Brand effects in search advertising*

Przemysław Jeziorski
University of California, Berkeley

Sridhar Moorthy
University of Toronto

September 11, 2014

*We thank Avi Goldfarb, Matthew Osborne, Tanjim Hossain, Kinshuk Jerath, and seminar audiences at University of Toronto and Northwestern University (2014 Theory + Practice in Marketing Conference) for useful feedback on this work. This research was supported by Grant #s 435–2013–0704 and 864–2007–0306 from the Social Sciences and Humanities Research Council of Canada to the second author. We also like to thank Microsoft Research for providing the data.

Abstract

We develop and estimate a dynamic Bayesian model of consumer response to sponsored search advertising. In our model, consumers use search ads to search the offers ("quality") of competing online retailers selling a branded product. Their sequential search is directed by two types of ad prominence: prominence due to ad position and prominence due to advertiser brand. Using individual-level click-stream data from Microsoft Live Search and measures of brand strength from Alexa.com, we estimate a structural model to recover the unobserved qualities of the retailers' offers, and estimate how ad position and advertiser brand interact in determining click-through-rate (CTR). We find that, contrary to previous assertions in the literature, ad position and advertiser brand strength are substitutes, not complements. Specifically, a weakly branded retailer increases its click-through-rate by as much as 150% in going from the bottom ad position to the top, while the corresponding increase for a strongly branded retailer is less than 50%. In addition, order of clicks and quality—which is essentially lower price—are positively correlated, suggesting that ad prominence, of either type, is not a significant source of market power for the search advertiser. These results suggest that in terms of gaining attention, search advertising is not very different from traditional TV or print advertising, but in its power to inform and persuade, search advertising is necessarily more limited.

1 Introduction

Paid search advertising is the advertising that occurs when a consumer enters a keyword in an online search engine and searches for something. The ads— sponsored links, as they are sometimes called—appear at the top and right-hand side of the results page. This form of advertising is relatively new; it didn't exist until 1998. However, already by 2003, it was the largest online ad medium, which it continues to be with nearly 40% of the market.[1]

While search advertising might seem to comprise equal amounts advertising and search, it is, in fact, mostly search. The ads themselves are fairly uninteresting objects—certainly nowhere near as rich as TV and print ads. Other than identifying the advertiser and providing a clickable link to the advertiser's website, they are relatively contentless. The advertiser pays for clicks, not for exposure.

This paper is about search within search advertising. When a consumer searches a keyword and the list of ads appears, how does she search among them? In what order does she click? On what does the clicking order depend? How much will the click-through-rate (CTR) of an ad improve if the advertiser moved up the list?[2] What do consumers' click patterns say about the quality of advertisers' offers?

The literature in economics and marketing has noted the importance of these questions, but definitive answers remain elusive. Edelman, Ostrovsky, and Schwarz (2007) and Varian (2007) simply assume that consumers click from top to bottom, independent of advertiser identities. More common is the assumption that CTR depends both on ad position and advertiser characteristics, independently. That is, $\text{CTR}_{ij} = \alpha_j \beta_i$, where α_j is an ad position factor and β_i is an advertiser factor (Aggarwal, Goel, and Motwani (2006), Katona and Sarvary (2010)). While there is empirical support for higher ad positions eliciting higher CTR in the aggregate (Brooks (2006), Rutz, Bucklin, and Sonnier (2012)), disaggregate

[1]Market share data from IAB Internet Advertising Revenue Reports April 2003 and April 2014. Oremus (2013) notes that a search engine called GoTo.com (later Overture) was the first to offer paid-search ads. Growth accelerated in 2002 when Google launched its own pay-per-click, auction-based search-advertising product called AdWords Select.

[2]Click-through-rate (CTR) is the number of clicks an ad generates as a fraction of the number of times it is presented.

analysis shows considerable variation around this pattern. For example, Jerath, Ma, Park, and Srinivasan (2011) demonstrate that particular advertisers in lower ranks get higher CTRs than advertisers above them. Jeziorski and Segal (2014) show that consumers do not always click from top to bottom, and that the CTR for a particular advertiser in a particular ad position depends on who else is advertising and in what positions.

In this paper we develop a structural model of consumer response to search ads and estimate it using individual-level clickstream data from Microsoft Live Search. In the process we shed light on how two different measures of ad prominence, ad position and advertiser brand, affect CTR. In our model, CTR is the outcome of a consumer search process through search ads (Athey and Ellison (2011)). What consumers are searching for is the advertiser with the best "quality." For branded keywords (such as "Nikon camera"), where the advertisers are typically retailers, "advertiser quality" translates to "retailer with the best deal," where "deal" includes not only price, but also delivery terms, and other retailer characteristics such as reliability and post-sales service.[3] If the ads themselves provided this information, there wouldn't be any need to search. But given the sparseness of information in the ads, the consumer has to click on the ads and visit the advertisers' websites to get more information. After each such visit, the consumer has to decide whether to search further, or to terminate the search, and buy from the best retailer visited until then.

In other words, what we have is a sequential search model. But it is not a random sequential search model. This is because the ads are not symmetric. For one thing, they vary by position. For another, the brand identities of the advertisers are different, and consumers likely have different perceptions of different advertisers. For example, one advertiser might be Amazon.com and another advertiser might be Adorama.com, and consumers may have quite different perceptions of these retailers' prices, delivery terms, reliability, etc. The search is thus best characterized as a directed search through a set of probability distributions parametrized by ad position and advertiser brand (Weitzman (1979)).

Our analysis reveals that many of the assumptions about CTR in the theoretical literature are not satisfied in the data. First, in nearly two-thirds of the impressions with multiple

[3]According to the Interactive Advertising Bureau, retailers are the largest category of search advertisers (IAB Internet Advertising Annual Report 2014).

clicks, the consumer clicks non-sequentially; a click on the jth link is generally *not* the jth click. Second, in nearly a third of the impressions with more than two clicks, the consumer changes direction: she proceeds upward after proceeding downward. These observations suggest the importance of advertiser identity and of context, and our subsequent results pin down more precisely how these things matter. We find that the independence assumption completely mischaracterizes the nature of the interaction between ad position and advertiser brand strength. Independence implies that the marginal CTR-value of a higher ad position, $(\alpha_j - \alpha_{j+1})\beta_i$, is increasing in advertiser brand strength. Instead, we find that stronger brands benefit less from ad position. The independence assumption also implies that *relative* marginal value of ad position, $(CTR_{ij} - CTR_{ij+1})/CTR_{ij+1}$, is independent of advertiser brand. We find that advertiser brand matters both in an absolute sense as well as in a relative sense. These results suggest that consumers' responses to search ads are not fundamentally different from their responses to TV and print ads. Moreover, we find that higher ad positions generate condirable amount of market power for non-branded advertisers and little market power for branded advertisers. It suggests that market power from the brand dominates and nullifies the market power from ad position.

The rest of the paper is organized as follows. In Section 2 we describe our data and the institutional setting. In Section 3 we provide model-free evidence on the questions of interest. These preliminary results motivate the structural model in Section 4, which we estimate using the procedures discussed in Section 5. Section 6 presents our results, and Section 7 concludes the paper.

2 Data and institutional details

When a user submits a search query ("keywords") to an online search engine, the result is a "search impression." A search impression is a list of *organic* and *sponsored* links ("paid-search ads"). Organic links appear on the left-hand side of the results page; sponsored links appear at the top, above the organic links, and on the right-hand side. Each search ad is a brief paragraph of text—perhaps two or three lines—of which the most notable part is the advertiser's (clickable) name and web address.

The search ads featured in an impression are determined in a generalized second-price auction (Edelman, Ostrovsky, and Schwarz (2007)). Businesses bid for advertising slots associated with particular keyword(s) by submitting a price per click they are willing to pay. The search engine weights the bids by proprietary quality scores and runs the auction, the outcome of which is an ordering of ads and a price-per-click for each advertising slot. Quality scores and bids are keyword- and advertiser-specific, but not consumer-specific. In other words, this is "untargeted" advertising. The search engine is compensated only if a consumer clicks on a sponsored ad.

Our search ads data come from Microsoft's search engine, *Live Search*, the precursor to today's Bing. According to `techcrunch.com`, in May 2008 Live Search had 9.1% of the U.S. online search market (compared to 61.6% for the market leader, Google). This translated to about 900 million search queries per month. In 2008, as part of its Beyond Search initiative, Microsoft made available to a limited set of academics, a data set containing 20 million search impressions chosen randomly from the ones that appeared over the three months, August 10-November 1, 2007. The sampling scheme involved selecting an impression at random from the log and then including all the other impressions displayed to the same user during the same session. Impressions that were part of longer user sessions thus had a proportionally higher probability of being in the data set than impressions from shorter sessions. However, since the vast majority of the sessions contained only one impression, we believe sample selection is essentially random. The average length of a session is about ten minutes.

Search impressions and user activity are well documented in the data. For each impression we have the list of organic links, the list of sponsored links, order of sponsored links, identity of advertisers, and time stamps of all clicks on organic and sponsored links in a session. However, there are no demographic data on users.

For the purpose of this paper we restrict our attention to "branded keywords" in a particular retail category, digital cameras. We chose this category because it provides a large number of search impressions, and is also fairly tight circumscribed as a category. Within this category, we examine three types of search keywords, those containing the words "Nikon," "Canon," and "Olympus." Advertisers advertising to these keywords are primarily online retailers selling these cameras. Our final sample consists of $28,153$ search impressions.

We supplement the Microsoft data with data from Alexa.com. Alexa ranks websites by their popularity. Its rank "is a measure of how a website is doing relative to all other sites on the web over the past 3 months. The rank is calculated using a combination of the estimated average daily unique visitors to the site and the estimated number of pageviews on the site over the past 3 months." We used Alexa's API to download daily rankings of each retailer during our sample period. Using the daily data we computed a 3-month average Alexa ranking for each retailer and used that as a measure of "advertiser brand strength" in our analysis.[4]

3 Reduced-form evidence

Table 1 contains descriptive statistics from our data set. 13% of impressions have at least one click, about the same as in other studies of sponsored search.[5] Additionally, a non-negligible 2.2% of impression have two or more clicks. Such impressions are important for this study because much of the consumer preferences are identified from data instances in which the consumer clicks on more than one ad.

Define "impressions with non-sequential clicks" as those where there is a click with click position not equal to ad position. For example, a consumer whose first click is on the second ad has made a non-sequential click, as has a consumer whose second click is on the first ad. Non-sequential clicks always involve "jumps" over some ads. We find that 63% of impressions with at least one click are impressions with non-sequential clicks. In other words, most ad impressions do *not* involve clicking from top to bottom, contradicting the assumptions in Edelman, Ostrovsky, and Schwarz (2007) and Varian (2007). Furthermore, in 28% of the impressions with at least two clicks, users click on a higher ad after clicking on a lower ad. This pattern contradicts the so-called cascade models (Craswell, Zoeter, Taylor, and Ramsey (2008)).

Finally, the last four rows of Table 1 show decreasing CTR as we go down the ad hierarchy.

[4]The variation of retailers' Alexa scores within the 3 months of our sample is negligible, suggesting that they might be good proxies for brand strength.

[5]For example, Jeziorski and Segal (2014) find that 17% impressions on non-retail keywords (weather, white pages, sex, games) have at least one click.

Impressions with no clicks	87.04%
Impressions with 1 click	10.75%
Impressions with 2 or more clicks	2.21%
Impressions with non-sequential clicks' (out of impressions with at least 1 click)	62.54%
Impressions with out-of-order clicks (out of Impressions with at least 2 clicks)	28.25%
CTR of the top slot	5.65%
CTR of the second slot	3.86%
CTR of the third slot	2.69%
CTR of fourth and below slots	1.07%

Table 1: Descriptive statistics.

The decrease is rather steep with fourth and below slots receiving five times smaller CTR than the top slot. This suggests either that high CTR is a natural property of high ad positions or that the top slots have "better" advertisers.

In order to disentangle these effects and to get an initial sense of advertiser-ad position interactions, we estimate a linear probability model (Table 2). The dependent variable is a click, and explanatory variables include advertiser fixed effects, ad position fixed effects, and interactions between ad position and a dummy variable representing "top" Alexa rated advertisers (Top 100 Alexa rank = 1, > 100 Alexa rank = 0). The analysis confirms the presence of strong ad position effects even after controlling for advertiser. Being in slot 1 increases CTR on average by 4-6 percentage points depending on keyword (relative to slots 6 and below), while the position effect is not statistically significant beyond position 3 for Nikon and Olympus, and beyond position 4 for Canon. In other words, the slot effects from the linear probability model are similar to the slot effects in Table 1, i.e., the CTR difference across slots can be explained by slot effects alone, suggesting that higher positions do not contain significantly better ads.

The bottom of Table 2 contains interactions between an advertiser's Alexa score and ad

	Nikon			Canon			Olympus		
Ad position	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1	0.055** (0.003)	0.055** (0.003)	0.052** (0.003)	0.052** (0.002)	0.052** (0.002)	0.050** (0.002)	0.043** (0.004)	0.044** (0.004)	0.046** (0.004)
2	0.033** (0.003)	0.032** (0.003)	0.032** (0.003)	0.030** (0.002)	0.032** (0.002)	0.033** (0.002)	0.025** (0.004)	0.027** (0.004)	0.027** (0.004)
3	0.017** (0.003)	0.016** (0.003)	0.017** (0.003)	0.019** (0.002)	0.020** (0.002)	0.022** (0.002)	0.014** (0.004)	0.017** (0.004)	0.015** (0.004)
4	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)	0.005** (0.002)	0.005** (0.002)	0.005** (0.002)	0.006 (0.004)	0.006 (0.004)	0.006 (0.004)
5	0.003 (0.003)	0.003 (0.003)	0.003 (0.003)	0.002 (0.002)	0.002 (0.002)	0.002 (0.002)	0.004 (0.004)	0.004 (0.004)	0.004 (0.004)
1×Top Alexa	-0.025** (0.008)	-0.022** (0.009)	-	-0.022** (0.003)	-0.025** (0.003)	-	0.008 (0.009)	0.005 (0.009)	-
2×Top Alexa	-	0.008 (0.008)	-	-	-0.014** (0.003)	-	-	-0.013 (0.009)	-
3×Top Alexa	-	0.011 (0.008)	-	-	-0.005 (0.003)	-	-	-0.016* (0.008)	-
1-3×Top Alexa	-	-	0.000 (0.006)	-	-	-0.015** (0.002)	-	-	-0.009** (0.006)

Standard error in parentheses, $*p < 0.05$.

Table 2: Results from a linear probability model that predicts clicks as a function of brand dummies (not reported) and position effects. The model includes interactions between a dummy for Top 100 Alexa score and position, thus, it allows for different position effects for retailers with known brands. Specifications (1), (2) and (3), contain interactions between, respectively, the highest, two highest, and three highest positions and Alexa score.

position. We estimate three specifications with these interactions and in each case find a statistically significant impact of advertiser brand strength on the size of the position effect. Namely, for the Nikon keyword, being in the top-100 of the Alexa classification nullifies the marginal effect of the top slot (the difference between the position 1 and position 2 effects). In other words, the retailers that enjoy high Alexa ratings obtain the same CTRs in positions 1 and 2. This doesn't mean, of course, that high Alexa-rated retailers get lower CTRs than low Alexa-rated retailers. In fact, the reverse is generally true: high Alexa-rated retailers tend to have larger retailer fixed effects than low Alexa-rated retailers. The impact of branding is even more dramatic for the Canon keyword. Here, top Alexa retailers enjoy the same CTRs at position 1, 2, and 3. Lastly, for the Olympus keyword, the effects of branding are less pronounced than for the Nikon keyword, although still significant. Here, even though we are unable to identify brand interactions with individual slots because of small sample sizes and lower presence of top Alexa retailers, we are still able to find an aggregate effect of branding on the top 3 slots. In particular, according to the specification (3), top Alexa retailers enjoy 1 percentage point lower gap between top three slots and lower slots, which is equivalent to nullifying the positive impact of slot 3 and decreasing the positive impacts of slot 1 and 2 by 20% and 33%, respectively.

In the reminder of the paper, for each search string, we focus our attention on the top-4 advertisers with the most clicks, pooling all the others into a catch-all "other advertisers" category.[6] Table 3 shows the CTRs of the chosen advertisers sorted by total number of clicks. Note that the CTRs are not monotonic. This is because some advertisers have more impressions than others and hence generate more clicks even with a lower CTR. Table 4 presents the average Alexa scores for the four most clicked advertisers and "other advertisers." We note that in general the most-clicked advertisers have higher Alexa scores than "other advertisers." However, amongst the most-clicked advertisers, we do observe advertisers with low Alexa ratings, such as advertiser 3 for Nikon and advertiser 1 for Canon. Additionally, note that between 7 and 9 percent of the low-CTR "other advertisers," have an Alexa rank of 100 and below, which places them among the most recognizable advertisers on the Internet. This shows the fallacy in using CTRs as a proxy for brand equity in the sponsored search market.

In the structural model developed in the next section, ad position and advertiser brand equity serve as signals of unobserved advertiser quality; they direct the consumer's search by shaping her priors. An important feature of the search data which enables identification of advertiser quality separately from ad position and advertiser brand perceptions is that the advertising in our sample is untargeted, conditional on the search string. Thus, while different

[6]As might be expected, the top-4 advertisers are often the same across keywords—after all, these are all camera brand keywords and camera vendors typically carry multiple brands of cameras. In our analysis, since we do not pool impressions across keywords, we treat the advertisers as distinct.

	Search string		
	Nikon	Canon	Olympus
Advertiser 1	5.9%	14.0%	3.0%
Advertiser 2	10.7%	2.6%	4.9%
Advertiser 3	11.5%	2.4%	4.4%
Advertiser 4	3.1%	3.4%	4.1%
Other advertisers	2.6%	2.4%	2.3%

Table 3: Click-through-rates of top-4 most clicked advertisers for each search string

	Search string		
	Nikon	Canon	Olympus
advertiser 1	$14,650$	$66,379$	24
advertiser 2	515	24	1117
advertiser 3	$66,379$	98	$16,455$
advertiser 4	24	540	540
Other advertisers	$314,537$	$255,582$	$253,532$
% of "other advertisers" ranked in Alexa Top-100	7	6	9

Table 4: Average Alexa ranks of retailers by search string (higher ranks correspond to lower popularity).

users searching the same search string will, in general, see different ads in different positions, what they see in each position will be a random draw from the same set of distributions (Athey and Nekipelov (2010)). Thus, we can treat the choice sets faced by each user as exogenous, and assuming that retailer characteristics do not change within our 3-month window, we can apply retailer-keyword fixed effects to control for unobserved advertiser heterogeneity (beyond the observed Alexa ratings). This identification strategy based on randomly varying choice sets is frequently used in estimating discrete choice demand systems (Ackerberg, Lanier Benkard, Berry, and Pakes (2007)). In our case, the identification is simpler because we do not need to estimate price coefficients. Further discussion of this issue and robustness analysis of the ratailer-keyword fixed effects estimator are presented in Section 8.

4 Model

Heterogeneous consumers are looking to buy a product from an online retailer. To shop for a retailer, they submit a search query to a search engine. The search engine provides a list of sponsored search ads in an impression. Consumers examine the list of advertisers and decide whether to click on an ad or not. A click on an ad takes them to the advertiser's website (the landing page). Consumers peruse the information on the website and then decide whether to buy the product from this retailer, or go back to the impression and click on another retailer's ad. Ultimately this process concludes when the consumer buys from one of the retailers visited, or, finding none of them acceptable, abandons the search and withdraws from the market. The abandonment option we call the "outside option."

Each impression is a list of N ads $a(1), \ldots, a(N)$, with $a(1)$ occupying the top advertising slot and $a(N)$ occupying the bottom advertising slot. Let q_a be the quality of advertiser a's offer. By "offer quality" we mean consumer surplus—utility minus price—offered by the advertiser on the product searched.[7] Consumers are uncertain about this quality after

[7]We are concerned with a retailer's offer quality on a *given* product, whose quality is independent of the retailer. This is because our data pertain to searches for "branded products," such as a Nikon camera. While there may be uncertainty about what the branded product offers, this uncertainty is common across retailers, and thus irrelevant to the retailer-choice decision.

looking at the advertiser's ad, because, as noted earlier, search ads do not provide much information. Clicking on an ad and visiting the retailer's website will alleviate some of the uncertainty, but not all. While the website will provide information about the offer's search attributes—attributes like price, delivery terms, returns policy, etc.—other aspects of the retailer's offer will not be resolved even after a website visit. For example, the consumer will not know whether the retailer will deliver as promised, or whether it will accept a return without hassle, etc.[8] Thus the appropriate decision criterion is maximize *expected* consumer surplus, not maximize consumer surplus.

Not all consumers interpret the consumer surplus offered by a retailer in the same way. Therefore we model consumer heterogeneity as follows:

$$q_{ai} = q_a + \epsilon_{ai},$$

where ϵ_{ai} is a user-specific shock distributed as a normal distribution. The outside option's quality is normalized to be equal to ϵ_{0i}. In addition, we allow for consumers to have different risk preferences. Let $u_i(q) = 1 - \exp(-\lambda_i q)$ denote consumer i's utility function; here λ_i is the coefficient of absolute risk aversion, which we represent as $\exp(\bar{\lambda} + \theta^{\lambda}\epsilon_i^{\lambda})$, with ϵ_i^{λ} distributed as a standard normal random variable. This formulation is convenient because expected utility from a retailer perceived to have a quality distribution distributed normally with mean μ and variance σ^2 has a simple closed- form expression

$$1 - \exp\left[-\lambda_i\left(\mu - \frac{\lambda_i\sigma^2}{2}\right)\right].$$

At the beginning, before a consumer enters a search query, she has generic uncertainty about what offers are "out there" in each of the N advertising slots. Denote these generic beliefs by a family of normal distributions $\mathbf{\Omega}^P = (\Omega_1^P, ..., \Omega_N^P)$, with means q_n^P $(n = 1, ..., N)$

[8]We thus differ from Jerath, Ma, Park, and Srinivasan (2011) model in at least two aspects. First, they conceptualize "quality" as a property of the retailer, whereas we conceptualize quality as "offer quality." In their model consumers know the qualities on offer even before clicking on an ad; a click is needed only to ascertain which retailer offers which quality. By contrast, we take a Bayesian approach: consumers start out with prior distributions about individual retailers' qualities, which they then update based on what they learn at the websites. Second, they assume that *some* consumers know advertisers' qualities even before clicking on any ads. We do not make such an assumption.

and variances $(\sigma^P)^2$. As search proceeds, perceived quality distributions get updated in light of new information. The new information comes in two stages. In the first stage, consumers observe the list of advertisers and the advertising slots they occupy. This information acquisition, we assume, is "simultaneous" and costless.[9] The second stage is when they click on individual ads and visit the advertiser's website. This information acquisition is sequential, and costly, involving a unit search cost, F_i. After each visit, deciding whether to search further amounts to solving a dynamic programming problem comparing the highest updated expected utility from the distributions searched so far against the value of searching further while incurring the search cost.

Stage 1: Observing ads. The first stage in the search process is observing the ads in the impression. For each advertisement, consumers update their beliefs $\mathbf{\Omega}^P$, using the information contained in the ads, namely, (i) advertiser identity, and (ii) ad content. We assume that the ad description provides an unbiased, but possibly imprecise, signal about unobserved retailer quality. In other words,

$$x_{ai} = q_{ai} + \eta_{ai}^D \epsilon_{ai}^D,$$

where ϵ_{ai}^D is a mean-zero Gaussian noise term signifying that the ad description is a noisy signal of underlying quality, and η_{ai}^D captures the intensity of that noise. We will later model η_{ai}^D as a function of the advertiser's brand strength.

After observing the ad signals, the generic distributions $\mathbf{\Omega}^P$ get transformed into advertiser-specific posterior distributions, $\Omega_i^D = (\Omega_{a(1)i}^D, ..., \Omega_{a(N)i}^D)$, via Bayesian updating; that is,

$$q_{a(n)i}^D = \left(\frac{q_n^P}{(\sigma^P)^2} + \frac{x_{a(n)i}^D}{(\eta_{a(n)i}^D)^2} \right) / \left(\frac{1}{(\sigma^P)^2} + \frac{1}{(\eta_{a(n)i}^D)^2} \right),$$

and

$$(\sigma_{a(n)i}^D)^2 = \left(\frac{1}{(\sigma^P)^2} + \frac{1}{(\eta_{a(n)i}^D)^2} \right)^{-1}.$$

[9]Effectively we are assuming that clicking behavior is dictated by consumers' perceptions of advertiser quality after observing all the ads, rather than a selective observation of a few ads (see Stage 1 below). In the absence of eye-tracking data, these stories cannot be separated. For example, we can't distinguish between people who don't observe the ads and hence don't click on them and people who observe the ads but decide not to click on any of them.

Note that if η_{ai}^D is large relatively to σ^P, then retailer-specific posteriors would be governed by position priors, which would generate a large position bias. On the other hand, if η_{ai}^D is negligible, we should see little position bias because users efficiently learn true quality from the ad which overrides the position-based prior.

Note that we allow η_{ai}^D to vary across consumers and across advertisers. Allowing η_{ai}^D to vary across consumers allows us to capture heterogeneity in position effects. Specifically, users with large η_{ai}^D (regardless of a) would always click from top to bottom no matter what ads are shown. Allowing η_{ai}^D to be advertiser-specific allows us to capture the possibility of some advertisers experiencing stronger position effects than others. Specifically, we allow η_{ai}^D to be different between advertisers in the Top 100 of Alexa rankings and the remaining advertisers. In particular, we set

$$\eta_{ai}^D = \bar{\eta}_1^D \mathbf{1}(\text{ALEXA}(a) \geq 100) + \bar{\eta}_2^D \mathbf{1}(\text{ALEXA}(a) < 100) + \theta^\sigma \epsilon_i^\sigma, \tag{1}$$

where ALEXA(a) is the Alexa score of advertiser a and ϵ_i^σ is user heterogeneity distributed as a standard normal random variable. This specification is motivated by the reduced form results reported earlier. As noted there, the position effect is smaller for more prominent advertisers. The structural model would produce this result if the estimation yields $\eta_1^D < \eta_2^D$. Other than that, we make no assumptions about the relationship between Alexa scores and the true quality of retailers.

Stage 2: Observing websites. Assume each click and visit to a website costs $F_i = \exp(F + \theta^F \epsilon_i^F)$, where ϵ_i^F is mean-zero normal random variable. Such a specification allows for heterogeneity, measured by θ^F, in the users' general attitude toward search advertising; users with large F_i would be less likely to click on the links no matter what ads are shown [10]. Let $\mathcal{C}$ be the set of clicked slots. For each $n \in \mathcal{C}$ consumer observes a website signal of the retailer's quality, i.e.,

$$x_{a(n)i}^W = q_{a(n)i} + \eta^W \epsilon_{a(n)i}^W$$

where $q_{a(n)i}$ is the true, but unknown, quality of retailer $a(n)$ with prior distribution $\Omega_{a(n)i}^D$ at this stage. The term $\epsilon_{a(j)i}^W$ is a signal noise distributed as a mean-zero normal random

[10] Note that very large values of F_i proxy users who do not click on any ads.

variable and η^W is a noise strength. The superscript W indicates information gained at the website. Following this observation, the updated distribution of retailers $a(n)$ quality for individual i is

$$q^W_{a(n)i} = \left(\frac{q^D_{a(n)i}}{(\sigma^D_{a(n)i})^2} + \frac{x^W_{a(n)i}}{(\eta^W)^2} \right) / \left(\frac{1}{(\sigma^D_{a(n)i})^2} + \frac{1}{(\eta^W)^2} \right),$$

and

$$(\sigma^W_{a(n)i})^2 = \left(\frac{1}{(\sigma^D_{a(n)i})^2} + \frac{1}{(\eta^W)^2} \right)^{-1}.$$

Note that the quality of the retailer is not fully revealed at a retailer's website, because, as noted earlier, online retailers have experience attributes as well as search attributes. Therefore, a consumer's buying (and not buying) decisions are all based on expected utility—the difference between websites visited and websites not visited solely being that for the former the expected utilities are based on updated distributions reflecting the search attribute information revealed in the websites whereas for the latter they are based solely on the ads.

It is convenient to denote $v_n(x^W_{a(n)i})$ as an expected utility of buying from retailer n after clicking and observing a website signal $x^W_{a(n)i}$. It can be shown that this expected utility retains a convenient exponential form, that is,

$$v_n(x^W_{a(n)i}) = \int u(q_{a(n)i}) dF(q_{a(n)i} | x^W_{a(n)i}) = 1 - \exp\left\{ -A_{ni} x^W_{a(n)i} + B_{ni} \right\},$$

where

$$A_{ni} = \lambda_i \frac{(\sigma^W_{a(n)i})^2}{(\eta^W)^2} \qquad B_{ni} = -\lambda_i \left[q^D_{a(n)i} \frac{(\sigma^W_{a(n)i})^2}{(\sigma^D_{a(n)i})^2} - \lambda_i \frac{(\sigma^W_{a(n)i})^2}{2} \right]$$

Because the user is assumed to buy from one of the websites or to take an outside option u_0 the utility of the set of clicked ads $\mathcal{C}$ is given by

$$U(\mathcal{C}) = \max\{u_0, v_n(x^W_{a(n)i})) : n \in \mathcal{C}\},$$

Note that $U(\mathcal{C})$ is a sufficient statistic to compute a continuation value. The continuation value is given by the following Bellman equation

$$V_i(\mathcal{C}, U) = \max\left\{ U, \max_{n \in N \backslash \mathcal{C}} EV_i\left(\mathcal{C} \cup n, \max\left\{ U, v_n(x^W_{a(n)i}) \right\} \right) - F_i \right\}, \qquad (2)$$

where the integration is performed with respect to the unknown website signal $x^W_{a(n)i}$. Order of clicks is determined by iterating on the above equation starting from $\mathcal{C} = \emptyset$ and $U = u_0$.

The above dynamic programming can be simplified using the results contained in Weitzman (1979). For each slot n, define the reservation utility $\bar{v}_n$ as

$$F_i = \int_{v_n^{-1}(\bar{v}_n)}^{\infty} \left[v_n(x^W_{a(n)i}) - \bar{v}_n \right] dF(x^W_{a(n)i})$$

It can be shown that the integration can be performed analytically and that $\bar{v}_n$ is a solution to

$$\begin{aligned} F_i = &- \exp(B_{ni}) \exp\left(-A_{ni} \left(q^D_{a(n)i} - \frac{A_{ni}(\eta^W)^2}{2} \right) \right) \times \\ &\left(1 - \Phi\left(\frac{v_n^{-1}(\bar{v}_n) - q^D_{a(n)i} + A_{ni}(\eta^W)^2}{\eta^W} \right) \right) + (1 - \bar{v}_n) \left(1 - \Phi\left(\frac{v_n^{-1}(\bar{v}_n) - q^D_{a(n)i}}{\eta^W} \right) \right), \end{aligned} \tag{3}$$

where Φ is a standard normal CDF. Given the set of reservation utilities $\bar{v}_n$, the optimal decision rule is relatively simple. Namely, the user clicks slots in the order of $\bar{v}_n$. After clicking a set of ads $\mathcal{C}$ the user stops clicking when the reservation utility of next potentially chosen ad slot is smaller than $U(\mathcal{C})$.

5 Estimation

We estimate the model using Simulated Generalized Method of Moments (Pakes and Pollard (1989)). Because the model allows for a rich persistent unobserved heterogeneity (including continuous types), the moments are computed using a nested full-solution approach. First we draw user-specific effects ϵ, λ_i, η^D_i, and ad signals $(x^D_{1i}, ..., x^D_{1N})$. Using these primitives, we compute the user's optimal policy by solving the Weitzman equation (3) numerically using a bisection method. Next, we draw website signals $(x^W_{1i}, ..., x^W_{1N})$ and simulate the optimal clicking behavior using the reservation utilities. We repeat this procedure 2500 times for each impression and obtain user-level micro moments as an average.

Our model has 28 unknown parameters; we identify them using the 73 conditional micro moments described at the end of this section. The moments are conditional on the search string and ad placements in an impression. Additionally, micro moments are clustered by time-stamp using $T = 599$ equally sized groups of $M = 47$ observations, which corresponds to 2-hour clustering.[11] Such a procedure allows for arbitrary correlation of random effects within each two-hour period. Formally, the estimator minimizes the objective function

$$Q_T(\theta) = \left[\frac{1}{T}\sum_{t=1}^{T} G_t(\theta)\right]' W \left[\frac{1}{T}\sum_{t=1}^{T} G_t(\theta)\right],$$

where G_t are clustered micro moments and W is a weighting matrix. In particular,

$$G_t(\theta) = \frac{1}{M}\sum_{m=1}^{M} \int_{\epsilon,\delta} g_{tm}(\theta|\epsilon,\delta)d(\epsilon,\delta) - \bar{g}_{tm},$$

where $\bar{g}_{tm}$ is the vector of micro moments observed in the data for impression m in cluster t. The integral of random coefficients is estimated using a frequency estimator with L IID draws from the distribution of (ϵ,δ).

Suppose that θ_0 is a true parameter vector. We make standard assumptions about the identification of the model, that is $E[G(\theta_0)] = 0$, $E[G(\theta_1)] \neq 0$ if and only if $\theta_1 \neq \theta_0$. Moreover we assume that the following limit statements are true

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\partial G_t(\theta_0)}{\partial \theta'} \xrightarrow{P} D$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t(\theta_0) \xrightarrow{d} N(0,\Omega)$$

Furthermore, suppose that $\hat{\Omega}_T^{-1}$ is a consistent estimator of Ω, then if we set the weighting matrix to $\hat{W}_T = \hat{\Omega}_T^{-1}$ our estimator is asymptotically normal with variance $(D'\Omega^{-1}D)^{-1}$ as T approaches infinity.

To obtain an estimator of Ω we need to make further assumptions about correlation of G_t across clusters; that is, we assume that G_t follows an MA(d) process. Considering that each cluster t covers roughly two hours of data we set d equal to 6. This way random effects and ad placements at the beginning of the day could be correlated with random effects and ad placements and the end of the day. Consequently, we use Newey and West (1987)'s estimator of Ω, given by

$$\hat{\Omega}_T = \left(1+\frac{1}{L}\right)\left[\hat{\Omega}_0 + \sum_{j=1}^{d}\left(1-\frac{j}{d+1}\right)(\hat{\Omega}_j + \hat{\Omega}_j')\right],$$

[11]We tried many different levels of clustering and find negligible differences in the asymptotic distribution of the estimates. For this reason, we decided to report the clustering level that makes the least assumptions about the distribution of the random effects.

where

$$\hat{\Omega}_j = \frac{1}{T-j} \sum_{t=j+1}^{T} G_t G'_{t-1}.$$

Note that we correct the variance by $1 + \frac{1}{L}$ because we use L simulations to compute the moments (see McFadden (1989) and Pakes and Pollard (1989)).

In practice, we perform the estimation in several steps. We start with a manually calibrated starting point, $\hat{\theta}^{(0)}$, in which all ads have zero quality and all standard deviations are 1. Also, we set $\hat{W}^{(0)}$ equal to the identity matrix. Next, we obtain $\hat{\theta}^{(1)}$ using GMM and obtain a new weighting matrix $\hat{W}^{(1)}$. We repeat these steps until convergence. To minimize the computation time we gradually increase the number of draws, L, until we reach 2500.

6 Results

Table 5 contains the estimates of consumer clicking cost (F_i) and coefficient of absolute risk aversion (λ_i). Average clicking cost is about -2.29, and its standard deviation is 9.63. The 90th, 70th and 50th (median) percentiles of the distribution are estimated to be -0.06, -0.21 and -0.53, respectively, all significantly greater than the mean, -2.29. The average consumer is risk averse with a coefficient of absolution risk aversion equal to 1.9. Again, as with the clicking cost, there is heterogeneity in this coefficient: the standard deviation of λ is estimated to be 6.93 and the median is 0.5.

Clicking cost		Risk aversion	
$E[F_i]$	$\text{std}[F_i]$	$E[\lambda_i]$	$\text{std}[\lambda_i]$
-2.29 (0.033)	9.63 (0.120)	1.90 (0.016)	6.93 (0.058)

Table 5: Estimates of clicking cost and risk aversion (distributed as log-normal)

In Table 6 we report estimates of the prior beliefs $\mathbf{\Omega}^P$ consumers have about the ads in each position. As can be seen, the position priors are monotonic with lower quality estimates assigned to lower positions. This is consistent with the position effect we observed in the raw click-through-rates. Note that the position prior declines steeply for positions 1 through 4

Ad position prior means q_n^P					std., σ^P
Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 5	
0.01 (0.047)	−0.11 (0.043)	−0.24 (0.042)	−0.49 (0.043)	−0.48 (0.036)	1.05 (0.050)

Table 6: Estimates of the ad position prior $\mathbf{\Omega}^P$

and subsequently flattens out. To evaluate the magnitude of the position priors it is useful to compare them to the outside option (clicking on an organic link or shopping outside the Microsoft Live platform). We find that the ad in position 1 is thought to be as good as the outside option while ads in lower positions are on average inferior to the outside option.

Table 7 contains information about the “true quality” of retailers, that is, retailer fixed effects. We note that the majority of the retailers have lower quality than the outside option with the exception of advertiser 3 in the Nikon keywords, advertiser 1 in the Canon keywords and advertiser 4 in the Olympus keywords. Overall, we observe large differences in quality among retailers, and in particular, these differences are much larger than the difference in the position effects. Moreover, the true quality of advertisers is usually lower than the position priors, which suggests that people overestimate the quality of retailers advertising on search engines before seeing them. We suspect that either our keywords have lower quality advertisers than the average keyword, or that sponsored search ads on Microsoft Live Search are of lower quality than sponsored search ads on other platforms.

It is interesting to compare the CTRs reported in Table 3 with the quality estimates reported in Table 7. We observe that these numbers do not align perfectly. One example is advertiser 4 in the Olympus search string. This retailer has the highest average quality, but it does not enjoy the highest CTR. Similarly, we observe a large difference in quality between advertisers 2 and 3 on the Nikon search string, but their difference in CTRs is negligible. The fundamental reason for these discrepancies is that CTRs are determined by the Weitzman (1979) “reservation values,” and these reservation values do not have a simple relationship to “average qualities.” In fact, this was the whole point of the Weitzman paper—that it is not, in general, optimal to search in the order of expected values. [There are at least two reasons for these discrepancies. First, observed CTRs are contaminated by position effects. Then, if

	Search string		
	Nikon	Canon	Olympus
advertiser 1	-0.33 (0.153)	0.99 (0.212)	-0.74 (0.079)
advertiser 2	-0.28 (0.115)	-1.19 (0.083)	-0.88 (0.103)
advertiser 3	0.86 (0.273)	-1.50 (0.074)	-2.04 (0.224)
advertiser 4	-1.26 (0.230)	-0.37 (0.192)	1.15 (0.177)
Other advertisers	-1.16 (0.097)	-1.35 (0.091)	-1.32 (0.075)

Table 7: Estimates of the average qualities of the retailers by search string.

Description signal std.			Website signal std., η^W
Alexa$\leq$ 100, $\bar{\eta}_1^D$	Alexa$>$ 100, $\bar{\eta}_2^D$	Heter. std., θ^σ	
1.28 (0.100)	2.02 (0.206)	0.35 (0.103)	0.89 (0.122)

Table 8: Estimates of the signaling model

ad position is correlated with (unobserved) true quality, CTR would provide a biased measure of quality. When estimating quality scores we correct for this endogeneity using advertiser fixed effects. Second, CTRs do not take into account branding effects described in Section 2, which might be important if branding is correlated with true quality. We correct for this bias by adding Alexa data, and allowing for Alexa score fixed effect specification described in equation (1).]

Table 8 contains estimates of the parameters of the ad and website signal. The main quantity of interest is the difference between $\bar{\eta}_1^D$ and $\bar{\eta}_2^D$, which describes the amount of extra information about quality carried by the advertiser brand. In particular, the precision of the ad signal for "branded" (Top Alexa) advertisers equals 0.78, while the precision of the ad signal for non-branded ads is only 0.5. These estimates imply about a 36% improvement in signal precision if the ad is from a branded advertiser. Comparatively, the precision of the position prior is about 0.95 and precision of the website signal is about 1.25. Thus, we can say that advertiser branding brings about as much extra precision to the ad signal as the difference in precision between the website signal and the position prior. It is consistent with our conjecture that the advertiser brand communicates some information at the ad stage itself, which, absent advertiser identification, would be gained only on the advertiser's website.

In the next section we conduct a series of counterfactual experiments to determine the CTR implications of the difference between branded and non-branded advertisers.

7 Counterfactuals

In order to quantify the impact of branding on consumer behavior at the sponsored search impression, we perform a series of counterfactual experiments. The structural model enables us to separate the impact of brand on the consumer preferences. In consequence, we are able to perform experiments that manipulate branding, while keeping all other characteristics of consumers and advertisers constant. Such manipulations identify the causal impact of branding on the CTRs and give the managers a tool to evaluate the value of the brand and prominence in the context of sponsored search advertising.

First we perform the experiments that allow to isolate the informational value of branding keeping the true quality of the advertiser constant. Usually it is hard to decouple branding from quality, and thus, manipulate the branding without affecting the quality. However, our structural model allows us to compute the click-through-rates of the same advertiser with and without the branding keeping everything else constant.

The Advertiser 3 in a "Nikon" search string is a website that obtains the highest CTR, despite being non-branded. We find that this advertiser would enjoy additional 5 percentage points CTR if he was branded. In comparison, a low quality branded Advertiser 4 would lose about 0.2 percentage point of the CTR is his branding was nullified. Similar regularity can be observed in a "Canon" search string. Endowing a high quality Advertiser 1 with a brand increases his CTR by 4 percentage points, while stripping the brand from Advertiser 2 penalizes him by about 0.3 percentage point. [12] These effects suggest that branding

[12]We believe that lower quality of chosen branded advertisers is a phenomena isolated to search strings in

and quality are compliments. The intuition is that, if the consumers are rational, then the informational function of the brand has more value to the firms that are able to communicate higher quality.

The next set of experiments are aimed at evaluating the interaction between branding and position for high quality advertisers. Such experiments are of a particular interest in two canonical situations: (i) if a high quality advertiser enjoys strong branding and wants to know if he should additionally invest in prominence (that is, bid aggressively for a top position), or (ii) if a high quality advertiser does not have a strong brand and wants to know if he should invest in prominence instead. Thus, we need to estimate the marginal impact of the position on the CTR[13] for strongly and weakly branded advertisers.

For the purpose of estimating the marginal value of the position we use a structural model to predict CTRs on a set of counterfactual impressions. In particular, we assemble impressions in which a chosen highest quality advertiser faces 3 identical low quality and weakly branded competitors ("Other domains") [14]. Subsequently, we manipulate the position and branding of the high quality advertiser and compare its CTRs between different scenarios. The branding manipulation consists of varying the advertisers' Alexa score between Top100 (strong brand) and below Top100 (weak brand). In practice, the branding manipulations are relatively straightforward and amount to changing the precision of the description signal from $\bar{\eta}_1^D$ to $\bar{\eta}_2^D$, while keeping other parameters constant.

The results of the experiments are presented in Table 9. The first row consist of manipulations for "Nikon" keyword. First column contains a figure describing differences between the CRT at positions 1-3 (top positions) and position 4 (bottom position) for Advertiser 3.

this study. In fact, if we examine all 12 domains, we find no relationship between branding and quality. To the contrary, if we examine other keywords such as "games", and "white pages", we find that branding is associated with higher quality. Despite this difference the key results of this paper are qualitatively the same when using "games" and "white pages," thus the relationship between quality and branding does not seem consequential for the external validity of this paper.

[13]We do not observe markups and use CTR as a proxy. This approach is a good approximation if the advertiser does not change its price when updating his sponsored search bid. In case the markups are observed they can be easily incorporated into the exercise.

[14]We tried to manipulate the quality and branding of competitors but the results and conclusions are qualitatively the same.

In order to demonstrate the interaction between branding and marginal value of higher position, we repeat the exercise for strong of weak branding of Advertiser 3. We find statistically significant differences between the marginal value of higher position and branding that are consistent with the reduced form results from Section 3. Namely, the prominence (higher position) is more valuable on the margin if advertiser has a weak brand. It confirms that prominence and branding are substitutes.

We note that the nominal differences of the marginal effects of prominence between weak and strong brand may be misleading because strongly and weakly branded ads receive different baseline CTR at position 4. Particularly, in the pay-per-click mechanism the relevant statistic for revenue maximization is the elasticity of CTR with respect to the position, or in other words, the percentage change in CTR when moving to a higher position. We compute these quantities and present them in the second row of Table 9. In this case, we note much larger differences in the value of prominence between strongly and weakly branded conditions. Namely a weakly branded advertiser can increase his CTR by nearly 150% by moving from position 4 to position 1, while strongly branded advertiser can increase its CTR only by about 50%. It suggests that (given similar markups) weakly branded advertisers have more incentive to bid aggressively than strongly branded ones.

We repeat the exercise for "Canon" and "Olympus" keywords and demonstrate that similar effects are present (see row 2 and 3 of Table 9). We find that "Olympus" keyword has a smaller difference between branding conditions that other keywords, which mirrors the reduced form results from Section 3 and provides a convenient validation of the structural estimation. We further note that the structural approach is more efficient in finding the effects of branding than the reduced form approach. In case of "Olympus" the branding interactions were difficult to establish with regressions, whereas with a help of theoretical model we able to uncover such regularities much easier. It demonstrates a usefulness of the structural approach for actual business decision making, that supplements classical reduced form regressions. That is, the approach is particularly useful in high noise environments (studying the effects of branding notoriously suffers from high noise), when the researcher wants to trade off making reasonable theoretical assumptions for greater efficiency.

Another way to assess the interaction between branding and position effects is to estimate

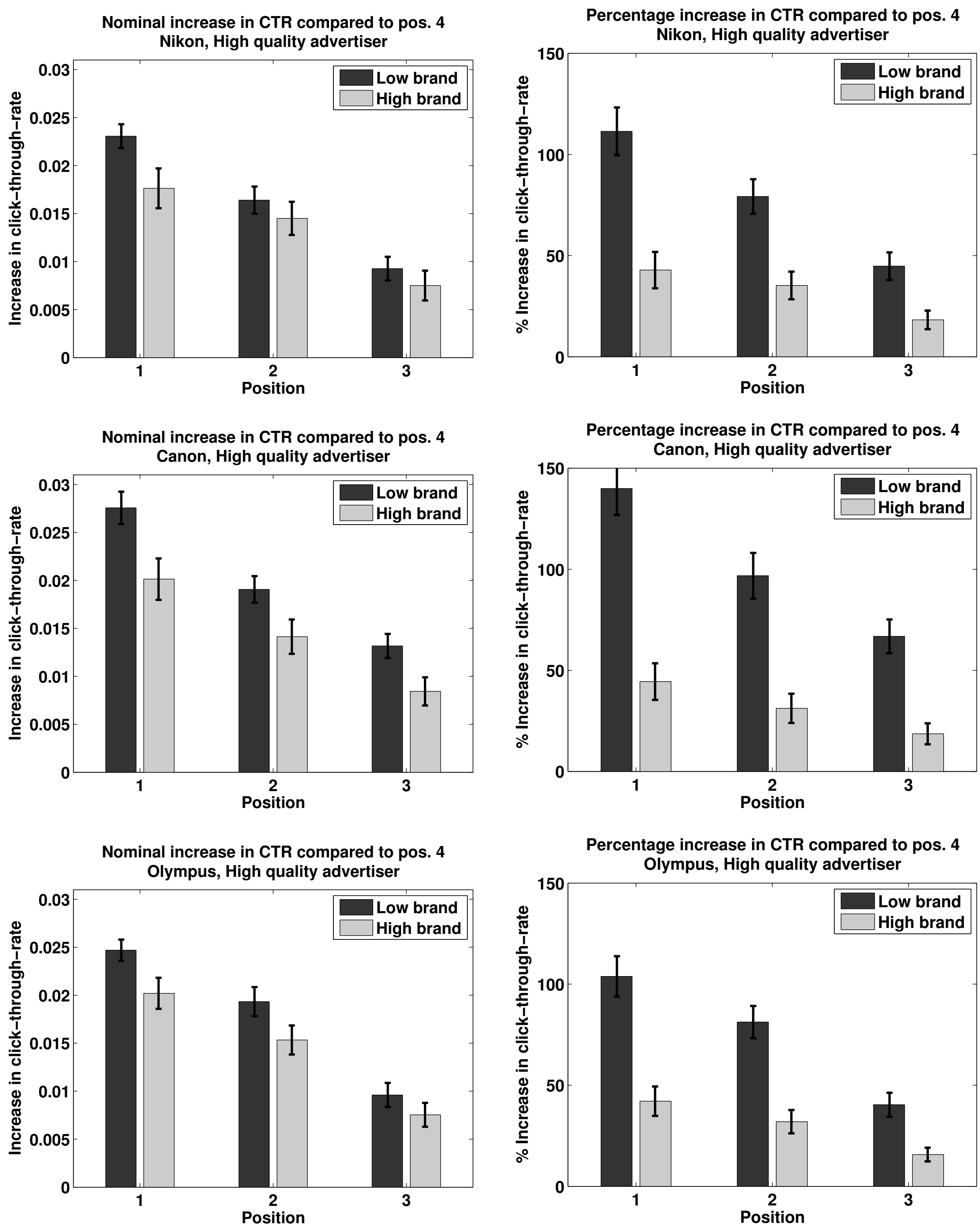

Table 9: The CTRs were calculated in the counterfactual scenarios where the chosen high quality retailer is endowed with either high or low Alexa score, and where competitors are endowed with low Alexa score. Standard errors were obtained by the parametric bootstrap from the asymptotic distribution of the estimates.

how branding affects market power generated by higher positions. In our context, we define market power as the ability to profitably lower the quality of the website. Specifically, the market power generated by the position is the ability to profitably lower quality when displayed on the higher position.

In order to directly measure the market power from position one would need to compare the optimal quality at the lower position to the optimal quality at the higher position for a given advertiser. In order to compute these optimal qualities one needs marginal cost information, which is unavailable in our data. Instead, we compute a proxy for the market power, that is, we compute by how much the advertiser can lower quality when moving to the higher position without losing click-through-rate. For that purpose, we compare two counterfactual impressions. In the first impression we place a focal high quality advertiser in position 2 and place 3 low quality competitors in positions $1, 3$ and 4. In the second impression we place the same focal advertiser in position 1 and lower its quality by a factor of Δ. We estimate the market power as Δ that generates the same CTRs of the focal advertiser in both impressons. We find that position generates considerable amount of market power if the focal advertiser is unbranded. For example, Advertiser 3 in "Nikon" keyword can lower its quality by 59% in position 1 and maintain the same CTR as in position 2. However, if this advertiser is branded, a higher position enables him to lower quality by only 14%. Similarly, Advertiser 1 in "Canon" keyword can lower the quality by about 50% and maintain the same CTR when unbranded, but only by about 20% when branded. We conclude that branded advertisers increase their market power by significantly less when moving up on the search page than the unbranded ones. Note that it does not mean that advertisers do not want to be branded, it simply means that the market power coming from the brand is dwarfing the market power coming from the position.

8 Robustness Analysis

There are scenarios in which using domain fixed effects could be problematic. For example, if the advertisements were targeted on the individual level, domain fixed effects would not be sufficient and domain-user fixed effects would be required. Also, if advertiser qualities were changing and were systematically correlated with changes in the position the estimation would require time-domain fixed effects. Narayanan and Kalyanam (2014) discuss the latter mechanism in detail and mention, among others: (i) strategic interactions caused by sophisticated bidding and bid throttling software, (ii) promotions conducted by the advertisers that may be associated with changes in their bidding strategy. However, according to anecdotal evidence, in 2007, advertisers changed their bids quite infrequently and automated bidding software was not available. Also, 2007 is the first year of operation of Microsoft Live market place. For this reason, the quality scores of advertisers are noisy, which creates considerable amount of exogenous ad placement variation.

In order to confirm that the domain fixed effects are appropriate to control for position endogeneity in our data, we conduct a robustness check using a regression discontinuity design inspired by Narayanan and Kalyanam (2014). Unfortunately, we cannot implement their exact design, because we do not observe the advertisers' quality scores. Instead, we single out impressions in which the same focal advertiser is presented in different positions, and that are generated within a short time interval of each other. Subsequently, we estimate a linear probability model after first-differencing the observations for these adjacent impressions. Under the mild assumption that the ad quality is constant within the short interval, this procedure produces estimates of position effects that are robust to position endogeneity. The results for a "Nikon"keyword, and a domain-fixed-effects-estimator for comparison, are presented in Table 10. We find that the position effects estimated using the first-differences are not economically different from the estimates obtained using domain fixed effects. Moreover, in only one out of nine comparison cases we find that this difference is statistically significant. We repeat the same analysis for "Canon" and "Olympus" keywords and arrive at similar results. Thus, in the remainder of the paper we apply domain fixed effects design.

9 Conclusion

This paper develops a structural model of consumers' response to search ads, and estimates it using individual-level clickstream data from Microsoft Live Search. Our model is an empirical operationalization of the search model seen in Edelman, Ostrovsky, and Schwarz (2007): the

	Within 5s	Within 1h	Within a day	Fixed effects
Position 1	0.055 (0.019)	0.049 (0.002)	0.058 (0.001)	0.052 (0.002)
Position 2	0.049 (0.014)	0.032 (0.002)	0.032 (0.001)	0.032 (0.004)
Position 3	0.009 (0.013)	0.017 (0.002)	0.014 (0.001)	0.017 (0.006)

Table 10: Robustness check of the fixed effects estimation using a regression discontinuity design applied to a linear probability model.

consumer's click-through behavior is modeled as a directed search among search ads. Our main contribution is to recognize that direction is provided not only by ad position, but also by brand identity. This recognition helps explain why consumers do not always click from top to bottom, and why particular advertisers in lower positions generate more clicks than other advertisers in higher positions. Thus, our results corroborate the "position paradox" discussed by Jerath, Ma, Park, and Srinivasan (2011).

Our most noteworthy results have to do with (1) how ad position and advertiser brand interact in determining click-through-rate, and (2) how click-through-rate correlates with unobserved advertiser quality. Regarding the first, we find that ad position and advertiser brand are substitutes, not complements, contradicting the common assumption in the theoretical literature (e.g., Aggarwal, Goel, and Motwani (2006), Katona and Sarvary (2010)). Marginal value of higher ad position is higher for a weaker brand, not for a stronger brand. Regarding the second result, we find that click-through-rate and advertiser quality are positively correlated. Since quality essentially means lower price for retailers, this result suggests that search advertisers are unable to gain market power from ad prominence of either kind.

Taken together these results point to the similarities and differences between search advertising and traditional forms of advertising such as TV or print. In gaining attention, it is well understood that the effectiveness of TV and print ads depends on how many other ads, and which other ads, are vying for attention. It is also well understood that the value of ad prominence—such as a front-page ad, or an ad in the Super Bowl—is higher for an up-and-coming-brand than for an established brand. It is reassuring that search ads do not deviate from these well-worn maxims. On the other hand, we are somewhat surprised by the finding that ad prominence for search ads doesn't seem to translate to market power. It may be that the consumer of search ads is more price-sensitive, or it may be that search ads just do not have the power to engage, inform, or persuade like TV or print ads.

References

Ackerberg, D., C. Lanier Benkard, S. Berry, and A. Pakes (2007): "Econometric tools for analyzing market outcomes," *Handbook of econometrics*, 6, 4171–4276.

Aggarwal, G., A. Goel, and R. Motwani (2006): "Truthful auctions for pricing search keywords," in *Proceedings of the 7th ACM conference on Electronic commerce*, pp. 1–7. ACM.

Athey, S., and G. Ellison (2011): "Position Auctions with Consumer Search," *The Quarterly Journal of Economics*, 126(3), 1213–1270.

Athey, S., and D. Nekipelov (2010): "A structural model of sponsored search advertising auctions," in *Sixth Ad Auctions Workshop*.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2), 242–262.

Brooks (2006): "The Atlas rank report: How search engine rank impacts traffic," Discussion paper, Atlas Institute.

Craswell, N., O. Zoeter, M. Taylor, and B. Ramsey (2008): "An experimental comparison of click position-bias models," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 87–94. ACM.

Edelman, B., M. Ostrovsky, and M. Schwarz (2007): "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, 97(1), 242–259.

Jerath, K., L. Ma, Y.-H. Park, and K. Srinivasan (2011): "A "position paradox" in sponsored search auctions," *Marketing Science*, 30(4), 612–627.

Jeziorski, P., and I. Segal (2014): "What Makes them Click: Empirical Analysis of Consumer Demand for Search Advertising," .

Katona, Z., and M. Sarvary (2010): "The race for sponsored links: Bidding patterns for search advertising," *Marketing Science*, 29(2), 199–215.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010): "Online Demand Under Limited Consumer Search," *Marketing Science*, 29(6), 1001–1023.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57(5), pp. 995–1026.

Narayanan, S., and K. Kalyanam (2014): "Position Effects in Search Advertising: A Regression Discontinuity Approach," .

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Oremus, W. (2013): "Google's Big Break," *Slate*.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–57.

Rutz, O. J., R. E. Bucklin, and G. P. Sonnier (2012): "A latent instrumental variables approach to modeling keyword conversion in paid search advertising," *Journal of Marketing Research*, 49(3), 306–319.

Varian, H. R. (2007): "Position auctions," *International Journal of Industrial Organization*, 25(6), 1163–1178.

Weitzman, M. L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, 47(3), pp. 641–654.