \tag{1.12}$$

$$z_{ij}^{N,Order} = \alpha_i^{N,Order} + \beta_1^{N,Order}\mathrm{R}etailer_i + \beta_2^{N,Order} Brand_i + \beta_3^{N,Order} Length_i + \varepsilon_{ij}^{N,Order} \tag{1.13}$$

To complete the model specification, we have

$$\varepsilon_{ij}^{P,Order} \sim N(0, \sigma_{P,Order}^2) \tag{1.14}$$

$$\varepsilon_{ij}^{N,Order} \sim N(0, \sigma_{N,Order}^2) \tag{1.15}$$

$$(\alpha_i^{P,Order}, \alpha_i^{N,Order})' \sim MVN(\overline{\alpha^{Order}}, \Sigma^{Order}) \tag{1.16}$$

3.3 Modeling Profit

Note that *Profit* can have both negative and positive values because the total revenues from an advertisement may be less than the total costs incurred for displaying that paid advertisement. Hence, we can use an ordinary least squares (OLS) regression to model the paid profit. We model the profit of the paid searches in the form of the following regressions:

[4]The Tobit model is an econometric method that describes the relationship between a non-negative dependent variable y_i and an independent variable (or vector) x_i. The model supposes that there is a latent (i.e. unobservable) variable y_i . This variable linearly depends on x_i via a parameter (vector) β which determines the relationship between the independent variable (or vector) x_i and the latent variable y_i (just as in a linear model). In addition, there is a normally distributed error term u_i to capture random influences on this relationship. The observable variable y_i is defined to be equal to the latent variable whenever the latent variable is above zero and zero otherwise. If the relationship parameter β is estimated by regressing the observed y_i on x_i, the resulting ordinary least squares estimator is inconsistent. [2] has proven that the likelihood estimator for this model is consistent.

$$y_{ij}^{P,\text{Profit}} = \alpha_i^{P,\text{Profit}} + \beta_1^{P,\text{Profit}} \text{Retailer}_i + \beta_2^{P,\text{Profitr}} Brand_i + \beta_3^{P,\text{Profit}} Length_i + \varepsilon_{ij}^{P,\text{Profit}} \quad (1.17)$$

Note that the profit in natural searches is always positive. This is because there are no direct advertising costs involved for the retailer for selling through natural listings, and hence profits are simply equal to revenues in this case. This implies that the data on profits from natural searches will be left censored. Hence, we use a Tobit specification to model the profit of the natural searches as follows:

$$y_{ij}^{N,\text{Profit}} = z_{ij}^{N,\text{Profit}} \qquad \text{if } z_{ij}^{N,\text{Profit}} > 0 \quad (1.18)$$

$$y_{ij}^{N,\text{Profit}} = 0 \qquad \text{if } z_{ij}^{N,\text{Profit}} \le 0 \quad (1.19)$$

As before, we model the latent buying intention from natural listings as follows:

$$z_{ij}^{N,\text{Profit}} = \alpha_i^{N,\text{Profitr}} + \beta_1^{N,\text{Profit}} \text{Retailer}_i + \beta_2^{N,\text{Profit}} Brand_i + \beta_3^{N,\text{Profit}} Length_i + \varepsilon_{ij}^{N,\text{Profit}} \quad (1.20)$$

To complete the model specification, we have the following:

$$\varepsilon_{ij}^{P,\text{Profit}} \sim N(0, \sigma^2_{P,\text{Profit}}) \quad (1.21)$$

$$\varepsilon_{ij}^{N,\text{Profit}} \sim N(0, \sigma^2_{N,\text{Profit}}) \quad (1.22)$$

$$(\alpha_i^{P,Profit}, \alpha_i^{N,Profit})' \sim MVN(\overline{\alpha^{Profit}}, \Sigma^{Profit}) \quad (1.23)$$

3.4 Results

We now examine the effect of keyword covariates at the mean level (see Tables 2a, 2b and 2c). The overall pattern of the results indicates that the presence of *retailer name*, *brand name* and the *length* of the keyword are associated with the decision to purchase, the amount of purchase and the firm's overall profit in any given week. Specifically, the presence of *retailer-specific* information is associated with an increase in the *Conversion* rate, the *Order Value* and the *Profit* in both forms of search advertising – paid and natural. In contrast, while the presence of a *brand name* is associated with an increase in *Conversion* rates, *Order Value* and *Profit* in natural search, it impact on any of these performance metrics in paid search is not statistically significant. Finally, the *length* of a keyword is negatively associated with an increase in the performance on all three metrics for natural search listings. In the case of paid search, the impact of the keyword length has a statistically significant and negative impact only on the *Order Value*. Thus, we see that longer keywords generally tend to have a detrimental affect on keyword performance such as conversion rates and profits.

Table 2a: Coefficient Estimates for Predicting Conversion[5]

	Intercept	*Retailer*	*Brand*	*Length*	σ^2	Σ^C
Paid	**-5.124**	**2.465**	0.228	-0.074	**4.301**	**1.245**
	(0.301)	**(0.259)**	(0.24)	(0.105)	**(0.314)**	**(0.313)**
Natural	**-7.370**	**0.562**	**0.488**	**-0.192**	**11.314**	**0.572**
	(0.537)	**(0.179)**	**(0.21)**	**(0.093)**	**(1.826)**	**(0.228)**

Table 2b: Coefficient Estimates for Predicting Order Value

	Intercept	*Retailer*	*Brand*	*Length*	σ^2	Σ^{Order}
Paid	**-2.997**	**3.803**	0.416	**-0.593**	**15.22**	**13.079**
	(0.78)	**(0.63)**	(0.47)	**(0.27)**	**(1.13)**	**(1.91)**
Natural	**-13.08**	**3.681**	**1.775**	**-1.154**	**48.16**	**11.260**
	(0.97)	**(0.69)**	**(0.63)**	**(0.29)**	**(2.7)**	**(2.14)**

Table 2c: Coefficient Estimates for Predicting Profit

	Intercept	*Retailer*	*Brand*	*Length*	σ^2	Σ^{Profit}
Paid	0.505	**2.144**	0.207	-0.177	**5.192**	**1.559**
	(0.276)	**(0.238)**	(0.167)	(0.101)	**(0.176)**	**(0.167)**
Natural	**-15.352**	**4.325**	**2.042**	**-1.344**	**66.049**	**15.059**
	(1.123)	**(0.799)**	**(0.737)**	**(0.339)**	**(3.675)**	**(2.831)**

How do these estimates translate into actual percentage changes? In 'Paid' search, the presence of retailer information in the keyword increases conversion rates by 131 %, an increase in length of the keyword by one word decreases order value by 7.7 % while the presence of retailer information in the keyword increases profit by 5.2 %. In 'Natural' search, the presence of

[5]Posterior means and posterior standard deviations (in the parenthesis) are reported, and estimates that are significant at 95% are bolded in Tables 2a -2c.

retailer information in the keyword increases conversion rates by 29.74 %, the presence of brand information in the keyword increases conversion rates by 42.93 %, and an increase in length of the keyword by 1 word decreases conversion rate by 5.41 %. In 'Natural' search, the presence of retailer information in the keyword increases order value by 67.61 %, the presence of brand information in the keyword increases order value by 45.19 % and an increase in length of the keyword by 1 word decreases order value by 20.01%. In 'Natural' search, the presence of retailer information in the keyword increases profit by 68.32 %, the presence of brand information in the keyword increases profit by 44.26 % and an increase in length of the keyword by 1 word decreases profit by 10.71 %. These results are summarized in Table 3 below.

Table 3: Summary of Percentage Effects of Keyword Covariates Based on Estimates from Tables 2a-2c. [6]

	Retailer	*Brand*	*Length*
Paid_Conversion Rate	131%	NA	NA
Paid_Order Value	70.4%	NA	-7.7%
Paid_Profit	52%	NA	NA
Natural_Conversion Rate	29.74%	42.93%	-5.45%
Natural_Order Value	67.61%	45.19%	-20.01%
Natural_Profit	68.32%	44.26%	-10.71%

To analyze whether the differences in the impact of different covariates on the performance metrics between 'Paid' and 'Natural' searches were statistically significant, we conducted pairwise t-tests. The analyses reveals that the presence of retailer information is associated with a bigger impact on paid search advertisements than natural search listings in predicting conversion rates. However, we cannot say anything conclusively about either the differential impact of retailer information or the impact of keyword length in predicting average order values between paid and natural searches.

4. 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]) 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 ([18]), we use data on individual search keyword advertising exposure. [24] 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 ([27]), click-through rates may be too imprecise for measuring the effectiveness of banners served to the mass market. Interestingly however, [24] found that banner advertising actually increases purchasing behavior, in contrast to conventional wisdom. A large literature in economics sees advertising as necessary to signal some form of quality ([16], [26]). There is also an emerging theoretical stream of literature exemplified by ([3] [8], [22], and [32]) that examines auction price and mechanism design in sponsored keyword auctions.

Despite the emerging theory work, very little empirical work exists in online search advertising that looks at conversions and profits. This is primarily because of difficulty for researchers to obtain such advertiser-level data. [5, 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*. [20, 21] investigate the relevance of sponsored and non-sponsored links for e-commerce queries on the major search engines. Other empirical work has so far focused on search engine performance ([4], [31]). Moreover, the handful of studies that exist in search engine marketing have typically analyzed publicly available data from search engines. [1] look at the presence of quality uncertainty and adverse selection in paid search advertising on search engines. [14] examine the factors that drive variation in prices for advertising legal services on Google. [30] studied the conversion rates of hotel marketing keywords to analyze the profitability of different campaign management strategies. Our previous work ([11], [12]) has analyzed the impact of different keyword covariates on sponsored search, and estimated the cross-selling potential from a keyword. In a related paper, [13] we estimate the inter-dependence between natural search listings and paid search advertisements and vice-versa, and conduct policy simulations to investigate if these two processes have a complementary or substitutive effect on each other's click-through rates.

However, none of these studies compared the performance of sponsored search with natural search by examining the impact of keyword content on performance metrics like conversion rates, order value and profit. 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, and to studying the performance of sponsored search advertisements. The domain of natural search listings has largely been ignored. We contribute to the literature by empirically comparing various performance metrics in sponsored search with natural search listings by estimating the impact of different keyword characteristics on paid and natural search listings.

[6] Percentage effects for statistically insignificant estimates in Tables 2a-2c are not computed and listed as NA.

5. CONCLUSIONS AND FUTURE WORK

In most search-based advertising services, a company sets a daily budget, selects a set of keywords, determines a bid price for each keyword, and designates 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 ([29]). In this regard, our analysis reveals that while retailer-specific information is more important than brand information in predicting conversion rates in both paid and organic search. This result can have useful implications for a firm's Internet paid search advertising strategy.

Our results can have implications on the issues related to search engine optimization (SEO) vs. search engine marketing (SEM) in particular because many advertisers engage in both kinds of activity. Our analysis suggests that most of the keyword-level characteristics have a stronger impact on the performance of natural search than paid search. This could shed light on understanding what the most "attractive" keywords from a firm's perspective are, and how it should invest in search engine advertising campaigns relative to search engine optimization.

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 precise data on competition since our data is limited to one firm and one industry. That is, we do not know the keyword ranks or other performance metrics of 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 by running more detailed policy simulations that incorporate competitive bid prices. Further, we do not have any knowledge of the other marketing variables such as any promotions during consumers' search and purchase visits. We also do not have data on the textual content in the copy of the ad and detailed content in the landing pages corresponding to the different keywords, although some evidence suggests that the presence of the keyword in the title of the ad is more important than that in the ad copy in influencing click-through rates ([25]). Future researchers can conduct various sorts of experiments to examine how the content of the ad copy interacts with keyword attributes to determine both consumer and firm behavior. We hope that this study will generate further interest in exploring this important emerging area in web search.

REFERENCES

[1] Animesh A., V. Ramachandran, S. Viswanathan. Quality uncertainty and adverse selection in sponsored search markets, *Working Paper*, University of Maryland, College Park, 2006.

[2] Amemiya, T. Regression analysis when the dependent variable is truncated normal. *Econometrica* 41 (6), 997–1016, 1973.

[3] Aggarwal, G., J. Feldman, and S. Muthukrishnan. Bidding to the Top: VCG and Equilibria of Position-Based Auction, arXiv: cs.GT/0607117 v1, 2006.

[4] Bradlow, E.T., and Schmittlein, D.C. The little engines that could: Modeling the performance of world wide web search engines, *Marketing Science*, 19(1), 43- 62, 2000.

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

[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] 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.

[9] Feng, J, Bhargava, H., and Pennock, D. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *Informs Journal on Computing*, 19(1), 2007, 137-148.

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

[11] Ghose, A., and Yang, S. An Empirical Analysis of Sponsored Search Performance in Search Engine Advertising. *Proceedings of the ACM WSDM* 2008.

[12] Ghose, A., and Yang, S. Analyzing Search Engine Advertising: Firm Behavior and Cross-Selling in Electronic Markets. *Proceedings of WWW* 2008.

[13] Ghose, A., and Yang, S. Impact of Organic Listings on Paid Search Advertising: Complements, Substitutes or Neither? Working Paper, New York University.

[14] Goldfarb, A., and C. Tucker. 2007. What makes search engine advertising valuable? Understanding search term variation in legal services," Working Paper, SSRN.

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

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

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

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

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

[20] 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.

[21] Jansen, B. J. The comparative effectiveness of sponsored and non-sponsored results for web ecommerce queries. *ACM Transactions on the Web*. 1(1), 3, 2007.

[22] Katona, Z., and M. Sarvary. The race for sponsored links: A model of competition for search advertising, Working Paper, INSEAD, 2007.

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

[24] 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.

[25] MarketingExperiments.com. PPC ad copy tested. *Marketing Experiments Journal,* 2005.

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

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

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

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

[30] Rutz, O., and R. Bucklin. A model of individual keyword performance in paid search advertising, Working Paper, 2007.

[31] Telang, R., P. Boatwright, and T. Mukhopadhyay. A mixture model for Internet search engine visits, *Journal of Marketing Research,* Vol. XLI (May), 206-214, 2004.

[32] Varian, H. Position auctions. *International Journal of Industrial Organization*, 25 (6) 1163-1178, 2006.

[33] Wooldridge, J. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press, 2002