What Makes them Click: Empirical Analysis of Consumer Demand for Search Advertising*

Przemyslaw Jeziorski[†] Ilya Segal[‡]

June 10, 2009

Abstract

We study users' response to sponsored-search advertising using data from Microsoft's Live AdCenter distributed in the "Beyond Search" initiative. We estimate a structural model of utility maximizing users, which quantifies "user experience" based on their "revealed preferences," and predicts user responses to counterfactual ad placements. In the model, each user chooses clicks sequentially to maximize his expected utility under incomplete information about the relevance of ads. We estimate the substitutability of ads in users' utility function, the fixed effects of different ads and positions, user uncertainty about ads' relevance, and user heterogeneity. We find substantial substitability of ads, which generates large negative externalities: 50% more clicks would occur in a hypothetical world in which each ad faces no competition. As for counterfactual ad placements, our simulations indicate that CTR-optimal matching increases CTR by 15% while user-optimal matching increases user welfare by 25% (and neither coincides with assortative matching). Moreover, targeting ad placement to specific users could raise user welfare by 60%. Finally, user welfare could be raised nearly 15% if they had full information about the relevance of ads to them.

*The authors are grateful to Microsoft Corp. for providing the data and computing facilities and hosting them during the summer of 2008. The second author also acknowledges the support of the Toulouse Network for Information Technology.

[†]Graduate School of Business, Stanford University, Stanford CA 94305

[‡]Department of Economics, Stanford University, Stanford CA 94305

1 Introduction

Over the past decade the Internet has become the dominant channel for consumer information about goods and services. A substantial fraction of this information is provided through Internet advertising. In 2007, Internet advertising revenues rose 26 percent to reach $21.2 billion, according to the *Internet Advertising Revenue Report* published by the Interactive Advertising Bureau and PricewaterhouseCoopers LLP[1].

To gain understanding of the online advertising market, compare alternative market structures and designs, and examine their welfare effects, it is important to understand the behavior of consumers in this market. Our paper makes a step in this direction, focusing on "search advertising," i.e., "sponsored links" that accompany results produced in response to the consumers' search queries. Search advertising accounts for 41% of the total Internet advertising revenues. It is viewed as the most effective kind of advertising because of its very precise targeting: a consumer's search string reveals a great deal about the products (s)he is likely to be interested in. This precise targeting allows to display only the most relevant ads, which in turn induces consumers to click on the ads. While the market for search advertising has recently received a lot of attention, not much is known about consumer behavior in the market. This paper makes a step towards remedying this problem.

Existing papers on search advertising postulate very simple and restrictive models of user behavior. For example, Edelman, Ostrovsky, and Schwarz (2007) propose a model that assumes that the CTR (clickthrough rate) on a given ad in a given position is a product of ad and position specific effects and does not depend on which other ads are displayed in the other positions. (Henceforth we will refer to this model as the "EOS model," which is also used in Edelman and Ostrovsky (2007), Varian (2006), Yenmez (2009), Gomes and Sweeney (2009), Ostrovsky and Schwarz (2009)). In the "cascade model" (Craswell, Zoeter, Taylor, and Ramsey (2008), Papadimitriou and Zhang (2008)), users consider the ads sequentially from top to bottom, deciding whether to click on the current ad and whether to continue clicking with ad-specific probabilities. These restrictive models have not been compared with actual user behavior. Also, as these models have not been derived from utility-maximizing behavior of users, they could not be used to evaluate user welfare.

This paper offers the first empirical investigation of user response to sponsored-search advertising that is based on a structural model of utility-maximizing user behavior. One advantage of a structural model over reduced-form models is that once the model's parameters are estimated

[1] http://www.scribd.com/doc/4787183/Internet-advertising-revenue-report-for-2007

and its fit with the data is established, the model can be used to predict user behavior for all conceivable counterfactual ad impressions. Another advantage of the model is that it quantifies the "user experience" on a sponsored-search impression as users' expected utility from the impression, and estimates this utility from the preferences of actual users revealed by their clicking behavior, rather than from the judgements of disinterested experts (as in Carterette and Bennett (2008)).[2] Improving user experience is crucial for the survival and growth of an internet platform, and our model can be used as a guide toward that goal.

Our dataset offers a selection of advertising impressions and user clicking behavior on Microsoft's *Live Search* advertising engine. The data contains a random selection of search sessions between August 10 and November 1, 2007. In each session, the user entered a search string and was then shown "organic" search results accompanied by advertisements ("sponsored links"). An advertising "impression" is an ordered list of sponsored links. (The first sponsored link is displayed at the top of the page in a highlighted box, while the others are displayed in a column to the right of the organic search results.) For each advertising impression, our data describes the ads clicked by the user and the times at which the clicks occurred.

Our estimation strategy is based on the fact that searches on the same search strings often generate different advertising impressions. We treat this variation in impressions as exogenous and uncorrelated with users' characteristics. Indeed, we have been assured that the impressions were not conditioned on the user's known characteristics or browsing history. We also make the crucial assumption that the characteristics of ads that determine users' values for them did not vary over our 3-month window. This assumption appears plausible for the four search strings we consider: "games," "weather," "white pages," and "sex".[3] In fact, it is easy to convince oneself of the large

[2] Dupret and Piwowarski (2008) quantify ad quality by calibrating a heuristic model of user behavior on real data. However, since their model is not based on utility maximization, it cannot be used to quantify user welfare.

[3] To understand the importance of this assumption, imagine that the preferences of users searching for "Paris Hilton" changed abruptly from looking for a hotel in the capital of France to looking for the infamous sex video, and that the advertising engine quickly responded to this preference change by changing the placement of ads. In this situation, our estimation strategy would be invalid: for example, it might wrongly find that putting an ad in the top position raises its CTR when in fact it may just be that the engine puts the most relevant ad at the top and there is no position effect for any given ad.

Microsoft plans to release a dataset in which ad impressions are truly randomized and independent of ad characteristics – an initiative known as the "adCenter challenge:"

http://research.microsoft.com/workshops/ira2008/ira2008_talk.pdf

Repeating our analysis on this dataset would eliminate any possible concerns about the endogeneity of impressions.

random component in ad placement by searching for the same search string several times in a row. The ad placement results from several fast-changing factors, such as advertisers' varying bids and budgets, the advertising engine's estimate of the ad's relevance based on its historical clickthrough rate (CTR), and explicit experimentation by the engine. We believe that at least on our search strings, this randomness swamps any possible shifts in the ads' relevance.

We begin by examining reduced-form evidence that contradicts the existing theoretical models and suggests some dimensions in which the models need to be enriched. In particular, the EOS model is contradicted by the prevalence of externalities across ads: the CTR on a given ad in a given position depends on which ads are shown in other positions. For example, the CTR of *Domain 1* in position 2 on the "white pages" search string is 18% if its competitor in position 1 is *Domain 3* (which is not a good match for "white pages"because offers yellow pages), but drops to 8% if the competitor is *Domain 2* (which is a specialized advertising company).[4] This difference is statistically significant. The "cascade model" is contradicted by the observation that 46% of the users who click on ads do not click sequentially on positions (1,2,. . .), and 57% of the users who click more than once do not "cascade," i.e., click on a higher position after clicking on a lower position. Also, the data exhibits certain kinds of externalities that could not emerge in the cascade model: the CTR on a given ad in a given position depends on which ads are shown below it, and the CTR on a given ad in position 3 given the two ads shown in position 1 and 2 still depends on the order in which the two ads.

Next, we formulate and estimate a structural model of rational user behavior that nests the existing models. In our model, a user chooses his clicks sequentially under uncertainty about the relevance of ads to him. The model is related to the literature on consumer search (e.g., Hong and Shum (2006), Hortacsu and Syverson (2004)), the closest work being Kim, Albuquerque, and Bronnenberg (2009), which estimates online search for durable goods at Amazon.com. The latter paper assumes full satiation: a consumer gets utility from at most one purchase. Our model instead parameterizes the degree of substitutability (satiation) among ads with a parameter R in a "Constant Elasticity of Substitution" utility function. For $R = 0$, user utility is the sum of the utilities derived from the clicked ads, and so there are no externalities across ads, as in the EOS model. At the other extreme, when $R = \infty$, user utility is the maximum of the values of the ads he clicks on, and so he derives utility from at most one ad, and the externalities are the most prominent

[4]The domain names are available in the dataset by Microsoft does not allow us to publish them to protect advertiser privacy.

(similar to Kim, Albuquerque, and Bronnenberg (2009)).

In addition to the substitutability parameter R, our model includes a number of other parameters. For each of the more common ads, we estimate two fixed effects: the probability of being relevant and the user's utility derived if it does turn out to be relevant. Thus, we separately estimate users' expectation of a given ad's value and their uncertainty this value. Also, we include fixed cost effects for clicking different positions, to capture the fact that the same ad receives more clicks when shown in a higher position. We also allow for user heterogeneity by incorporating a user-specific random utility effect whose variance we estimate. This effect proves important to fit the data, in which some users click on many ads and others on few or none.

We find that externalities are both statistically and economically significant. Our estimate of the substitutability parameter R is 0.55. Using the estimate, we predict that the CTR on most domains in the hypothetical world without externalities would have been substantially higher than their actual CTR. We predict that the total number of clicks in our dataset would have been 50% higher had satiation been absent. We find evidence of user uncertainty: if this uncertainty were resolved prior to clicking, consumer welfare would be nearly 15% higher (and there would be no significant effect on the total CTR).[5]

We use our estimated model to predict user behavior on counterfactual ad impressions and generates impressions that maximize the total CTR or the expected user welfare. It is well known that in the EOS model, the total CTR is maximized by assortative matching of higher-quality ads to better positions. The same is true in our model without externalities ($R = 0$), furthermore, in this model the same assortative matching also maximizes expected user welfare (where "quality" is interpreted as expected quality, and user uncertainty about quality does not matter). However, in the estimated model with satiation ($R > 0$), such assortative matching, while providing substantial improvement over the status quo, does not exactly maximize either the total CTR or the expected user welfare. Namely, we find that assortative matching would raise user welfare by 18% and the total CTR by 17%. However, the user-optimal placement policy would raise user welfare by 25% while the CTR-maximizing policy would raise the CTR by 18% (the two policies no longer coincide). The intuition for how assortative matching can be improved upon for users is that for two ads with the same expected quality it is strictly optimal to put the ad with higher user uncertainty in a higher

[5]Note that in our model of expected utility maximization, cardinal utility has empirical meaning: impression A being x% better than impression B means that the user is indifferent between receiving impression B for sure and receiving impression A with probability $\frac{x}{100+x}$ and no ads at all with the complentary probability.

position so as to reduce the user's cost of learning its relevance. (Similar changes raise the CTR but to a smaller extent.) This intuition is similar to the idea of accounting for valuing "diversity" in ranking organic search results by advantaging results that are less likely to be relevant but are very useful if they do prove to be relevant.[6]

It has also been suggested that user experience and CTR could be improved by targeting the impressions to individual users, e.g., based on their browsing history or demographics (E.g., see Radlinski and Dumais (2006) for a discussion of targeting and ad diversity.) We can bound above the gain in user welfare and CTR that could be achieved by such targeting, by simulating "first-best" targeting based on the users' individual characteristics. We find that user-optimal first-best targeting could raise user welfare by 60%, while CTR-optimal first-best targeting would raise the total CTR by 48%.

Athey and Ellison (2007) (henceforth AE) model user learning about general ad relevance in the course of a search session: upon learning the relevance of a clicked ad, the user updates his beliefs about the relevance of the other ads in the same impression.[7] Our paper ignores this kind of updating. There are two reasons for this: (1) Such updating would generate positive "informational externalities" across ads – i.e., an ad would benefit from having better ads in the same impression. Empirically we find that the overall externalities are instead negative, and it would be difficult to separately identify both satiation externalities and informational externalities from the available data. (2) We believe such updating to be a long-run rather than a short-run phenomenon. As consumers use a given search engine frequently, we don't expect much learning about relevance to occur in the course of a single session (as assumed in Athey and Ellison (2007)). While long-run

[6]These counterfactuals can be reliably interpreted as "short-run" counterfactuals. If an alternative placement policy is implemented in the long run, one may ask if this will change any of the model's parameters. We think that the only parameters that may be subject to change are the position fixed effects, which capture the fact that users are much more likely to click on the same ad in a higher position. If these fixed effects stem from (psychic) clicking costs, then they should not depend on the matching policy. If instead these effects are due to users' expectations that higher positions contain more relevant ads, these expectations will be altered in the long run under the new matching policy.

[7]In their basic model, the ads' texts are uninformative, and so the CTR on a given ad depends on the information learned from clicking on the preceding ads, but not on the ad itself. User behavior in this model is similar to that in the "cascade" model, with the added feature that the probability of continuing after clicking a given ad depends not just on this ad's quality but also on the qualities of the ads above it (which determine user beliefs about the quality of subsequent ads). Just as the cascade model, the AE model is inconsistent with non-sequential and non-cascading clicks and with externalities from below.

learning over the course of many sessions may prove to be very important, we are unable to observe it in our data which does not keep track of user histories (for privacy reasons).

The paper is organized as follows. In Section 2 we describe the dataset and examine some reduced-form evidence. Section 3 describes the model. Section 4 describes identification and estimation. Section 5 discusses the estimation results. Section 6 simulates counterfactual matching policies. Section 7 concludes.

2 The data and its preliminary analysis

2.1 Data Description

Our dataset offers a selection of advertising impressions and user behavior on Microsoft's *Live Search* advertising engine. As of May 2008, *Live Search* had 9.1% of the U.S. online search market (as compared to the market leader Google's 61.6%).[8] This modest market share nevertheless translated into about 900 million search queries per month. This enormous data is generally not available to external researchers, primarily for fears of compromising user privacy. However, in 2008, Microsoft created a DVD with a sample of user search and advertising data, cleaned up to eliminate privacy-compromising information. This DVD was distributed to a few dozen recipients of Microsoft's external research grants, as well as to a small number of other researchers, including the authors of this paper.

The data on the DVD contains a random selection of search sessions between August 10, 2007 and November 1, 2007. In each session, the user entered a search string and was then shown "organic" search results accompanied by advertisements ("sponsored links"). An advertising "impression" is an ordered list of sponsored links. The first sponsored link is displayed at the top of the page in a highlighted box, while the others are displayed in a column to the right of the organic search results. For each sponsored link the user was shown a text display containing the advertiser's domain name as well as brief advertising copy. For example, one ad produced in response to a search for "weather" reads

```
Local Weather Forecast
Get Live Weather Forecasts & More With The Free Weather Toolbar
Weather.alot.com
```

[8] http://www.techcrunch.com/2008/05/22/the-empire-strikes-back-our-analysis-of-microsoft-live-search-cashback/

For each advertising impression, our data describes the ads clicked by the user and the times at which the clicks occurred.[9]

The sample of impressions on the AdCenter DVD was randomly generated from the search engine's complete log file. 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. We were informed that the average length of a session is about 10 minutes. Impressions that were part of longer user sessions have a proportionally higher probability to be in the data set than shorter ones. Since the vast majority of the sessions contain only one impression, we believe that sample selection issues are not of great importance.

Each data point was screened for privacy protection by Microsoft's technical team. On top of the privacy screening, each search string was "normalized." We do not have full information about the transformations employed, since this information is proprietary to Microsoft. However, we have been assured that normalization did not involve anything more complicated than converting the case of letters and getting rid of special characters, articles, and prepositions. We tried to minimize the impact of such transformations by the choice of search strings to analyze.

The subset of the dataset that we examine contains the impressions produced on 4 search strings (exact match): "games", "weather", "white pages" and "sex". These are the search strings that produced the most sponsored-ad clicks in the data, with the exception of searches for domain names and the "yellow pages" string. We did not consider searches for domain names like "google" or "myspace" because we believe that such searches commonly arise when a user either (i) mistakenly types a domain name into the search box, or (ii) types an incomplete domain name in the browser's address bar, forgetting an extension such as ".com," and is redirected by the browser to the search engine. The user's behavior in such situations may not be typical of his behavior following intentional searches. We also excluded the searches for "yellow pages" since we did not find enough variation in the impressions on this query to estimate our model.

We matched the impressions on the selected search strings to clicks on these impressions, applying a couple of sanity rules. We dropped impressions with the same unique impression id since we believe that those were due to errors in the data generation process. Similarly, when observing more than one click on the same link in an impression, we kept only the first click. Since the vast

[9] Advertising domains often experiment by varying the text of the advertising; we ignore this issue by ignoring the text and treating all ads with the same domain as identical. To the extent the text matters to consumers, it will be subsumed in our noise terms.

majority of repeat clicks occur within seconds of the first click (e.g., 84% occur within 10 seconds), we believe that the repeat clicks are either user errors or attempts to reload the web site following technical problems. If there are any repeat clicks that are not user errors or technical problems, they are effectively assumed not to affect the user's payoff (i.e., yield a zero marginal utility and have a zero marginal cost), which would justify dropping them. Our final dataset contains 92136 impressions, of which 17.7% have at least one click and 1.4% have at least two clicks.

2.2 Non-cascade clicks

Our dataset exhibits several features of user behavior that are not captured by the theoretical models in the existing literature. Namely,

- 46% of users who click do not click in the sequential order of positions, i.e. (1,2,...).
- 57% of users who click more than once do not "cascade," i.e., click on a higher position after clicking on a lower position.

These findings are inconsistent with the cascade model or with the AE model, both of which predict "cascades," and the latter also predicts sequential clicks. These findings demonstrate the importance of user heterogeneity. This is confirmed by having different orders of clicks by different users facing the same impression.

We model heterogeneity by letting users have different preferences over ads. Formally, we do this with a user$\times$ad random utility effect.

2.3 Rich Externalities

Another important observation from the data is the prevalence of externalities: the CTR on a given ad in a given position depends on which ads are shown in the other positions. These externalities immediately violate the EOS model or any other model in which users' decisions of clicking on different ads in an impression are independent of each other. Also, some of these externalities are inconsistent with the cascade model.

The externalities are evident by examining the conditional probabilities of clicking on a given ad in a given position under various assumptions about the ads displayed in the other positions. For example, Table 1 presents evidence for "externalities from above": the CTR on a given link displayed in position 2 conditional on the competitor displayed in position 1. (We were only able to conduct this analysis for the most popular ads, for which there were enough observations with desired impressions.) Comparison of the CTRs suggests negative externalities from the competitors. Namely, *Domain 1* in "white pages" string prefers a competitor at the position above that is of lower relevance (*Domain 3,* which does not have any white-page information) rather than higher relevance (any of the *Domains 2,3*). The same conclusion obtains in the "weather" search string: Having *Domain 3* (which does not have any weather information) as the above competitor is better than having *Domain 2*. All this evidence is suggestive of negative externalities, which may be attributed to users being satiated after clicking on good advertisements (in an extreme case of satiation, a user might not derive any benefit from a second ad – e.g., he may be fully satisfied with a single weather report).

More evidence of negative externalities is presented in Table 2. In this table, we repeat the exercise from the previous paragraph, but this time conditioning on the presence of certain competitors *below*. We can see that the differences are again statistically significant. Note that the average number of competitors is lower in the impressions that have *Domain 1*, which acts in the opposite direction to the negative externality, so if we had enough observations to control for the number of ads in the impression we expect the differences to be even larger. An important implication of externalities from below is rejection of cascading models (including the AE model) in which users always make clicking decisions going sequentially from top to bottom. Instead, users appear to exhibit more rationality, examining many ads before decided which ones to click on.

Another interesting observation that is inconsistent with the basic cascade model is that switching the ads in the top two positions affects the CTR of the ad in position 3. We were able to perform this analysis for one impression configuration on the "weather" search string. (The number of relevant observations in the other cases was fewer than 300, and in search strings "games" and "sex" there were no relevant observations at all.) The CTR of *Domain 1* in position 3 conditional on having *Domain 3* in position 1 and *Domain 2* in position two is 0.0434. When we switch the top two ads, the CTR drops to 0.0077. The difference is significant with 0.05 test size. To perform the test we used the asymptotic Wald test. The test statistic (distributed as standard normal) was 2.193. As we mentioned earlier we believe that *Domain 3* is not a very relevant domain for "weather", while *Domain 2* is very relevant. Thus, matching the better competitor domains with the higher position has a negative externality on a lower ad. This externality can again be attributed to user satiation: matching the better domain with the higher position increases the likelihood of user clicking on the better domain, making him more satiated and less likely to click on the third ad.

In addition to the externalities caused by satiation, we may also expect externalities caused by user learning about the quality of ads (as in Athey and Ellison (2007)). In contrast to satiation, we would expect learning to generate positive externalities: seeing one relevant ad would raise user expectation about the relevance of ads in general and make him more likely to click on other ads. Since the overall externalities exhibited in the data are negative, it appears that satiation is a more important source of externalities than learning. It would be difficult to identify these two effects separately given out data set: We cannot tell if a user stops clicking because he is satiated by the ads he has clicked on or because he is discouraged by their poor quality. One way to distinguish these two effects would be by using the data on "conversions" (i.e., purchases or follow-up requests) following the clicks. Another way would be to consider long-run learning about the general quality of ads across different search strings (where satiation is not an issue). Since we do not currently have data on conversions or on user histories, we cannot undertake either approach.

2.4 User Heterogeneity

Another interesting feature of the data is positive correlation between clicks on different positions in a given impression. We found this correlation by looking at impressions with a given (the most popular) ad is shown in position 1 and examining the correlation between clicking on this ad and clicking on any other ad in the impression. In a model without satiation in which a user's values for different ads are drawn independently (such as the EOS model), the correlation would be zero. In a world with satiation but with independent draws, the correlation would be negative. However, Table 3 demonstrates that the actual correlation is in some cases positive and statistically significant and in others statistically insignificant. For example, when *Domain 1* is displayed in position 1 on the "weather" search string and the user clicks on it, the probability of clicking on any other position is 5.1%, while if the user does not click on it, the probability of clicking on any other position is 3.4%, and this difference is highly significant. Similar significant positive correlation is found in the "white pages" search string, while in the other two search strings we find no significant correlation.

To explain these correlations, we model "vertical" heterogeneity of users, which makes some users more likely than others to click on any ad. For example, some users can have higher utilities for all ads (e.g., due to higher beliefs about the relevance of sponsored search advertising) or lower costs of clicking on ads (e.g., due to lower opportunity cost of time). We capture this vertical heterogeneity with a random user utility effect. The heterogeneity has to be large enough to offset the negative correlation among clicks created by satiation and in some cases even to generate positive correlation. This positive correlation is also needed to explain disproportionate numbers of multiple-clicks observations ("bundles"). Namely, our model without satiation and with independent clicks (which is then equivalent to the EOS model) would predict only 911 bundles of 2 clicked ads versus 1157 in the data, and only 20 bundles of 3 clicked ads versus 188 in the data. Introducing satiation only increases this discrepancy, and so we need to add vertical heterogeneity of users to fit the data better.

3 The Model

Consider a user i who faces an impression $a = (a_1, ..., a_N) \in A^N$, where N is the number of ads in the impression, A is the set of all possible ads that could be displayed, and $a_n \in A$ is the ad displayed in position n. Each ad $a \in A$ is characterized by a pair (p_a, v_{ai}), where $p_a \in [0, 1]$ is the probability that domain a will be prove to be relevant to him upon clicking, and v_{ai} is user i's value derived from clicking on the ad when it turns out to be relevant to him (when it is not relevant, the value is 0). The users learns whether an ad is relevant only upon clicking on it. The user also incurs a cost f_n from clicking on an ad in position n.

The timing of the user's decision problem is as follows:

(i) The user observes the impression $(a_1, ..., a_N)$ and the pairs (p_a, v_{ai}) for all ads in the impression.

(ii) The user either clicks on a position c in the impression that he hasn't clicked on yet or stops clicking (exits).

(iii) The user observes whether the clicked ad a_c is relevant to him or not.

(iv) Go to (ii).

We assume the user is a rational and forward-looking expected-utility maximizer and knows all the parameters. His decision problem can then be modeled as a dynamic programming problem whose payoff-relevant state can be summarized with two disjoint subsets $C_R, C_{NR} \subset \{1, ..., N\}$ of clicked positions that turned out to be relevant and irrelevant, respectively. The optimal continuation value of user i in state (C_R, C_{NR}), which we denote by $V_i(C_R, C_{NR})$, is governed by the following Bellman equation:

$$V_i\left(C_R, C_{NR}\right) = \max\left\{U_i(C_R), \max_{c\in\{1,..,N\}\setminus(C_R\cup C_{NR})}\left[p_{a_c,i}V_i\left(C_R\cup c, C_{NR}\right) + \left(1-p_{a_c,i}\right)\left(C_R, C_{NR}\cup c\right)\right]\right\}, \tag{3.1}$$

where $U_i(C_R, C_{NR})$ is the user's utility from stopping in state (C_R, C_{NR}). We postulate this utility to take the form

$$U(C) = \left(\sum_{n\in C_R} v_{a_n,i}^{1+R}\right)^{1/(1+R)} - \sum_{n\in C_R\cup C_{NR}} f_n, \tag{3.2}$$

where R is a parameter that captures the substitutability of different ads to the user.

We assume that the value of user i for a given ad a is generated as

$$v_{ai} = q_a + \varepsilon_{ai} + \delta_i,$$

where q_a is the fixed "quality" effect of ad a, ε_{ai} is a random shock to user value for a given ad, and δ_i is a random effect in user value for ads. We assume that ε_{ai} is drawn from an exponential distribution whose decay parameter is normalized to 1 (i.e., the c.d.f. is $F\left(\varepsilon_{ai}\right) = 1 - e^{-\varepsilon_{ai}}$). As for δ_i, it is drawn from a normal distribution whose standard deviation σ is a parameter to be estimated.

This model is rich enough to nest the following special cases:

- $R = 0$ (additively separable utility): The user's clicking decisions on different ads are then independent, and there are no externalities across ads. If in addition user random effects are absent (i.e., $\sigma = 0$), the clicks on the different positions are statistically independent, and the CTR on ad a in position n is $\Pr\left\{p_a q_a + \varepsilon_{ai} - f_n \geq 0\right\} = F\left(f_n - p_a q_a\right) = \max\left\{e^{p_a q_a}e^{-f_n}, 1\right\}$. Thus, provided that each ad receives a CTR less than one in any position (which is certainly true empirically), our model nests as a special case the EOS model, in which the CTR is the product of the ad fixed effect ($e^{-p_a q_a}$) and the position fixed effect (e^{-f_n}).[10] This nesting is the key motivation for us adopting the exponential distribution of errors ε_{ai}, and it also allows a simple quantitative interpretation of the estimated fixed effects on the CTR. In the EOS case, a consistent estimate of the fixed effects $p_a q_a$ and f_n can be obtained with an OLS regression of the logarithm of CTR on the ad and position dummies. Note that user uncertainty about

[10]If $\sigma > 0$ but small, the random variable $\varepsilon_{ai} + \delta_i$ can be approximated in the relevant upper tail with an exponential distribution, and the CTR can be approximated with the EOS multiplicatively separable form. Still, the model would be distinguishable from the EOS model by predicting a positive correlation between clicks on different positions.

relevance cannot be identified in this model - only the *expected* quality of ad a, $p_a q_a$ can be identified. Note also that since only the differences $p_a q_a - f_n$ are identified in the EOS model, the fixed effects f_n and $p_a q_a$ are identified only up to a constant.

- Perfect substitutability: $R = \infty$. In this case, the user's utility asymptotes to $U_i(C_R, C_{NR}) = \max_{n \in C_R} v_{a_n,i} - \sum\limits_{n \in C_R \cup C_{NR}} f_n$, i.e., the user derives utility from at most one ad (for example, he derives no benefit from viewing a second weather forecast.). This nests the classical consumer search model (e.g., Kim, Albuquerque, and Bronnenberg (2009)). In this model, user uncertainty about relevance matters: for example, if he has no uncertainty ($p_a = 1$ for all a) he will click on at most one ad; otherwise he may click on many ads. We can also approximate "cascade models" by assuming that position clicking costs f_n increase sharply in position n relative to any variation in ad quality, which induces users to click positions top to bottom.

We also allow for the case of $R < 0$, in which the clicks are complements rather than substitutes.

4 Estimation and Identification

We estimate the model using the Simulated Generalized Method of Moments based on Pakes and Pollard (1989).

Moments are computed using a nested dynamic programming approach. First we draw user specific effects ϵ, δ. We compute the user's optimal policy by solving system (3.1) by backward induction. The solution produces a stochastic process of ordered clicks for each user. We take 1000 draws from this stochastic process and compute the user-level average moments. We repeat the ϵ, δ draws 100 times and take the average of user-level moments. Each iteration of the estimation algorithm amounts to solving about one million dynamic programming problems.[11]

Our model has 46 unknown parameters and identifies them using 78 moments. Our parameters are divided into 3 groups:

- position fixed effects f_n, domain qualities q_a and relevance probabilities p_a,
- the standard variation σ of the user random effect,
- the satiation parameter R and the domain/position normalizing constant.

[11] Computations were possible because of supercomputer resources provided by Microsoft Corp.

We discuss the identification of all those groups separately.

We can identify the position fixed effects f_n and the domains' expected qualities $p_a q_a$ up to a constant even if $R = 0$. The moments that identify these parameters are the CTRs of domains, the CTRs of positions, and the positions/domain CTRs. Thus, we include the probabilities of clicking on each position from 1 to 5 conditional on each search string, the probabilities of clicking on each domain conditional on each search string, and the probabilities of clicking on each of the top 2 domains in each of the top 3 positions. (We dropped the moments that proved to have a close to zero variance.) In the data we observe the same domains placed in different positions, which allows us to identify the fixed effects: We can can identity position fixed effects f_n by comparing the CTR on the same given domain in different positions, and we can identify ad expected qualities $p_a q_a$ by comparing the CTRs of different ads in the same position. (When $R > 0$, we also have to control for the ad's competitors.) Under our assumption that user/position noise is distributed exponentially with decay parameter 1, the fixed effects can be interpreted as factors in the CTR.[12]

To identify the standard variation σ of the user random effect we include the unconditional probabilities of bundles of 2 and of 3 clicks. Increasing σ increases the correlation of clicks on different ads in the same impression, and so increases the probabilities of clicking bundles. (For parametric identification we use the functional form assumption that user specific errors have a normal distribution with mean 0.)

One the main contributions of this paper is identifying the user satiation parameter R and separating utility from cost. For this purpose, we use two additional sets of moments. The first set is composed of conditional probabilities similar to those presented in Table 1. For each search string, it consists of the following 3 moments:

- The probability of clicking on the most popular domain in position 2 conditional on the second most popular domain being in position 1;
- The probability of clicking on the most popular domain in position 2 conditional on the third most popular domain being in position 1;
- The probability of clicking on the second most popular domain in position 2 conditional on the most popular domain being in position 1.

[12] Since the number of clicks on positions 6 and 7 is very small we assume that cost of clicking on those are respectively 10% and 30% higher than on position 6, these numbers don't affect the estimation.

We dropped a couple of such moments that had zero observations in the sample. We did not include similar conditional probabilities for other impressions due to the small number of observations with such impressions.

The second set of moments identifying R consists of probabilities of continuing clicking after clicking on a given domain. We have 3 such moments per search string for the 3 most popular domains. The satiation parameter is identified from these moments, since more satiation means lower probabilities of continuing clicking. Given our assumed functional forms, the parameter R as well as the normalizing constant separating domain utilities and position costs are both identified. Identification is driven by the fact that moving a constant from costs to utilities and increasing R produce different curvature of incremental utility of subsequent clicks as a function of the already clicked links.

When $R \neq 0$, we can separately identify domain relevance probabilities p_a and their quality parameters q_a using the domain-specific continuation probabilities. Indeed, reducing p_a while raising q_a while holding the product fixed reduces increases the probability that the user continues clicking after clicking on domain a (while it does not affect the probability of clicking on domain a when R is close to zero). Intuitively, when we observe a domain with a high CTR but also a high probability of continuing clicking after clicking it, we attribute the "discrepancy" to high user uncertainty about the domain – i.e., low relevance probability p_a offset with a high quality q_a if relevant.[13]

We argued in the Section 2 that the cascade model and the AE models are not very realistic because of wide presence of non-ordered clicks. To make sure that our model explains this phenomenon we include, in addition to the already discussed moments, the probabilities of clicking on a link in a higher position conditional on clicking on a link in a lower position for each search string.

We perform moment weighting using a consistent estimate of the optimal weighting matrix, which in this case is the inverse of the asymptotic covariance matrix of the moment conditions. Estimation was done in 3 steps: (1) The moments conditions were evaluated at the starting point to get the initial weighting matrix, (2) A minimization routine (using initial weighting matrix) was performed and a consistent estimate of the optimal weighting matrix was computed, and (3) We obtained final estimates by minimizing the weighted sum of squared sample moment conditions.

To perform nonlinear optimization we used the Levenberg-Marquard gradient method[14] with a 10^{-9} tolerance factor. The starting point for the estimation was a consistent estimator of the

[13] An alternative explanation for the discrepancy is that users hold incorrect prior beliefs about the domain's quality. It would be difficult to distinguish this explanation from our model of user uncertainty.

[14] Uses software developed by the University of Chicago, as Operator of Argonne National Laboratory.

constrained model with $R = \sigma = 0$. In this special case, the model is separable, so we obtained consistent estimates of $p_a q_a$ and f_n by regressing the logarithm of the domain/position CTRs on the domain and position dummies. Because the cost and utility in the restricted model are identified only up to a constant, we normalized the cost of clicking on the top position to be 0. We drop this normalization when estimating the full model.

5 Results

The estimates of the main model are presented in Tables 4, 5 and 6. Table 4 presents the estimated position clicking costs for each search string. Table 5 presents the estimated quality measures of selected domains, organized by search string. Finally Table 6 contains the estimates of the satiation parameter R and the user heterogeneity parameter σ.

Table 4 presents our estimates of clicking costs on positions 1 to 5 in the four chosen search strings. (As mentioned earlier, we assume that positions 6 and 7 have respectively 10% and 30% higher clicking cost than position 5.) To interpret the magnitude of those numbers, recall that the utility of not clicking anything is normalized to 0. The fact that users face an exponential shock to their utility means that reducing the cost of a position by 1 increases the CTR of the position by a factor of e.

As expected, higher positions have lower cost of clicking. By exponentiating the cost differences we obtain the ratios of CTRs on different positions in the EOS world of $R = 0$. For example, in the "games" search string, the CTR of a given ad in position 1 is $\exp(3.4 - 1.4) \approx 7$ times higher in position 1 than in position 5. In the "weather" search string, the ration is $\exp(4.1 - 0.7) \approx 30$.

We do not know whether to attribute users' reluctance to click on lower positions to their bounded rationality that creates a high "psychic" cost of clicking on them, or to their rational expectations that ads placed in lower positions have lower quality, as e.g. in the AE model. We cannot answer this question given the available data. Answering this question would be important to predict long-run responses to changes in the ad allocation policy. Yet, we can be agnostic about this question in analyzing user behavior for a given allocation policy, or their short-run response to a change in the policy.

Our separation of utility from cost also enables us to compare the costs of clicking on ads under different keywords. For example, it turns out that people searching for "weather" find it relatively cheap to click of sponsored links, as opposed to those searching for "white pages". This cost heterogeneity of search strings may be due to the selection of different users in different searches and also to competition with the "organic search" results: If some keywords have better organic search results than others it would manifest itself in our model as a higher cost of clicking on sponsored search results. Unfortunately, we do not observe organic search links for the impressions we analyze, so we cannot test this hypothesis.

It is also interesting to note the heterogeneous cost differences between positions. This observation is important for optimizing bidding strategy in the keyword auctions. For example, the biggest percentage jump in cost between position 1 and 2 is observed for "weather". It suggests that there is a lot of extra value for winning slot number 1 vs. 2. At the same time for "sex" this difference is much smaller so an advertiser might benefit from bidding less and taking position 2.

Table 5 presents the estimates of qualities and relevance probabilities of selected domains for each keyword. In each keyword, we have selected the 4 most-clicked domains and pooled all the other domains, assuming they have the same quality. We can now supplement our reduced-form evidence negative externalities from Section 2 with structural estimates that provide us with quantitative guidance about the relative qualities of the domains. The advantage is that now we do not need to guess which domains are stronger and which are weaker competitors.

For example, in the "games" search string, the largest number of clicks is received by the Microsoft-owned *Domain 1*, yet the structural model yields that this domain has the lowest quality of the top 4. The structural model attributes the large number of clicks on this domain to its frequent placement in top positions (which presumably was done by Microsoft to promote the service). The same phenomenon is observed for Microsoft's *Domain 3* in the "weather" search string. Moreover, the description of this domain suggests that is service with maps, so users can be uncertain if it contains weather. This explains the domain's relatively low relevance probability.

We investigated the domains advertised on the "sex" string and found that only *Domain 1* is directly relevant to the search query. *Domain 2* is a general Internet shopping web site, *Domain 3* is a health nutrition store and *Domain 4* is a spam domain with no content other than sponsored links. Our estimates of domain qualities are consistent with these findings. However, it is interesting that *Domain 4* is estimated to have a relatively high quality. We cannot reveal the domain name due to Microsoft's privacy restrictions, but we can say that it is very well chosen, suggesting success in sexual life. We think that many users are disappointed by the domain after clicking on it, which explains the domain's lowest relevance probability among the "sex" domains.[15]

[15] In our model, users are induced to click on this domain because they rationally expect that its quality will be

Our estimates of satiation and of user heterogeneity are presented in Table 6. The interpretation of the standard deviation σ is that different users' probabilities of clicking on a given ad in a given position may differ on average by a factor of $\exp \sigma \simeq 9.5$.

To interpret the quantitative significance of the externality parameter R, we performed two counterfactual exercises. In the first exercise, we consider a hypothetical impression with only two advertisements and we compute the effect of satiation on the CTR of the advertiser in position 2. That is, we calculate the probability of the advertiser in slot 2 getting clicked when the user is not satiated by the ad in slot 1 (e.g., when a very low quality ad is placed in position 1), and compare it with the CTRs with satiation for different actual competitors placed in slot 1. Table 7 presents the results. The biggest losses due to satiation occur in the "sex" string, on ads that compete with *Domain 1* in position 1. For example, the CTR of Domain 3 in position 2 would be almost 3 times higher if it did not compete with *Domain 1* in position 1. On the other hand, *Domain 1* itself, being a high-quality ad, does not suffer much from externalities: its CTR in position 2 would have been only 3%-9% higher had it faced no competition from position 1.

The second counterfactual exercise is performed on the actual data. We simulate the CTRs of selected domains in the observed impressions in the hypothetical world without satiation (i.e., in which $R = 0$) and compare the results with the actual empirical CTRs. The simulation results are presented in Table 8. Unlike in the previous exercise, the size of the loss now depends not only on the domain's own quality but also on how often it faces strong competitors in the impression. A good example is given by comparing *Domain 1* in "games" to *Domain 4* in "white pages". Both of those domains have similar CTRs, however *Domain 4* gains much more in the counterfactual. While in general better domains tend to lose less due to externalities, the magnitude of the loss varies by search string. We also calculate that the total number of clicks in our dataset would have increased by 50% had satiation been absent.[16]

We can also quantify the effects of user uncertainty about relevance by considering the counter-

high if it proves to be relevant for them. Alternatively, the same data can be explained with a model (estimated in an earlier version of this paper) in which users are deceived by the domain's name into expecting the domain to be highly relevant and are systematically disappointed upon clicking on it. It would be difficult to distinguish between the models of rational user learning and incorrect priors from the available data. Note, however, that the policy implications of the two models would be quite different: high placement of "uncertain" domains may benefit rational users by facilitating their learning, while high placement of "deceptive" domains would hurt users.

[16]We cannot estimate the loss of advertiser profits caused by externalities, because of lack of click conversion data. This issue is left for further research.

factual in which this uncertainty is resolved before the user starts clicking. (For example, the search engine can reduce uncertainty by offering longer website descriptions, user comments, or experts opinions.) Although it is straightforward that eliminating user uncertainty will raise user welfare, a priori it is not clear how it would affects the total CTR. Table 8 presents the CTR effects on each domain of removing uncertainty about the relevance of ads. We note that the ads that benefit from this change are the ones with the greatest uncertainty (lowest relevance probability), while the ads with relevance probability close to 1 receive lower CTRs in the counterfactual regime. The last row of Table 10 presents the effects of removing uncertainty on user welfare and the total CTR: the counterfactual raises user welfare by nearly 15% and also yields a slight increase in the total CTR.

6 Counterfactual Matching Policies

This section presents the outcomes of simulations that compare user welfare and the total CTR for counterfactual matching policies of ads to positions. (Since we do not observe the advertisers' bids, we use the total CTR is our proxy for the search engine's revenue.) In particular, it is interesting to consider the matching policy that maximizes the users' expected utility and a potentially different policy that maximizes the total CTR.

A natural candidate matching policy is *Assortative Matching (AM)*, in which the ads are displayed in the decreasing order of their expected quality $p_a q_a$. (This policy is feasible for Microsoft provided that they know the qualities of the different ads. We suspect that Microsoft has some estimates of quality though they might not be perfect.) This policy in fact maximizes the total CTR and users' expected utility in the cases of our model without externalities:

Proposition 6.1. *If $R = 0$ and each ad receives a CTR less than one in each position, then assortative matching maximizes both the total CTR and the user's expected utility.*

Proof: It suffices to show that this is true conditional on any given realization of the user's random effect δ_i: this will imply that the same is true on expectation over δ_i.

Recall that the CTR on an ad of expected quality $\bar{q} = pq$ in a position with cost f is $\pi(\bar{q}, f) = \Pr\{\bar{q} + \varepsilon > f\} = \max\{e^{\bar{q}-f}, 1\} = e^{\bar{q}} \cdot e^{-f}$. Since this function is supermodular in $\left(e^{\bar{q}}, e^{-f}\right)$, a well-known result implies that assortative matching maximizes the total CTR.

As for the user's expected utility from having ad with expected quality q in position with cost

f, it can be computed as

$$\int_0^{\infty} \max\{\bar{q} + \varepsilon - f, 0\} e^{-\varepsilon} d\varepsilon = \int_{f-q}^{\infty} (\varepsilon - f + \bar{q}) e^{-\varepsilon} d\varepsilon = e^{\bar{q}-f} = \pi(q, f).$$

So in this benchmark model the user's expected utility coincides with the CTR and is again maximized by assortative matching.

When $R > 0$, the conclusion no longer holds, and we can find examples in which the total CTR or expected user utility are not maximized by AM. The intuition for how assortative matching can be improved upon for users is that for two ads with the same expected quality $p_a q_a$ it is strictly optimal to put the ad with the lower relevance probability p_a in a higher position so as to reduce the user's cost of learning its relevance. (Similar changes raise the CTR but to a smaller extent.)

Example 6.1. *Suppose that $R > 0$. There are two ads: $A = \{1, 2\}$, with ad 2 having no relevance uncertainty: $p_2 = 1$. The position clicking costs are $f_2 > f_1$. Compare the two possible impressions: $(1, 2)$ and $(2, 1)$. A user can have four possible types of optimal strategies: (a) always click on both ads, (b) always click on zero ads, (c) always click on one ad, and (d) click on the uncertain ad 1, and then click on ad 2 if and only if ad 1 proves to be irrelevant. (Of course, the optimal strategy may depend on the impression as well as the user's realized utility.) The expected payoffs from strategies (a) and (b) are the same on the two impressions. Since strategy (c) yields the same payoff as if $R = 0$, the expected payoff from this strategy is maximized by assortative matching, by the above proposition. However, the payoff from strategy (d) is maximized on impression $(1, 2)$, since with probability $1 - p_1$ it avoids clicking on ad 2. Thus, for parameter values for which strategy (b) is sufficiently likely to be optimal to the user on both impressions, it is optimal to display the uncertain ad above the certain ad, even if the certain as has higher expected quality ($q_2 > p_1 q_1$).*

We simulate both welfare- and CTR-optimal matching policies on our data, and compare them to both assortative matching and the actual data. As shown in Table 10 , we find that assortative matching would raise user welfare by 18% and the total CTR by 17%. However, the user-optimal placement policy would raise user welfare by 25% while the CTR-maximizing policy would raise the CTR by 18% (the two policies no longer coincide).

We also examine the improvements that could be achieved by "first-best" targeting, i.e., conditioning the impressions on the user's utility characteristics ϵ_i, δ_i. This approximates the situation in which the search engine uses information about the consumers like search history or demographics to tailor the impression. As shown in Table 10, moving towards first-best welfare-maximizing raises the user's expected utility by 59% from the actual data, and raises the total CTR by 31%. If we instead implement CTR-maximizing first-best targeting, we increase the CTR by 48%, without any loss to utility . We take it a an evidence that there are extra profit opportunities from exploring user level targeting that are also beneficial for the consumers. Microsoft does have access to substantial information about users' browsing habits stored in "cookies" on their computers, and this information is especially rich for users who have opened "Microsoft Passport " accounts (special accounts that offer a gateway to e-mail, Internet communicator, and many other services). To the best of our knowledge, Microsoft does not yet target sponsored search results to individual users. However, it is now pretty common to target display ads within webpages (in particular Yahoo! and Google-Doubleclick are known for doing this.) Our analysis of full-information targeting can be viewed as an upper bound on what can be achieved with targeted advertising.

7 Conclusion

This paper provides empirical evidence of externalities among ads in sponsored-search advertising, of user heterogeneity, and of user uncertainty regarding the relevance of ads to them. The evidence is provided using both reduced-form tests and a structural models of expected utility-maximizing users. The advantage of the structural model is that it allows us to make counterfactual predictions for different ad placement regimes and to quantify "user experience" as the average user's expected utility. We find that an alternative ad placement policy could raise user welfare by 25%, and that the increase could go up to 59% if information is available to target the placement to specific consumers. This suggests a large potential for ad targeting based on user level covariates, such as demographics or previous search history.

The interpretation of our counterfactuals depends on our attribution of on position fixed effects – namely, if these effects would be affected in the long run by implementation of an alternative ad placement policy. If the fixed effects are position-specific clicking costs that do not depend on the matching policy, then our counterfactuals are valid in the long run as well as in the short run. If instead these fixed effects are due to users' expectations that higher positions contain more relevant ads, these expectations will be altered in the long run under the new matching policy, which is not accounted for in our model. While we believe it is important to study long-run user learning of position-specific ad relevance, the dataset on the *Beyond Search* DVD does not allow is to do it as it does not track users over time. This is an important direction for future research.

References

Animesh, A., V. Ramachandran, and S. Viswanathan (2007): "An empirical investigation of the performance of online sponsored search markets," in *ICEC '07: Proceedings of the ninth international conference on Electronic commerce*, pp. 153–160, New York, NY, USA. ACM.

Athey, S., and G. Ellison (2007): "Position Auctions with Consumer Search," Discussion paper, UCLA Department of Economics.

Carterette, B., and P. N. Bennett (2008): "Evaluation measures for preference judgments," in *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 685–686, New York, NY, USA. ACM.

Craswell, N., O. Zoeter, M. Taylor, and B. Ramsey (2008): "An experimental comparison of click position-bias models," in *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pp. 87–94, New York, NY, USA. ACM.

Dupret, G. E., and B. Piwowarski (2008): "A user browsing model to predict search engine click data from past observations.," in *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 331–338, New York, NY, USA. ACM.

Edelman, B., and M. Ostrovsky (2007): "Strategic bidder behavior in sponsored search auctions," *Decis. Support Syst.*, 43(1), 192–198.

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.

Ghose, A., and S. Yang (2008): "An empirical analysis of sponsored search performance in search engine advertising," in *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pp. 241–250, New York, NY, USA. ACM.

Gomes, R., and K. Sweeney (2009): "Bayes-Nash Equilibria of the Generalized Second Price," *working paper.*

Hong, H., and M. Shum (2006): "Using Price Distributions to Estimate Search Costs," *RAND Journal of Economics*, 37(2), 257–275.

Hortacsu, A., and C. Syverson (2004): “Product Differentiation, Search Costs, And Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds,” *The Quarterly Journal of Economics*, 119(2), 403–456.

Kim, J., P. Albuquerque, and B. J. Bronnenberg (2009): “Online Demand Under Limited Consumer Search,” *SSRN eLibrary*.

Ostrovsky, M., and M. Schwarz (2009): “Optimal Auction Design: A Field Experiment,” *working paper*.

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

Papadimitriou, C. H., and S. Zhang (eds.) (2008): *Internet and Network Economics, 4th International Workshop, WINE 2008, Shanghai, China, December 17-20, 2008. Proceedings*vol. 5385 of *Lecture Notes in Computer Science*. Springer.

Radlinski, F., and S. Dumais (2006): “Improving personalized web search using result diversification,” in *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 691–692, New York, NY, USA. ACM.

Varian, H. R. (2006): “Position Auctions,” *International Journal of Industrial Organization*.

Yenmez, M. B. (2009): “Pricing in Position Auctions and Online Advertising,” *working paper*.

A Tables and Graphs

Competitor	CTR Domain 1
Domain 2	0.0763 (0.0060)
Domain 3	0.1842 (0.0138)
Domain 4	0.1078 (0.0240)

Competitor	CTR Domain 1
Domain 2	0.0189 (0.0020)
Domain 3	0.0535 (0.0038)
Domain 4	no observations

Table 1: Conditional Click-Through Rates on domains www.whitepages.com in search string ”white pages” and try.weatherstudio.com in search string “weather” when placed in position 2 given different domains in position 1. In parentheses we give standard errors. The estimates all have asymptotic normal distributions.

Domain	Regime	CTR	Number of observations	Avg. number of ads	Diff. of CTRs
Domain 2	With	0.1051	5061	6.1577	0.074^{***} (0.009)
	Without	0.1785	2112	7.1089	
Domain 3	With	0.1558	1560	7.1071	0.067^{***} (0.013)
	Without	0.223	2022	7.632	
Domain 4	With	0.1546	304	7.2993	0.019 (0.032)
	Without	0.1739	253	7.2885	

Table 2: CTRs of different domains in position 1 with and without having Domain 1 as competitor in any of the lower positions. One, two and three stars mean statistical significance with respectively 10%, 5% and 1% level.

	Search string/domain at top position			
	games Domain 1	weather Domain 1	white pages Domain 1	sex Domain 1
Clicking on top pos.	0.051	0.046	0.17	0.037
Not clicking top on pos. 0	0.034	0.043	0.116	0.045
Difference	0.017^{***} (0.005)	0.003 (0.006)	0.054^{***} (0.009)	0.008 (0.006)

Table 3: Probability of clicking on any other ad conditional on clicking and not clicking on top position

	Search string			
	games	weather	white pages	sex
Position 1	−1.409911 (0.147992)	−0.733881 (0.068356)	−1.906546 (0.048001)	−1.944481 (0.077962)
Position 2	−1.416051 (0.139611)	−1.275386 (0.057181)	−2.399289 (0.084595)	−2.115889 (0.080576)
Position 3	−1.795665 (0.094092)	−1.685258 (0.046972)	−2.985527 (0.052608)	−2.395046 (0.109008)
Position 4	−3.114202 (0.147293)	−3.511444 (0.138787)	−4.756276 (0.139401)	−3.134371 (0.308461)
Position 5	−3.370750 (0.257361)	−4.091068 (0.212180)	−5.606055 (0.129365)	−3.858798 (0.253015)

Table 4: Estimates of clicking cost in the baseline model

		Search string			
		games	weather	white pages	sex
Domain 1	Quality	−2.096701 (0.094661)	−3.128748 (0.069997)	−0.601720 (0.094814)	−0.404398 (0.094930)
	Prob.	0.965571 (0.082814)	0.955901 (0.046424)	0.978472 (0.042071)	0.975745 (0.096698)
Domain 2	Quality	−1.819920 (0.105851)	−4.019671 (0.076267)	−0.791732 (0.043651)	−1.945452 (0.092016)
	Prob.	0.764561 (0.075614)	0.965770 (0.073166)	0.831009 (0.026427)	0.939859 (0.031526)
Domain 3	Quality	−1.390381 (0.168104)	−3.729956 (0.121728)	−0.856630 (0.184161)	−2.465742 (0.068158)
	Prob.	0.826856 (0.072353)	0.716832 (0.036785)	0.783813 (0.043513)	0.900242 (0.040996)
Domain 4	Quality	−1.760887 (0.142918)	−3.012410 (0.074585)	−0.747621 (0.155905)	−1.429058 (0.103623)
	Prob.	0.714405 (0.047272)	0.948836 (0.026740)	0.933203 (0.044351)	0.893868 (0.065818)
Domain 5	Quality	−3.009994 (0.064489)	−4.840702 (0.084988)	−2.571644 (0.053933)	−2.842649 (0.088202)
	Prob.	0.963028 (0.044564)	0.970167 (0.042465)	0.971973 (0.042459)	0.905035 (0.048214)

Table 5: Estimates of domain quality and probabilities of relevance

R	σ
0.55 (0.02)	2.25 (0.067)

Table 6: Estimates of the structural parameters: the satiation parameter R and the standard deviation σ of the user random effect.

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.116	0.100	0.121	0.126	0.136
(d2)	0.093	-	0.089	0.107	0.114	0.124
(d3)	0.155	0.166	-	0.169	0.164	0.175
(d4)	0.085	0.101	0.084	-	0.107	0.118
(d5)	0.045	0.054	0.048	0.056	-	0.133

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.062	0.065	0.049	0.070	0.075
(d2)	0.022	-	0.030	0.021	0.033	0.037
(d3)	0.019	0.022	-	0.018	0.027	0.032
(d4)	0.057	0.067	0.070	-	0.075	0.079
(d5)	0.011	0.013	0.015	0.010	-	0.052

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.146	0.159	0.129	0.171	0.183
(d2)	0.057	-	0.084	0.068	0.111	0.124
(d3)	0.045	0.063	-	0.054	0.090	0.104
(d4)	0.081	0.111	0.118	-	0.147	0.160
(d5)	0.016	0.022	0.025	0.018	-	0.108

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.227	0.236	0.225	0.236	0.243
(d2)	0.033	-	0.076	0.059	0.078	0.086
(d3)	0.019	0.035	-	0.031	0.046	0.056
(d4)	0.050	0.091	0.099	-	0.106	0.115
(d5)	0.013	0.025	0.030	0.022	-	0.089

Table 7: Predicted CTR on a domain in position 2 conditional on different competitors in position 1 and conditional on a “dummy competitor” inposition 1 who creates no satiation.

		Search string			
		games	weather	white pages	sex
Domain 1	True data	0.094	0.073	0.129	0.256
	R=0	0.126	0.082	0.158	0.271
Domain 2	True data	0.042	0.022	0.086	0.072
	R=0	0.068	0.032	0.123	0.098
Domain 3	True data	0.119	0.024	0.063	0.030
	R=0	0.152	0.035	0.095	0.049
Domain 4	True data	0.047	0.012	0.099	0.115
	R=0	0.074	0.019	0.139	0.131
Domain 5	True data	0.063	0.012	0.031	0.031
	R=0	0.105	0.019	0.062	0.056

Table 8: Counterfactual domain CTRs if there are no externalities, i.e $R = 0$

		Search string			
		games	weather	white pages	sex
Domain 1	True data	0.094	0.073	0.129	0.256
	No uncertainty	0.087	0.072	0.121	0.254
Domain 2	True data	0.042	0.022	0.086	0.072
	No uncertainty	0.052	0.022	0.100	0.073
Domain 3	True data	0.119	0.024	0.063	0.030
	No uncertainty	0.121	0.026	0.078	0.032
Domain 4	True data	0.047	0.012	0.099	0.115
	No uncertainty	0.059	0.013	0.102	0.120
Domain 5	True data	0.063	0.012	0.031	0.031
	No uncertainty	0.062	0.012	0.029	0.034

Table 9: Counterfactual domain CTRs if all uncertainty about ad quality of resolved prior to clicking decisions

	Utility	CTR
True data	0.231	0.181
Assortative matching	0.273	0.210
Max U	0.289	0.209
Max CTR	0.276	0.214
First best utility targeting	0.368	0.237
First best CTR targeting	0.232	0.268
No uncertainty	0.264	0.187

Table 10: Repositioning and no uncertainty counterfactual utilities and CTRs