An Empirical Investigation of the Performance of Online Sponsored Search Markets

Animesh Animesh
R.H. Smith School of Business
University of Maryland,
College Park, MD 20815. USA
301-405-8587
animesh@umd.edu

Vandana Ramachandran
R.H. Smith School of Business
University of Maryland,
College Park, MD 20815. USA
301-405-8587
vramacha@rhsmith.umd.edu

Siva Viswanathan
R.H. Smith School of Business
University of Maryland,
College Park, MD 20815. USA
301-405-8587
sviswana@rhsmith.umd.edu

ABSTRACT
Online sponsored search advertising has grown to become the dominant form of online advertising with the last few years, with Yahoo! and Google being the leading market providers. While firms bid for better placement in the listing of search results on both these sponsored search markets, Yahoo! and Google have traditionally employed different mechanisms to determine the placement of the advertisements. Using data from these sponsored search markets, this study compares the relative performance of these sponsored search markets and also examines whether intervention by the search intermediary impacts the outcomes in these markets. Our preliminary analysis uncovers interesting differences in the quality-position relationships across the two markets. We find that these differences are further validated by the results of the quantile as well as non-linear regression analyses.

Categories and Subject Descriptors
H.3.3 [**Information Systems**]: Information Systems Storage and Retrieval – *Information Search and Retrieval*.

General Terms
Performance, Design, Economics, Human Factors.

Keywords
Sponsored Search, Quality Uncertainty, Online Advertising.

1. INTRODUCTION
S*ponsored search* have emerged as a viable alternative to *organic* (algorithm-based) search as well as to traditional advertising, raising several issues of interest to researchers as well as practitioners. The increasing presence of sponsored search results in search engines presents a new kind of problem in the digital realm. An inherent conflict of interest arises in sponsored search (also known as paid-placement or keyword advertising), where search intermediaries deliver information about sellers and their offerings, but are paid by those same sellers they "certify". On the one hand, while "sponsored search" can potentially bias the search results, thereby reducing the value of online search to consumers, on the other, the validity of the sponsored search model is evinced by its growing popularity as well as the new complementary markets it has spawned in the wake of its success. Given the expanding importance of sponsored search mechanisms for the emerging economic and competitive landscape online, it is important to gain a better understanding of the implications of sponsored search mechanisms for consumers, advertisers, search intermediaries, as well as for policy makers.

Sponsored search advertising is the fastest growing of all online advertising formats, accounting for more than 40% of the total online advertising dollars spent by companies in the U.S. [9]. Despite the phenomenal growth of sponsored search markets, a majority of online consumers (62%) are unaware of the distinction between *sponsored search* results and *organic search* results. Even among those who are aware of the sponsored search results, the majority believe that firms ranked higher in the sponsored search results are of higher quality than those ranked lower in the search listings. These beliefs also directly translate into a higher number of clickthroughs that firms on the top receive, compared to others lower down the listing. According to a recent eye motion study searchers scanning a listing from top to bottom were found to pay more attention on the advertisements appearing on the top of the listing [15]. In addition, as indicated by a recent study by DoubleClick [4], more than 30% of total purchases are made from sellers listed at the top of the search listings. Given that consumers visit (and purchase from) the sellers at the top of the search listings, the firms appearing on top of these search listings stand to benefit disproportionately more than their counterparts lower down the search listings. Clearly being on top of the sponsored listings is beneficial to *all* firms. However, consumers as well as search intermediaries stand to benefit only if the sellers listed on top are also of higher quality. Given consumer awareness and beliefs, sponsored search mechanisms where low quality bidders are placed at the top of the search listings can adversely affect consumer welfare and lower the utility of such markets for consumers. On the other hand, it is possible that these sponsored search markets may be self-correcting. Consequently, it is important to understanding the quality-position correlation of sellers using these markets as it has significant implications for consumer welfare as well as the future of sponsored search markets. The relationship between a

ICEC'07, August 19–22, 2007, Minneapolis, Minnesota, USA.

seller's quality[1] and its ranking in the sponsored search results can differ across product categories and also across the different online sponsored search markets. The *relative performance* of sponsored search mechanisms therefore remains an empirical question that needs to be validated, and is one that we examine in this study.

As is well known, the two most popular sponsored search markets are Yahoo! and Google. While in the case of Yahoo!, the advertiser's rank/position in the sponsored listings is determined purely based on the bid amount (a pure market mechanism), Google provides a more "regulated" market where the ranking of a seller is based on a complex score that takes into account the "effectiveness of the advertisement" (its ability to attract clicks), in addition to the bid amount. The existence of two different sponsored search markets, one with, and the other without intervention by the search intermediary, provides a unique opportunity to compare their relative performance and understand their welfare implications. Specifically, in this study we address the following questions: (1) Are sellers that are ranked higher in the sponsored search listings of higher quality than sellers lower down the listings? (2) More importantly, does this relationship between a seller's rank and its quality *differ* across products with differing degrees of quality uncertainty? (3) Finally, are these *differences* if any, consistent across the two sponsored search markets?

2. RESEARCH CONTEXT

As noted earlier, sponsored search advertising is the dominant and the fastest growing of all online advertising formats including display advertising, search advertising, sponsorship, referral, email, rich media, slotting, email, and classifieds. According to a recent survey by Search Engine Marketing Professional Organization, sponsored search was a $5.75 billion industry in 2005 and is expected to nearly double to $11.1 billion in 2010. Google, with a U.S. market share of nearly 37%, followed by Yahoo! at 30%, operate the two most popular sponsored search markets online. However interestingly, they adopt different mechanisms in ranking their sponsored search results. Firms that wish to be listed on Yahoo!'s or Google's sponsored search listings (in response to a keyword search initiated by a consumer) can bid on the keywords related to their offerings. On Yahoo!, the higher the advertiser's bid per-click in the auction, the higher the placement the advertisement receives in the listing of sponsored search results triggered by a keyword query. However in the case of Google, the position of a firm's/advertisement's listing is a function of the advertiser's bid per-click as well as its clickthrough rate (CTR), i.e. the number of clicks the advertisement gets when displayed. If a seller (advertisement) fails to generate sufficient clicks from users, it is penalized and is moved lower down the list. The differences between Yahoo! and Google's sponsored search markets therefore brings to the forefront questions relating to the effectiveness of the search intermediary's intervention (or the lack thereof) in determining the placement of paid advertisements. Our study investigates this issue by comparing the performance of the two sponsored search markets.

[1] Quality in our study refers to the quality of the seller/advertiser, which is typically a multi-dimensional construct. We use multiple measures of advertiser-quality as described later.

3. THEORETICAL BACKGROUND

As noted earlier, the nascence of online markets combined with the lack of adequate quality signals exacerbates the problem of information asymmetries between buyers and sellers. Of particular importance for online purchases is the amount of information available to consumers regarding sellers and their offerings and the uncertainty resulting from any information asymmetries. Given our focus on quality uncertainty, we seek to examine if sellers of products/product-categories characterized by varying degrees of quality uncertainty exhibit *differences* in their bidding behavior that manifests in the outcomes of sponsored search auctions.

A Framework for Product Categories: We draw upon Nelson's [11] classification of products into *Search*, *Experience* and *Credence* (SEC) goods. This classification is particularly relevant in this context as it captures the underlying uncertainty consumers face while purchasing these products. The SEC framework has been widely used in the economics and marketing literature to understand consumer search behavior as well as firms' advertising strategies [5]. According to this framework, attributes of products can be analyzed in terms of three properties – *search, experience,* and *credence* [3, 11]. *Search* goods have characteristics that are identifiable through inspection and prior to purchase. *Experience* goods, on the other hand, have features that are revealed only through consumption. *Credence* goods differ from *Experience* and *Search* goods in that consumers can never be certain of the (long-term) quality and value of *Credence* goods purchased even from ex post observations and use. Typical examples of each include books, music CDs, and television for *Search*; brokerages, cruises, and healthcare for *Experience*; and psychics, tax services and counseling for *Credence* goods. It is pertinent to note here that the boundaries between these categories can be fuzzy, and the categories are best considered to represent regions in a continuum. What is important though is that the defining characteristic underlying this segmentation — pre-purchase quality uncertainty — increases from *Search* to *Experience* to *Credence* goods, as quality becomes more expensive to judge [3].

Advertising and Seller Quality: As mentioned earlier, sponsored search mechanisms are one of the fastest growing online advertising models. Advertising is clearly one important mechanism which can serve to reduce information asymmetries and help improve the efficiency of the market [5]. Most of the existing work in advertising has focused on traditional media such as televisions, radio, magazines and newspapers. Results from analytical models suggest that advertising expenditures should be positively related to quality [6]. However, Schmalensee [14] and Comanor and Wilson [2] show that lower quality firms, under certain conditions will advertise more as compared to high quality sellers. Empirical research [for instance see, 10] examining advertising in traditional media is also inconclusive about the relationship between seller quality and advertising intensity. Our research extends this stream of research to online markets. Since online markets are characterized by higher uncertainty and risk of adverse selection, we seek to examine the relationship between the quality of online advertisers and their level of advertising as indicated by their bidding intensity for search keywords.

Although online advertising has seen impressive growth and continues to grow rapidly, research in this field is lacking. Extant research on online advertising can be broadly classified into two categories: the first, examining the impact on online advertising on consumer behavior and attitudes, and the second, identifying factors that impact the effectiveness of online advertising. Our research complements these papers by comparing two of the largest markets for online advertising.

4. DATA AND ANALYSIS

4.1 Data

We collect data from two different sponsored search mechanisms – AdWords and Overture – used respectively by Google and Yahoo!. Following the Search, Experience and Credence framework commonly used in marketing literature, we selected a total of 36 keywords representing products, twelve each in the three categories. The classification of keywords into *Search*, *Experience*, and *Credence* categories is adapted from prior research (for instance, see [5, 11]). For each of these keywords, we collected data on advertisers' positions or ranks achieved on listings (for both Yahoo! and Google) from the sponsored search results, once every day for a period of 60 consecutive days in late 2004. Of these 36 keywords, not all of them received sufficient bids from keyword advertisers across both these search mechanisms. We therefore restricted our focus to keywords that had a sufficient number of advertisers bidding for keywords representing the specific products, and also discarded any bidders that bid less than 30 days for each of the keywords. This helps to ensure that our data is devoid of noisy or sporadic behavior patterns exhibited by sellers. After maintaining the same number of keywords across good types, our final dataset consists of 9 keywords in each category, as listed in Table 1.

Table 1. Classification of Products as per SEC Framework

Search	Experience	Credence
Apparel	Auto Insurance	Cosmetic Surgery
Books	Brokerages	Counseling
CD	Cruises	Security Systems
Cell Phones	Event Planning	Pest Control
Flight Tickets	Healthcare	Psychics
Laptops	Jewelers	Tax Services
Refrigerators	Martial Arts	Therapy
Television	Moving & Storage	Used Cars
Toys	Perfumes	Vacation

The data on advertiser/seller quality was gathered from Alexa.com, which collects detailed usage data from the millions of users that contribute this information by using the online Alexa toolbar. Alexa then makes publicly available data on aggregate website statistics that we use for the purposes of our study. Prior research has employed Alexa data as a proxy of website quality [12] as well as a proxy for firm's brand equity or social capital [13]. The measures in our study are classified into three groups: a) the outcome of sponsored search auctions or the POSITION obtained by firms, b) the QUALITY of the firms, and c) the PRODUCT_TYPE.

4.2 Measures

The *dependent variable* of interest is the POSITION (rank) advertisers receive on paid search lists on Yahoo!'s and Google's sponsored search results. In keeping with industry studies that find that consumers typically do not search beyond the first page of search results[1], we restrict our focus to the top fifteen search listings for each keyword. Within each keyword category, advertising firms are first ordered by their average rank/position in the sponsored search listings over the period of our data collection (not including the days that they did not bid) and the top fifteen ranked firms are then selected to form a smaller subset. We use two measures to depict the POSITION achieved by advertisers. The continuous variable *AvgRank* is a measure of the average POSITION obtained by the advertising firm in the paid search listing over all the days it bid. *OrdRank* represents the POSITION measure in discrete form, i.e. firms are first ordered based on the rank averaged by them over the 60-day observation period, and then ranked from 1(top) to 15 (bottom).

The main *independent variable* of interest in the study is the QUALITY of the advertiser. As noted earlier, advertiser-quality is a multi-dimensional construct. As a result we use three different measures/proxies of advertiser QUALITY, which are particularly relevant in online settings. The greater the number of page views a seller's website attracts, and higher the proportion of all web users that visit the website, the higher we may regard the seller's quality to be. Alexa refers to these two numbers as page view (fraction of all the page views by toolbar users that go to a particular site, per million) and page reach (percentage of all Internet users who visit a given website), respectively. A new variable, *TrafficRank*, is then computed by Alexa that collectively represents both page view and reach. For *TrafficRank*, a lower value indicates higher quality and vice versa. It is important to note that search engines like Google and Yahoo! are just one of the several sources that drive traffic to a website, the others being links from other websites, word-of-mouth, referrals by friends, and random surfing, among others. Consequently, a seller's ranking (POSITION) on Google or Yahoo! sponsored listings is neither synonymous with, nor the primary determinant of its *TrafficRank* on Alexa. This is further confirmed by our tests for endogeneity and the robustness checks (for more details see Section 5). A second measure is provided by the number of incoming links to a website, *Inlinks*. Originally popularized by search engines such as Google and Yahoo!, links pointing to a website are now commonly used as a measure of quality. An incoming link is considered as a positive recommendation by the originator of that link, so the more the *Inlinks for* an advertiser the more important it is considered to be. The third variable, *Ratings*, is calculated by averaging over the scores provided by customers who visit sellers' websites and rate them on their purchase and shopping experiences. *Ratings* are measured on a scale from 1 to 5. The use of three distinct measures reinforces the robustness of our findings. It is pertinent to note here that a positive (negative) relationship between QUALITY and POSITION depicts the presence of higher quality firms in higher (lower) positions.

Search, *Experience,* and *Credence* are binary (dummy) variables that represent PRODUCT_TYPE increasing in pre-purchase quality uncertainty. Finally, we collect information on the *age of the firm* from Alexa. AGE is measured as the number of days the firm has existed online and serves as a control.

4.3 Empirical Analyses

Our primary objective in this study is to examine how the relationship between seller QUALITY and POSITION varies across the different product categories and across the two sponsored search markets. We conduct a series of increasingly sophisticated analyses. We use an OLS model as a benchmark in our analysis. Interaction terms are created using centered main effects variables, QUALITY and PRODUCT_TYPE to minimize multi-collinearity. The OLS model uses a continuous dependent variable, *AvgRank*. Our dependent variable is however naturally ordered as it measures the outcome or position in sponsored listings, and we therefore also repeat the analysis for an ordered dependent variable, *OrdRank* using ordered probit regressions. Further, we control for the age of the firm in order to account for the possibility that newly established, not necessarily low quality, firms may have lower *TrafficRank*. After accounting for missing values from the top fifteen ranked firms for the 27 product keywords, our total sample for the analysis of Yahoo! data is 353 (for OLS) and 350 (for ordered probit) observations. The corresponding numbers for Google data are 274 and 272. We specify the following equations, 1a and 1b, with age normalized quality measures, N_QUALITY, for both Yahoo! and Google.

(1a) $\mathbf{AVGRANK} = \gamma_1 + \gamma_2\mathbf{N\_QUALITY_i} + \gamma_3\mathbf{PRODUCT\_TYPE_i} + \gamma_4\mathbf{N\_QUALITY_i * PRODUCT\_TYPE_i} + \varepsilon_i$

(1b) $\mathbf{ORDRANK} = \delta_1 + \delta_2\mathbf{N\_QUALITY_i} + \delta_3\mathbf{PRODUCT\_TYPE_i} + \delta_4\mathbf{N\_QUALITY_i * PRODUCT\_TYPE_i} + \varepsilon_i$

It is further possible that unobserved variables relating to each keyword affect the outcomes observed in the above analyses. While the above analyses assume that the observations on the independent variables are not systematically correlated with the error terms, the observations, and subsequently residuals, within each keyword may not be independent. We specify additional models to deal with this structural complexity using clustering and fixed effects at the SEC sub-sample level. The results of these analyses are consistent and are not presented here for sake of brevity.

We are further interested in examining if and how the relationship between seller QUALITY and POSITION in the search listings differs *across* POSITIONS, or levels of the dependent variable. We do so by employing *quantile regression* analysis [7] to models 1a and 1b. While OLS regression estimates the regression coefficient at the conditional mean of the regressor's distribution, quantile regression can provide parameter estimates at *different quantiles of the dependent variable*. This enables us to examine the variation in the effect of independent variables on the dependent variable at different quantiles. Thus, quantile regression allows for presence of heterogeneity in the QUALITY-POSITION relationship across different POSITIONS in the sponsored search listings.

Our analysis thus far assumed a linear relationship between advertiser's QUALITY and POSITION in the search results. In this section we test for the presence of non-linear relationship between the seller QUALITY and POSITION in the search listings. We conduct separate regressions for each product category (i.e. *Search*, *Experience,* and *Credence*) for both Yahoo! and Google, as specified in 2a and 2b.

(2a) $\mathbf{AVGRANK} = \alpha_1 + \alpha_2\mathbf{N\_QUALITY_i} + \alpha_3\mathbf{(N\_QUALITY_i)}^2 + \varepsilon_i$

(2b) $\mathbf{ORDRANK} = \beta_1 + \beta_2\mathbf{N\_QUALITY_i} + \beta_3\mathbf{(N\_QUALITY_i)}^2 + \varepsilon_i$

5. RESULTS

Significant correlations exist among the three measures of QUALITY for Yahoo and Google respectively: *TrafficRank* and *Inlinks* (0.84 and 0.85), *TrafficRank* and *Ratings* (0.62 and 0.62), and finally, *Inlinks* and *Ratings* (0.60 and 0.57), all of which are significant at the p<0.001 level. Among the three however, the most direct and comprehensive quality measure available is *TrafficRank*, and therefore we present regression results using it as our primary measure of seller quality. The results remain consistent across *Inlinks* and *Ratings,* as discussed in Section 5. The results from the baseline OLS regression analyses are presented in Table 2.

Table 2. Regression Analyses

	Yahoo (N = 350)		**Google (N =272)**	
	Y1a	**Y1b**	**G1a**	**G1b**
Model	**AVG RANK**	**ORD RANK**	**AVG RANK**	**ORD RANK**
Constant	**8.60** **(0.23) *****		**4.73** **(0.15) *****	
N_QUALITY	**0.47** **(0.24)***	**0.14** **(0.06)****	**0.40** **(0.15) *****	**0.21** **(0.07) *****
EXPERIENC	-0.25 (0.26)	-0.09 (0.06)	-0.05 (0.17)	0.04 (0.08)
CREDENCE	-0.16 (0.27)	-0.05 (0.06)	-0.18 (0.18)	-0.06 (0.09)
QUALITY X EXPERIENC	**-0.50** **(0.29)***	**-0.13** **(0.07)***	0.12 (0.17)	0.06 (0.08)
QUALITY X CREDENCE	**-1.30** **(0.27) *****	**-0.28** **(0.07) *****	0.05 (0.17)	0.06 (0.08)
F or $^{\$}X^2$ (5)	**5.53*****	**23.73 *****	**1.89 ***	**13.18 ****
$^{\$}R^2$	0.075	0.013	0.034	0.009
Adj. R^2	0.061		0.016	

* p<0.1 level, ** p<0.05 level and *** p< 0.01 level

Parentheses contain standard errors

\$ Pseudo-R^2 and χ^2 reported for ordered probit using ORDRANK

It should be noted here that since the left out category among product types is search, the coefficient of QUALITY in the regression equations represents the effect of one unit of change in QUALITY on the POSITION in listings for *Search* goods. The two interaction terms between QUALITY and *Experience* and QUALITY and *Credence* goods then are a measure of how much this association changes for *Experience* and *Credence* goods, *relative to* Search goods.

We first compare the corresponding baseline models across Yahoo! (Y1a, Y1b) and Google (G1a, G1b) depicted in Table 2. Across all the aforementioned models, we find that QUALITY is *positively correlated* with average POSITION obtained by the firm in the sponsored search listings for *Search* goods on both search mechanisms. More interestingly, we find that the coefficients of the interaction between PRODUCT_TYPE and seller QUALITY are *negative* and significant in the case of *Experience* and *Credence* goods in all the four models for Yahoo!, but not for Google. These findings suggest that the relationship between seller QUALITY and POSITION achieved in listings are significantly different for *Experience* and *Credence* goods, as compared to *Search* goods for Yahoo!. In the case of Google, however, there appears to be no significant differences across the three product categories.

While the above results focus on the *relative* differences in the relationship between QUALITY and POSITION in listings across the three product categories, it is important to examine the *absolute* relationship between QUALITY and POSITION for each of the three product categories. We test for such a relationship across all three PRODUCT_TYPES using tests of linear combinations (as depicted in Table 3) for Yahoo! and Google, corresponding to the models in Table 2. These tests assess whether the absolute coefficient of quality on position in listings is different from zero across *Search*, *Experience*, and *Credence* categories.

Table 3. Absolute coefficients for Quality- Position

	Yahoo (N = 350)		**Google (N =272)**	
	Y1a	**Y1b**	**G1a**	**G1b**
Model	**AVG RANK**	**ORD RANK**	**AVG RANK**	**ORD RANK**
Search	**0.44 (0.24)***	**0.14 (0.06)****	**0.40 (0.15)*****	**0.21 (0.07)*****
Experience	-0.05 (0.40)	0.01 (0.10)	**0.53 (0.23)****	**0.27 (0.11)****
Credence	**-0.86 (0.37)****	-0.14 (0.09)	**0.46 (0.22)****	**0.27 (0.11)****
F-test for equality of interaction coefficients	**F(3,344= 9.21 *****	**χ^2 (3) = 23.63 *****	**F(3,266) = 2.66 ****	**χ^2 (3) = 9.22 ****

* p<0.1 level, ** p<0.05 level and *** p<0.01 level

Parentheses contain standard errors

$ Pseudo-R^2 and χ^2 reported for ordered probit using ORDRANK

These results for Yahoo! in Table 3, which are mostly consistent across models Y1a and Y1b, suggest that the coefficient for QUALITY is positive and strongly significant for *Search* goods. On the hand, the coefficient for *Credence* goods is negative and significant in the linear regressions with *AvgRank*. The corresponding coefficients for *Experience* goods lie in between those of *Search* and *Credence*, and are not significantly different from zero. Table 3 suggests that different outcomes are evident on Google's sponsored listings. The coefficients for all three product types are positive and significant across G1a and G1b. Results of our analyses using clustering and fixed effects models are robust. However, since the Hausman (1978) null is not rejected, the random keyword effects model is a better choice for our setting.

Next, we assess findings from quantile regression. Our findings show that the results from the linear regression hold at almost all the position ranges (i.e. from position 1 to position 15). We display the quantile graphs in Figure 1 for the two extreme cases for Yahoo! – *Search* (Fig. 1a) and *Credence* goods (Fig. 1b), and for *Credence* goods for Google (Fig. 1c), since the other two are very similar. The coefficients of Quality-Position are positive (the graph in 1a lies above 0) across all quantiles for search goods for Yahoo!, and the interaction coefficient (depicting the difference between *Search* and *Credence* goods) is consistently negative (the graph in 1b lies below 0) across all quantiles.

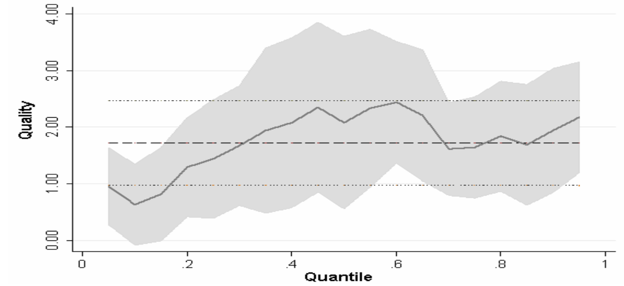

Figure 1a: Yahoo! Search Goods

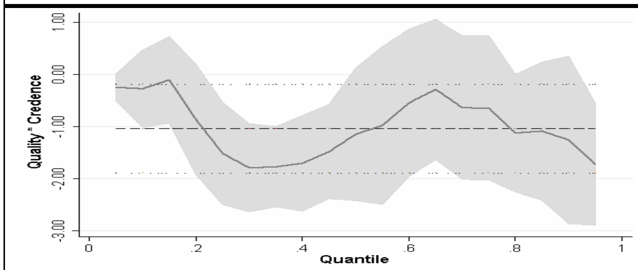

Figure 1b: Yahoo! Credence Goods

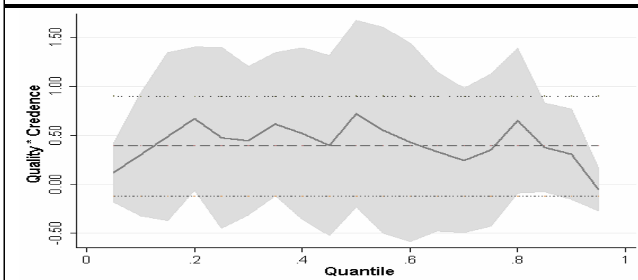

Figure 1c: Google Credence Goods

Figure 1. Graph depicting quantile regression results

The advantage of quantile regression here lies in indicating how the strength of the negative correlation between QUALITY and POSITION for *Credence* goods changes across quantiles of sponsored search listings. As shown in Figure 1, the negative correlation appears to be the strongest in the top positions of sponsored listings, reinforcing our earlier findings.

Last, we examine the results from the non-linear model. The results of this analysis are depicted in Figure 2a for Yahoo! and Figure 2b for Google. As indicated by the graphs in these figures, the quadratic terms are not significant for the two extreme product categories - *Search* and *Credence* goods, but the relationship is markedly non-linear for *Experience* goods. The coefficients for the squared quality variable in the model are not significant for search and credence goods, in both Google and Yahoo!. However, squared term for *Experience* goods is significant in both Yahoo! and Google, suggesting a non-linear relationship. The coefficients of the squared QUALITY term in the equation 2a are 0.69 ($p < 0.10$) and 0.47 ($p < 0.05$) for Yahoo! and Google respectively. In equation 2b, the coefficients of squared QUALITY term are 0.21 ($p < 0.05$) for both Yahoo! and Google.

Figure 2a: Yahoo

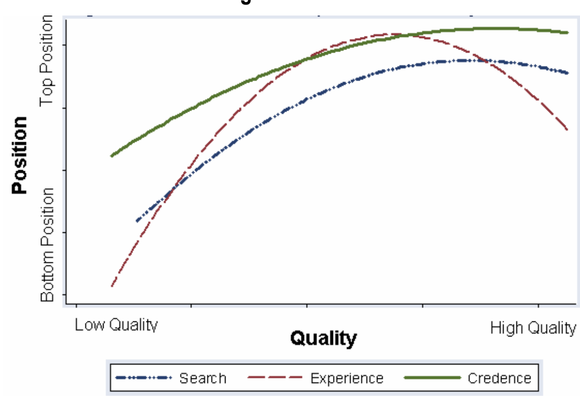

Figure 2b: Google

Figure 2. Quadratic Relationship Between Quality and Position Across SEC Goods

5.1 Robustness Checks

Tests for Endogeneity: The issue of endogeneity arises as an artifact of our particular measure of seller quality- *TrafficRank*. It is possible that *AvgRank*, the measure of the advertiser's POSITION in sponsored search listings may affect its *TrafficRank*, the measure of seller's QUALITY in our study. In other words, better positions in the search listings could increase traffic to the advertiser's site. However, our key findings of a negative correlation between QUALITY and POSITION for *Experience* and *Credence* goods in the case of Yahoo! imply that this relationship is in fact, reversed – i.e. *better* positions in the search listings are actually correlated with *lower* traffic. Thus, these results are likely to be strengthened in the absence of any potential endogeneity.

If our model suffers from recursive endogeneity (i.e. *AvgRank* affects *TrafficRank* and *TrafficRank* affects *AvgRank*), OLS would be insufficient. We therefore test for the endogeneity of *TrafficRank* using the Wu-Hausman F test and the Durbin-Wu-Hausman chi-sq test. These tests examine the null hypothesis that *TrafficRank* is exogenous by checking for a statistically significant difference between the OLS and 2SLS estimates of its beta coefficient when regressed on *AvgRank*. The OLS model is as specified earlier. In the 2SLS model, we use the variable AGE as the instrument for *TrafficRank*. Theoretically, the AGE of a website would be correlated with its traffic (QUALITY), and therefore can be used to predict the latter. It is unlikely though, that AGE has a direct impact on the POSITION of the seller on sponsored listings.

From the 2SLS analyses, we find that the first-stage F is highly significant for both Google ($F(5,266)=51.34$, $p<0.01$) and Yahoo ($F(5,344)=39.62$, $p<0.01$); the F-statistic is much higher than the minimum value of 10 [16]. The corresponding coefficients of AGE are also significant ($p<0.01$) indicating that AGE is both a valid and relevant instrument. However, since neither the Wu-Hausman F test or the Durbin-Wu-Hausman χ^2 test is rejected, we fail to reject the null that *TrafficRank* is exogenous. Based on these analyses, we find that OLS is unbiased, consistent and the more efficient estimator for our model. Therefore we focus on the OLS estimates in our discussions. Further, the point estimates are qualitatively unaffected if we use 2SLS.

Additional Tests for Robustness: In this section, we address several sensitivity concerns that might arise from the measures and models used in our analyses, by conducting appropriate checks of robustness [2]. First, we examine the robustness of our findings by using several different measures of *TrafficRank* computed over different time periods. We then re-assess our models using several combinations of the three quality measures described earlier. We also control for the presence of different types of sellers in the sponsored search listings, and finally, we analyze the sensitivity of our results to subsets of keywords in each product category.

Alexa provides for each seller *TrafficRank* measures computed over 3-months, 1-month, 1-week and 1-day, along with an instantaneous measure, which is the one we used in the analyses presented above. We repeat our analyses using these measures of *TrafficRank* collected over different time periods, and find that our results are robust. The coefficient for the QUALITY for *Search* goods on POSITION is directionally consistent in all cases. *Experience* interaction coefficients for Yahoo across the 3-months, 1-month, 1-week and 1-day models are all negative and significant ($p<0.10$ or better); *Credence* interaction coefficients for Yahoo! are also negative and significant ($p<0.01$ or better) respectively. The corresponding interaction coefficients for Google are all insignificant. This suggests that our findings of a negative correlation in the case of *Experience* and *Credence* goods for Yahoo! are robust to changes in the *TrafficRank* measure. We also obtain consistent results when we test for clustering and keyword effects using the 3-month *TrafficRank* measure.

Next, we consider the two other available measures of seller quality - *Inlinks*, and *Ratings*. We first use *Inlinks* and *Ratings* as separate measures of QUALITY. Across all these models, the coefficient of QUALITY on POSITION is significant for *Search* goods for Yahoo ($p<0.10$ or better) and for Google ($p<0.05$ or better). Moreover, the QUALITY coefficients for *Credence* goods is significant and directionally opposite to Search ($p<0.01$ or better); while the corresponding QUALITY coefficients for *Experience* goods are directionally consistent, but mostly insignificant in the case of Yahoo!. Both interaction coefficients are not significantly different from that for *Search* goods in the case

[2] The results from these additional analyses are available from the authors upon request.

of Google. These results reinforce the validity of our original findings.

Another check we conduct ensures that our results are not affected by the presence of different types of sellers such as retailers/manufacturers and infomediaries, in the sponsored search listings for a given keyword. Despite the presence of a mix of seller types, it is important to note that our results are based on an ordinal ranking of the sellers in the sponsored search listings; this ordinal ranking of sellers is still preserved with each type, and our results would therefore still hold. However, to test this, we include dummies to control for the presence of a mix of seller types. We find that the dummy for infomediaries is insignificant for Yahoo, but significant for Google ($p<0.05$). After controlling for their effects, we observe that the main QUALITY coefficient for *Search* goods continues to be significant for Google ($p<0.01$) but is interestingly insignificant for Yahoo. Further, while the product type interactions are insignificant for Google, both the *Experience* and *Credence* interaction coefficients are still directionally consistent for Yahoo, but only the latter coefficient is significant ($p<0.01$).This suggests that in the worst case, the QUALITY-POSITION relationship may be reversed even for Search goods in Yahoo!'s listings.

Finally, we examine the sensitiveness of our models to our choice of keywords. We reran our analyses presented in Table 2 (with 9 keywords each) using several combinations of 7 and 5 keywords each across Search, Experience and Credence goods. We obtain consistent results that reinforce the differences in the relationships for Experience and Credence goods across Yahoo! and Google.

6. DISCUSSION AND IMPLICATIONS

Our results indicate that the sponsored search markets suffer from adverse selection; however the intensity of adverse selection differs across markets as well as product characteristics. While adverse selection was almost non-existent in the market for *Search* goods, the unregulated sponsored search mechanism used by Yahoo! suffered from problems of adverse selection for *Experience* and *Credence* goods. However, Google's intervention mechanisms of ranking bidder advertisements (by moderating the advertiser's willingness to pay with its performance measured by clickthrough rates) seem to be capable of circumventing the problem of adverse selection for *Search* as well as *Credence* goods. While adverse selection issues in the case of *Experience* goods are not as severe as in the case of *Credence* goods, Google's intervention mechanism does not seem to alleviate this issue.

This could adversely impact consumer welfare particularly for uninformed consumers and consumers who trust the search results provided by these search engines. This risk has been identified by consumer advocacy groups such as Consumer Reports WebWatch, as is evident from the following excerpt from their report [8]:

> *"... trust in search engines may make them (online consumers) vulnerable while online, as they are largely unaware such navigation sites often accept fees in exchange for giving advertiser Web pages prominent placement on their search results pages."*

While online markets can improve consumer welfare by lowering search costs, in the presence of uncertainty regarding unobservable quality characteristics, sellers can distort or hide information, leading to adverse selection. Thus, it is possible that the higher costs of adverse selection counteract the benefits gained from lowered search costs for consumers. The sponsored search market provides an excellent test bed to examine issues of adverse selection. As advertising channels, online search mechanisms such as Yahoo! and Google lack traditional differentiators of firm quality, forcing consumers to seek out alternate sources of information, such as ranking of advertisements on paid listings. Our study is among the first to examine the performance outcomes of these important and powerful online advertising mechanisms.

Our findings also add to the existing research that examines the efficacy of online markets for different product categories. In particular, the comparison of the two most popular sponsored search mechanisms allows us to illustrate their differential effectiveness in abating adverse selection. Our study also contributes to the literature on advertising by testing traditional theories in emerging channels. Just as eBay resorts to user-feedbacks and Amazon to reviews and ratings to alleviate adverse selection problems, our findings suggest that online sponsored search mechanisms may be able to decrease the negative impacts of adverse selection in markets with high pre-purchase uncertainty by providing alternate signals of quality about advertisers.

Our findings also have significant implications for the providers of search services. Sponsored search listings that are biased can adversely affect consumer welfare; in addition they can also drive out higher quality firms, and eventually, reduce the profitability of the intermediary as well. Search intermediaries would benefit by providing better information regarding their paid search mechanism and incorporating reputation mechanisms to aid consumers in their decision-making for online purchases. Our findings indicate that additional signals of quality can potentially help to improve the efficiency and welfare properties of the sponsored search markets by reducing adverse selection. Provision of such additional quality information such as ratings and reviews from Bizrate.com and Epinions.com alongside the search listings can help reduce the risk faced by the consumer and improve consumer welfare.

7. CONCLUSION

It is appropriate to discuss some of the limitations of our study. First, adverse selection in markets associated with unobservable seller characteristics such as quality has typically been known to be difficult to measure empirically. This is because reliable quality signals are hard to come by. Alexa is the only publicly available source for website *TrafficRank*. *TrafficRank* is calculated by aggregating the traffic generated by the site from among a subset of online users. Our measure of quality is therefore only accurate to the extent that this segment of consumers is representative of the broader online population. The use of two other measures of seller quality, namely the number of incoming links and ratings provided by a subset of online consumers provides us with a way to triangulate our findings. More importantly, our focus is on the *relative* performance of sponsored search and our key results highlight interesting differences across the three different product categories as well as across the two sponsored search mechanisms using the same measures of QUALITY and POSITION.

Second, traffic generated by the website is usually a function of how long the website has been in existence. The use of web-site traffic as a measure of quality would be problematic and confound the results of our study, if these new entrants and niche marketers were high quality sellers. We address this issue by normalizing quality by the age of seller, or the length of its existence online.

Further, we find that all three measures of quality are highly and significantly correlated. While traffic rank and incoming links could be affected by the newness or niche focus of the website, the third measure – ratings, is less likely to be influenced by age. These measures therefore help reduce any potential confounding effects. Third, while the distinction among *Search*, *Experience,* and *Credence* goods is well established in theoretical literature, in reality all goods have search, experience and credence attributes, albeit to varying extents. While our classification largely conforms to prior literature, what is more important is that these products are characterized by increasing uncertainty for consumers purchasing them.

Our study opens doors for plenty of future research opportunities. Sponsored search auctions for keywords, though growing rapidly, are still in their infancy. Despite the nascence of sponsored search/advertising mechanisms, there exist significant differences between traditional advertising formats and sponsored search formats. Of particular interest is the fact that sponsored search advertising is a *performance-based* advertising model where firms incur an expense only when consumers click on the links to their websites. Thus firms' advertising expenses are closely linked to their revenues from potential sales to online consumers. In comparison, advertising in traditional print/broadcast media is characterized by fixed costs and further removed from any potential sales. Thus the two advertising formats (traditional vs. sponsored search) differ in their risk to advertisers. Future research should examine the implications of these differences on the incentives for (low vs. high quality) firms to advertise. The keywords in our study were chosen based on their popularity (as published on publicly available sources). It would be interesting to examine whether our results hold for "niche" or less popular keywords with fewer bidders. It would also be useful to examine the bidding patterns for keyword combinations as well as for keywords representing brands (such as "Sony Vaio" or "Dell Inspiron") rather than generic products.

Our study demonstrates the usefulness of SEC framework in studying the impact of different market mechanisms on the market effectiveness. Future research can employ the same SEC framework to study the dynamics of bidding strategies of the firms over a period of time. A more extensive analysis of such bidding dynamics promises to shed light on relative competition across different product categories as well as the existence of strategic groups within product categories.

Further, online seller quality is multidimensional, and possible extensions to our work may consider the impacts of several other alternate dimensions of quality such as website quality, seller trustworthiness, etc. Future studies should also examine consumer behavior in response to the sponsored search phenomena. Laboratory studies designed to analyze the differential search strategies adopted by consumers would help understand how consumer search across different search formats. Studies of this nature are sparse, given the novelty of the phenomenon. Whether findings of studies relating to consumer behavior in traditional channels translate well to online settings is an empirical question yet to be answered.

8. ACKNOWLEDGMENTS

We would like to thank the NET Institute for generous financial support. This paper has greatly benefited from comments from anonymous reviewers and participants at the 2006 NET Institute Workshop.

9. REFERENCES

[1] BBC News. Search users 'stop at page three', 2006.

[2] Comanor, W.S. and Wilson, T.A. Advertising and Competition: A Survey. Journal of Economic Literature, 17, 2.(1979) 453-476.

[3] Darby, M.R. and Karni, E. Free Competition and the Optimal Amount of Fraud. Journal of Law and Economics, 16, 1.(1973) 67-88.

[4] DoubleClick. Search Trend Report Q4 2005, 2006.

[5] Ekelund, R.B., Mixon, F.G. and Ressler, R.W. Advertising and Information: An Empirical Study of Search, Experience and Credence Goods. Journal of Economic Studies, 22, 2.(1995) 33-43.

[6] Kihlstrom, R.E. and Riordan, M.H. Advertising as a signal. Journal of Political Economy, 92, 3.(1984) 427–450.

[7] Koenker, R. and Bassett, G. Regression quantiles. Econometrica, 46, 1.(1978) 33-50.

[8] Marable, L. Consumer Reaction to Learning the Truth About How Search Engines Work, Consumer Reports WebWatch, 2003.

[9] Markoff, J. and Ives, N. Web Search Sites See Clicks Add Up to Big Ad Dollars, New York Times, 2005.

[10] Moorthy, S. and Zhao, H. Advertising Spending and Perceived Quality. Marketing Letters, 11, 3.(2000) 221-233.

[11] Nelson, P. Information and Consumer Behaviour. Journal of Political Economy, 78, 2.(1970) 311-329.

[12] Palmer, J.W. Web site usability, design, and performance metrics. Information Systems Research, 13, 2.(2002) 150-167.

[13] Palmer, J.W., Bailey, J.P. and Faraj, S. The Role of Intermediaries in the Development of Trust on the WWW: The Use and Prominence of Trusted Third Parties and Privacy Statements. Journal of Computer-Mediated Communication, 5, 3.(2000).

[14] Schmalensee, R. A Model of Advertising and Product Quality. Journal of Political Economy, 86, 3.(1978) 485-503.

[15] Sherman, C. A new F-word for Google search results, Search Engine Watch, 2005.

[16] Staiger, D. and Stock, J.H. Instrumental variables regression with weak instruments. Econometrica, 65, 3.(1997) 557-586.