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

Przemyslaw Jeziorski[†] Ilya Segal[‡]

April 26, 2012

Abstract

We study users' responses to sponsored-search advertising using consumer-level data from Microsoft's Live AdCenter distributed in the "Beyond Search" initiative. We estimate a dynamic 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 quality of advertising. We estimate the substitutability of ads in users' utility function, the fixed effects of different ads and positions, as well position-specific priors about quality. To match the clicking patterns in the individual-level data, we allow rich user-level unobserved, persistent heterogeneity. We find substantial substitutability of ads, which generates large negative externalities: 81% more clicks would occur in a hypothetical world in which each ad faces no competition. As for counterfactual ad placements, our simulations indicate that clickthrough-optimal matching increases clickthrough rate (CTR) by 10.8%, and user-optimal matching increases user welfare by 16.7%. Moreover, targeting ad placement to specific users could raise user welfare by 63.7%. Finally, user welfare could be raised by 4.47% if users 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.

[†]Haas School of Business, UC Berkeley, Berkeley CA 94720

[‡]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% 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," that is, "sponsored links" that accompany results produced in response to the consumers' search queries. Search advertising accounts for 41% of total Internet advertising revenues. It is viewed as the most effective kind of advertising because of its precise targeting: a consumer's search string reveals a great deal about the products, he is likely to be interested in. This precise targeting allows a display of only the most relevant ads, which in turn induces consumers to click on them. Although search advertising has recently received a lot of attention, researchers know little about consumer behavior in this market. This paper makes a step towards remedying this problem.

Existing papers on search advertising postulate 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), Edelman and Schwarz (2010)). 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 user behavior, 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 and

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

its fit with the data is established, we can use the model 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 judgments 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 contain a random sample 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, and the others are displayed in a column to the right of the organic search results.) For each advertising impression, our data describe the ads the user clicked 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 by Microsoft 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.[34]"In fact, it is easy to convince oneself of

[2]Dupret and Piwowarski (2008) quantify ad quality by calibrating a heuristic model of user behavior on real data. However, because 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 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 the engine may simply put the most relevant ad at the top, and no position effect is present 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.

[4]Additionally, we test for stationarity by examining how the predictive power of the model changes over time. We

the large 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 prevalence of externalities across ads contradicts the EOS model: 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* at position 2 on the "white pages" search string is 18% if its competitor at position 1 is *Domain 3* (which is not a good match for "white pages"because it offers yellow pages), but drops to 8% if the competitor is *Domain 2* (which is a specialized advertising company).[5] 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," that is, 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 at position 3, given the two ads shown in position 1 and 2, still depends on the order in which these two latter ads are presented.

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 quality of ads. 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 (2010), 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 parametrizes the degree of substitutability (satiation) among ads with a parameter R in a "Constant Elasticity of Substitution" utility function. Additionally, we allow R to vary across users. For $R_i = 0$, user utility is the sum of the utilities derived from the clicked ads, and in this case, no externalities are present across ads, as in the EOS model. At the other extreme, when $R_i = \infty$, 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

find the average predicted probability of choosing the observed bundle of ads does not depend on the time stamp.

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

the externalities are the most prominent (similar to Kim, Albuquerque, and Bronnenberg (2010)).

In addition to the substitutability R, we endogenize the drop in the CTR associated with lower positions. We allow it be a consequence of two factors: scrolling cost and users' expectations about the quality of ads at different positions. We capture the former effect by position fixed effects and the latter by a Bayesian signaling model. In this model, users have priors about the quality of ads on each position and receive signals about the quality of a particular ad from its description. Moreover, we allow for priors to be different across users to control for the fact that different people might have had different ad experiences in the past. We argue identification, and jointly estimate these two mechanisms under the assumption that long-run user learning about positions is unbiased.

For each of the more common ads, we estimate their quality fixed effects and 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. In addition, we allow for classical preference shocks.

We find that externalities are both statistically and economically significant. Our estimate of the mean substitutability parameter R is 0.55, with substantial heterogeneity across users. Using the estimate, we predict the CTR on most domains in the hypothetical world without externalities would have been substantially higher than their actual CTR. We predict the total number of clicks in our dataset would have been 81% higher had satiation been absent. Moreover, we find evidence of user uncertainty: if this uncertainty were resolved prior to clicking, consumer welfare would be 4% higher with 6% higher overall CTR.[6]

We use our estimated model to predict user behavior on counterfactual ad impressions, and generate 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 not true in our model even without externalities ($R_i = 0$), because of learning from positions. To investigate this issue, we simulate short, and long-run counterfactuals of welfare-optimal and CTR-optimal matching policies and compare them to the data and assortative matching according to simple OLS-type estimates. When computing the long-run effects, we leverage on the fact that we explicitly estimate the expectations about quality of ads on each position separately from (psychic) clicking costs. Consequently, we presume that in

[6]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 complementary probability.

the long-run, users exactly learn the new search engine placement policy collapsing the priors to degenerate distributions. We find that in the short run, the assortative matching provides about 10.6% improvement in welfare and 6.76% improvement in the CTR. In comparison, welfare-optimal placement provides 16.7% welfare improvement and CTR-optimal matching provides 10.8% CTR improvement (welfare and CTR-optimal matchings are in practice very similar to each other but different than assortative matching). These results are robust to allowing for long-run adjustments.

Because the search engine provider might consider improving description signals (providing more space, introducing user or expert reviews), we measured the welfare effects of removing uncertainty. We find that if the users know the quality of an ad exactly ex-ante clicking, their welfare goes up by 4.5% and CTR goes up by 6%. We view these gains as fairly modest compared the ones achieved by different matching policies.

Researchers have also suggested that targeting the impressions to individual users, for example, based on their browsing history or demographics could improve user experience and CTR (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 such targeting could achieve, 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%, whereas CTR-optimal first-best targeting would raise the total CTR by 53%. Moreover, we find that these gains become more significant in the long run. In particular, welfare can be raised by 63% and CTR can be raised by 68%.

Athey and Ellison (2011) (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, keeping position priors fixed within the impression. More explicitly, we assume that realized ad quality ex-post clicking does not affect priors of other positions in the short-run for two reasons: (1) such updating would generate positive "informational externalities" across ads; that is, an ad would benefit from having better ads in the same impression. Empirically we find that the overall externalities are instead negative, and separate identification of both satiation externalities

[7] In the authors' basic model, the ads' texts are uninformative, 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). Like the cascade model, the AE model is inconsistent with non-sequential and non-cascading clicks and with externalities from below.

and informational externalities from the available data would be difficult. (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 (2011)). Although long-run learning over the course of many sessions may prove to be 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. Section 2 describes the dataset and examines 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. Microsoft distributed this DVD was distributed to a few dozen recipients of the company'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, and 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

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

`Local Weather Forecast`

`Get Live Weather Forecasts & More With The Free Weather Toolbar`

`Weather.alot.com`

For each advertising impression, our data describes the ads the user clicked 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. The average length of a session is about 10 minutes. Impressions that were part of longer user sessions have a proportionally higher probability of being in the data-set than shorter ones. Because the vast majority of the sessions contain only one impression, we believe sample selection issues are not of importance.

Microsofts technical team screened each data point for privacy protection, and each search string was "normalized." We do not have full information about the transformations employed, because this information is proprietary to Microsoft. However, the company assured us that the 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 examine contains the impressions produced on four search strings (exact match) – "games," "weather," "white pages," and "sex"– 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 such as "google" or "myspace" because we believe 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" because 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, apply-

[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.

ing a couple of sanity rules. We dropped impressions with the same unique impression id because we believe they were due to errors in the data-generation process. Similarly, when we observe more than one click on the same link in an impression, we kept only the first click. Because the vast majority of repeat clicks occur within seconds of the first click (e.g., 84% occur within 10 seconds), we believe the repeat clicks are either user errors or attempts to reload the web site following technical problems. If any repeat clicks are not user errors or technical problems, we effectively assume they do not 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 $92,136$ 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 a couple of features of user behavior the theoretical models in the existing literature do not capture, 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, 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 introduce a user$\times$ad random utility effect, which captures differences in users' tastes.

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 to click 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.

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 CTRs on domain 1 in search string "white pages" and domain 1 in search string "weather" when placed at position 2 given different domains at position 1. We give standard errors in parentheses. The estimates have asymptotic normal distributions.

For example, Table 1 presents evidence for "externalities from above": the CTR on a given link displayed at position 2 conditional on the competitor displayed at 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 the "white pages" string prefers a competitor at the position above to be of lower relevance (*Domain 3,* which does not have any white-page information) rather than higher relevance (any of the *Domains 2,4*). We obtain the same conclusion 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; for example, he may be fully satisfied with a single weather report).

Table 2 presents more evidence of negative externalities. 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 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 (than the negative externality), so if we had enough observations to control for the number of ads in the impression, we expect the differences would be even larger. An important implication of externalities from below is the 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 deciding on which ones to click.

Another interesting observation that is inconsistent with the basic cascade model is that switch-

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 at position 1 with and without having Domain 1 as competitor in any of the lower positions. One, two and three stars mean statistical significance at the 10%, 5%, and 1% levels, respectively.

ing the ads in the top two positions affects the CTR of the ad at 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 the search strings "games" and "sex" contained no relevant observations at all.) The CTR of *Domain 1* in position 3 conditional on having *Domain 3* at position 1 and *Domain 2* at 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, and the test statistic (distributed as standard normal) is 2.193. As we mentioned earlier, we believe *Domain 3* is not a relevant domain for "weather,"whereas *Domain 2* is. 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 the 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 expect externalities caused by user learning about the quality of ads (as in Athey and Ellison (2011)). In contrast to satiation, we would expect learning to generate positive externalities: seeing one relevant ad would raise the user's expectation about the relevance of ads in general and make him more likely to click on other ads. Because the overall externalities exhibited in the data are negative, satiation appears to be more important source of externalities than learning. Identifying these two effects separately given our data-set would be difficult: 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

	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

between 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). Because 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 shown at position 1, and examining the correlation between clicking on this ad and 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 at position 1 on the "weather" search string and the user clicks on it, the probability of clicking on any other position is 5.1%, whereas 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. We find similar significant positive correlation in the "white pages" search string, but no significant correlation in the other two search strings.

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, so we need to add vertical heterogeneity of users to better fit the data.

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 at position n. The ad $a \in A$ is characterized by a quality measure v_{ai}, which is user i's value derived from clicking on the ad. The quality is unknown to the user prior to clicking and is fully revealed right after clicking.

The user learns about the quality of an ad from its position and description. Before reading the descriptions, the user forms priors about the quality of ads at each position. The priors are position-keyword specific and are a result of the user's prior experiences with search advertising. By making the priors both keyword and position specific, we allow users to be aware of heterogeneity across populations of ads placed at each position within each keyword. To save on notation, we denote both positions and keywords by n. Formally, we assume prior belief about quality of ads at the position n has a normal distribution with the mean $\bar{q}_n$ and the variance $\bar{\sigma}_{ni}$. We assume the priors are generated by an unbiased cross-impression learning process, but we are agnostic about the details of that process. The variance of the prior is allowed to vary across users to capture the fact that some users are more and others are less experienced; therefore, their priors should have thinner or fatter tails. The ad description provides an additional unbiased signal about the true ad quality v_{ai}. Formally, the description is a random variable $x_{ai} = v_{ai} + \nu_{ai}$, where ν_{ai} is a mean zero normally distributed noise, with variance σ. [10]

Additionally, each user incurs a cost f_n from clicking on an ad at position n. This cost represents both the behavioral cost of scrolling down as well as the opportunity cost created by organic search

[10]Theoretically, our model facilitates identification of the ad-specific variance σ_a. In practice, we failed to reject that these variances are the same with a 1%-size test.

results.[11] Because the organic results vary between keywords, we allow this cost to be keyword specific. The important distinction between f_n and position priors is that the position costs are exogenous and not subject to learning in the long-run. In contrast, one could reasonably expect position priors to change in the long-run to respond to changes in ad-placement policy.

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

(i) The user searches for a particular keyword and forms priors about the quality of ads at each position $((\bar{q}_1, \bar{\sigma}_{ni}), ..., (\bar{q}_N, \bar{\sigma}_{Ni}))$.

(ii) The user observes the impression $(a_1, ..., a_N)$ and the quality signals $(x_{1i}, \ldots, x_{Ni})$ for all ads in the impression. The user forms posterior beliefs about the qualities of each ad.

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

(iv) The user observes the true quality v_{ai} of a clicked ad a_c.

(v) Go to (iii).

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 a set $C \subset \{1, ..., N\}$ of clicked positions and a set $Q = \{q_{ai} : a \in C\}$ of observed ex-post qualities. The optimal continuation value of user i in state (C, Q), which we denote by $V_i(C, Q)$, is governed by the following Bellman equation:

$$V_i(C, Q) = \max \left\{ U_i(C, Q), \max_{c \in \{1,..,N\} \setminus C} EV_i(C \cup c, Q \cup q_{ci}) \right\} \tag{3.1}$$

The expectation is taken with respect to the posterior of q_{ci}, which is a random variable distributed as normal with the mean

$$\left(\frac{x_{ai}}{\sigma^2} + \frac{\bar{q}_n}{\bar{\sigma}_{ni}^2} \right) / \left(\frac{1}{\sigma^2} + \frac{1}{\bar{\sigma}_{ni}^2} \right)$$

and the variance

$$\left(\frac{1}{\sigma^2} + \frac{1}{\bar{\sigma}_{ni}^2} \right)^{-1},$$

where n is a position of the ad a. $U_i(C, Q)$ is the user's utility from stopping in state (C, Q). We postulate this utility to take the form

[11] For further analysis of higher-order interactions between organic and sponsored search links, we refer the reader to Yang and Ghose (2010).

$$U_i(C,Q) = \left(\sum_{n \in C} q_{a_n,i}^{1+R_i} \right)^{1/(1+R_i)} - \sum_{n \in C} f_n, \tag{3.2}$$

where R_i is a parameter that captures the substitutability of different ads to the user. It is generated as $R_i = R + \sigma^R \eta_i$, where η_i is standard normal.

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

$$v_{ai} = \bar{v}_a + \varepsilon_{ai} + \delta_i,$$

where $\bar{v}_a$ 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 $F(\varepsilon_{ai}) = 1 - e^{-\varepsilon_{ai}}$). As for δ_i, it is drawn from a normal distribution whose standard deviation σ^δ is a parameter to be estimated.

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

- $R_i = 0$ (additively separable utility), $\sigma = 0$ (no uncertainty): The user's clicking decisions on different ads are then independent, and no externalities are present across ads. If in addition user random effects are absent (i.e., $\sigma^\delta = 0$), the clicks on the different positions are statistically independent, and the CTR on ad a at position n is $\Pr\{\bar{v}_a + \varepsilon_{ai} - f_n \geq 0\} = F(f_n - \bar{v}_a) = \max\{e^{\bar{v}_a} e^{-f_n}, 1\}$. Thus, provided that each ad receives a CTR less than one in any position (which is certainly true empirically), our model nests the EOS model as a special case, in which the CTR is the product of the ad fixed effect ($e^{-\bar{v}_a}$) and the position fixed effect (e^{-f_n}).[12] This nesting is the key motivation for our adoption 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 $\bar{v}_a$ 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 relevance cannot be identified in this model – only the quality of ad a, $\bar{v}_a$ can be identified. Note also that because only the differences $\bar{v}_a - f_n$ are identified in the EOS model, the fixed effects f_n and $\bar{v}_a$ are identified only up to a constant.

[12] If $\sigma^\delta > 0$ 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.

- Perfect substitutability: $R_i = \infty$. In this case, the user's utility asymptotes to $U_i(C,Q) = \max_{n \in C} q_{a_n,i} - \sum_{n \in C} f_n$; that is, the user derives utility from at most one ad (e.g., he derives no benefit from viewing a second weather forecast.). This nests the classical consumer search model (e.g., Kim, Albuquerque, and Bronnenberg (2010)). In this model, user uncertainty about relevance matters: for example, if he has no uncertainty ($\sigma = 0$), he will click on at most one ad; otherwise, he may click on many ads. We can also approximate "cascade models" by assuming position clicking costs f_n increase sharply at 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). Because the model allows for a rich persistent unobserved heterogeneity (including continuous types), the moments are computed using a nested dynamic programming approach. First we draw user specific effects (ϵ, δ), satiation parameter R_i, and user-specific priors $((\bar{q}_1, \bar{\sigma}_{ni}), ..., (\bar{q}_N, \bar{\sigma}_{Ni}))$ as well as ad-quality signals coming from the descriptions $(x_{1i}, ..., x_{1N})$. Using these primitives, we compute the user's optimal policy by solving system (3.1) by backward induction. The solution produces an optimal policy as a function of a set of clicked ads as well as observed qualities after clicking. We compute the user-level moments by running this policy forward and assuming users observe the true quality ex-post. We repeat the draws 100 times and take the average of user-level moments. Each iteration of the estimation algorithm amounts to solving about 10 million dynamic programming problems. [13] The integration of the right-hand-side of the Bellman equation is performed using Gauss-Hermite quadrature with five nodes. Therefore, the quality of each clicked ad can be discretized into six endogenous grid points that depend on the draw of random effects and ad-description signals: five nodes of the quadrature centered at the posterior and a node at the true quality needed for forward simulation. We find it is enough to consider bundles of three choices, because bundles of four are extremely rare in the data, and including them shifts the moments by the negligible amount. In this case each dynamic program has 74,640 state points.

We do not explicitly model cross-visit user learning. Instead, we infer users' position priors using

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

the assumption that the cross-impression learning generates unbiased priors and that ad placement is stationary. Under the first assumption, $\bar{q}_n = E[q_{ai}|n]$. Under the second assumption, we can estimate the mean of the position prior by

$$\hat{q}_{ni} = \sum_a \phi_{na} \bar{v}_a + 1,$$

where ϕ_{na} is the observed frequency of placing an ad a on the position n. The prior variances are modeled as $\bar{\sigma}_{ni} = \bar{\sigma}_n + \epsilon_i^P$, and ϵ_i^P is a normally distributed random effect with a standard deviation σ^P. This specification allows for some users to have stronger position effects than others. In the estimation, we allow $\bar{\sigma}_n$ to be different across keywords, but not across positions within the keyword.[14]

Our model has 49 unknown parameters and identifies them using 78 moments. Our parameters are divided into four groups:

(i) domain mean qualities $\bar{v}_a$,

(ii) position fixed effects f_n, precisions of the priors $\bar{\sigma}_n$, and a heterogeneity of priors precisions σ^P

(iii) the standard deviation σ^δ of the user random effect,

(iv) the satiation parameters (R, σ^R), the domain/position normalizing constant, and the precision of the description signals σ.

We discuss the identification of all groups separately.

We can identify between effects of parameters in the group (ii) and the domains' expected qualities $\bar{v}_a$ up to a constant even if $R_i = 0$. The moments that identify these parameters are the CTRs of domains and the CTRs of positions. Thus we include the probabilities of clicking on each position from 1 to 5 conditional on each search string and the probabilities of clicking on each domain conditional on each search string (we dropped the moments proven 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 identify effects of positions (joint effects of priors and position costs) on clicking by comparing the CTR of the same domain at different positions. Similarly, we

[14]The model would allow us to identify position-specific variance using perfect data. It would require seeing many click events on high-quality ads placed on low positions. These events are rare in our data, both because we do not see enough impressions with high-quality ads at the bottom, and because bottom ads are rarely clicked.

can identify ad qualities q_a by comparing the CTRs of different ads in the same position. (When $R_i > 0$, 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 position effects can be interpreted as factors in the CTR.[15]

Position fixed effects f_n can be separately identified from the effects of position-specific priors by looking at the interaction between positions and the advertisements. If the heterogeneity in CTRs placed in different positions comes from position priors, good ads should benefit from top positions less than bad ads. Conversely, good ads should lose from bottom positions more than bad ads. On the other hand, if the CRT heterogeneity comes from position fixed effects, we should not see the above interactions. To capture these effects, we include domain-position-specific CTRs for the top two advertisers placed at each of the top three positions.

To identify the standard variation σ^δ of the user random effect, we include the unconditional probabilities of bundles of two and of three clicks. Increasing σ^δ 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.) These moments also help to pin down the variance of R_i, because the higher variance generates more bundle clicks.

One of the main contributions of this paper is identifying the users' 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, the set consists of the following three moments:

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

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

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

with such impressions.

The second set of moments identifying R consists of probabilities of continuing clicking after clicking on a given domain. We have three such moments per search string for the three most popular domains. The satiation parameter is identified from these moments, because 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 the precision of the description σ and ad-quality parameters $\bar{v}_a$ using the domain-specific continuation probabilities, because σ directly determines the variance of the posterior. Indeed, increasing σ and raising $\bar{v}_a$ while holding the certainty equivalent of the posterior fixed increases the probability that the user continues clicking after clicking on domain a (whereas it does not affect the probability of clicking on domain a when R is close to zero), because the click is coming more from the position prior, which can cause the user to be disappointed ex-post. Intuitively, when we observe a domain with a high CTR of an ad but also a high probability of continuing clicking after clicking it (it happens usually on high positions), we attribute the "discrepancy" to high user uncertainty about the domain; that is low σ offsets with a high quality $\bar{v}_a$.[16]

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 ensure 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 three steps: (1) we evaluated the moment conditions at the starting point to get the initial weighting matrix, (2) we performed the minimization routine (using initial weighting matrix) and we computed a consistent estimate of the optimal weighting matrix, and (3) we obtained final estimates by minimizing the weighted sum of squared sample moment conditions.

To perform nonlinear optimization, we used the combination of Nelder-Mead and Levenberg-

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

	Search string			
	games	weather	white pages	sex
Position 1	−0.96 (0.011)	−0.48 (0.002)	−1.57 (0.002)	−1.62 (0.004)
Position 2	−1.31 (0.004)	−1.03 (0.002)	−2.14 (0.003)	−1.69 (0.004)
Position 3	−1.73 (0.003)	−1.58 (0.004)	−2.41 (0.005)	−2.05 (0.005)
Position 4	−3.34 (0.011)	−3.31 (0.009)	−4.68 (0.009)	−2.73 (0.030)
Position 5	−3.26 (0.019)	−3.74 (0.011)	−4.93 (0.015)	−3.44 (0.046)

Table 4: Estimates of clicking cost in the baseline model

Marquard gradient method[17] with a 10^{-9} tolerance factor. The starting point for the estimation was a consistent estimator of the constrained model with $R = \sigma = 0$. In this special case, the model is separable, so we obtained consistent estimates of $\bar{v}_a$ 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

Tables 4, 5, 6, and 7 present the estimates of the model. 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. Table 6 contains the estimates of the satiation parameter R and the user heterogeneity parameter σ^δ. Finally, Table 6 presents the estimates of the Bayesian signaling model.

Table 4 presents our estimates of clicking costs on positions 1 to 5 in the four chosen search strings. (As mentioned earlier, we assume positions 6 and 7 have a 10% and 30% higher clicking cost, respectively, 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.

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

	Search string			
	games	weather	white pages	sex
Domain 1	−2.64 (0.012)	−3.05 (0.002)	−0.31 (0.002)	−0.56 (0.006)
Domain 2	−1.92 (0.006)	−3.57 (0.002)	−1.08 (0.003)	−1.94 (0.007)
Domain 3	−1.35 (0.008)	−4.21 (0.002)	−1.54 (0.004)	−2.74 (0.004)
Domain 4	−2.29 (0.009)	−2.80 (0.004)	−0.74 (0.004)	−1.47 (0.009)
Domain 5	−2.98 (0.006)	−4.35 (0.005)	−2.14 (0.005)	−3.04 (0.007)

Table 5: Estimates of domain quality and probabilities of relevance

Satiation parameter		Preference shock
Mean (R)	Std. dev. (σ^R)	Std. dev. (σ^δ)
0.55 (0.001)	0.18 (0.003)	2.23 (0.002)

Table 6: Estimates of the user specific random effects: satiation parameters and preference shocks

Std. dev. of the prior				User random effect	Description noise
games	weather	white pages	sex	std. dev. (σ^P)	std. dev. (σ)
0.96 (0.005)	0.93 (0.001)	0.91 (0.004)	0.99 (0.008)	0.14 (0.002)	0.39 (0.001)

Table 7: Estimates of the learning process: standard deviations of the position priors for each keyword, user level random effect, and a standard deviation of the noise in a description signal

As expected, higher positions have a lower cost of clicking. By exponentiating the cost differences, we obtain the ratios of CTRs on different positions in the EOS world of $R_i = 0$. For example, in the "games" search string, the CTR of a given ad at position 1 is $\exp(3.26 - 0.96) \approx 10$ times higher at position 1 than position 5. In the "weather" search string, the ratio is $\exp(3.74 - 0.48) \approx 26$.

Because we estimate the learning model separately from the position fixed effects, we know the above reluctance to click on lower positions is due to users' bounded rationality that creates a high "psychic" cost of clicking on them, as opposed to expectations about the quality of positions that is not coming from sponsored search. This fact is important when predicting long-run responses to changes in the ad-allocation policy, because users would not update their cost in response.

Our separation of utility from cost also enables us to compare the costs of clicking on ads under different keywords. For example, people searching for "weather" find clicking on sponsored links to be relatively cheap, 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, the difference 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.

Also note the heterogeneous cost differences between positions. This observation is important for optimizing bidding strategy in the keyword auctions. For example, "weather" exhibits the biggest percentage jump in cost between position 1 and 2. This jump suggests that winning slot number 1 versus 2 carries extra value. At the same time, the percentage jump for "sex" is much smaller, so an advertiser might benefit from bidding less and taking position 2.

Table 5 presents the estimates of mean ex-post qualities $\bar{v}_a$ of selected domains for each keyword. In each keyword, we have selected the four 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 being 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 Microsoft-owned *Domain 1* receives the largest number of clicks, 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 to the domain's frequent placement

in top positions (which presumably was done by Microsoft to promote the service). We observe the same phenomenon for Microsoft's *Domain 3* in the "weather" search string, which might be due to domain's description as a service with maps; therefore, users might correctly think the domain does not contain weather.

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, interestingly *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 it is very well chosen, suggesting success in sexual life.

Table 6 presents our estimates of satiation and of user heterogeneity are presented in . The interpretation of the standard deviation σ^δ is that different users' probabilities of clicking on a given ad in a given position may differ on average by a factor of $\exp \sigma^\delta \simeq 9$.

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.044	0.042	0.045	0.061	0.112
(d2)	0.104	-	0.070	0.093	0.109	0.152
(d3)	0.127	0.104	-	0.112	0.128	0.185
(d4)	0.089	0.064	0.055	-	0.090	0.134
(d5)	0.050	0.042	0.040	0.044	-	0.115

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.039	0.044	0.032	0.045	0.074
(d2)	0.028	-	0.036	0.027	0.037	0.054
(d3)	0.019	0.022	-	0.019	0.025	0.042
(d4)	0.044	0.052	0.060	-	0.061	0.084
(d5)	0.019	0.020	0.023	0.019	-	0.042

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.127	0.147	0.113	0.162	0.282
(d2)	0.053	-	0.089	0.060	0.105	0.223
(d3)	0.039	0.042	-	0.040	0.070	0.175
(d4)	0.063	0.087	0.107	-	0.123	0.243
(d5)	0.030	0.038	0.040	0.036	-	0.155

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.168	0.179	0.153	0.179	0.258
(d2)	0.043	-	0.082	0.053	0.090	0.138
(d3)	0.026	0.037	-	0.034	0.049	0.097
(d4)	0.056	0.092	0.110	-	0.113	0.172
(d5)	0.021	0.029	0.036	0.027	-	0.091

Table 8: Predicted CTR on a domain in position 2 conditional on different competitors in position 1 and conditional on a “dummy competitor” in position 1 that creates no satiation.

		Search string			
		games	weather	white pages	sex
Domain 1	True data	0.104	0.090	0.171	0.268
	R=0	0.124	0.102	0.284	0.325
Domain 2	True data	0.061	0.040	0.112	0.107
	R=0	0.120	0.057	0.213	0.166
Domain 3	True data	0.163	0.030	0.071	0.039
	R=0	0.217	0.038	0.149	0.088
Domain 4	True data	0.055	0.015	0.124	0.156
	R=0	0.100	0.046	0.237	0.207
Domain 5	True data	0.014	0.006	0.013	0.019
	R=0	0.033	0.012	0.042	0.052

Table 9: Counterfactual domain CTRs if there are no externalities, i.e, $R = 0$

To interpret the quantitative significance of the externality parameters R and σ^R, we performed two counterfactual exercises. In the first, we consider a hypothetical impression with only two advertisements, and we compute the effect of satiation on the CTR of the advertiser at 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 low-quality ad is placed at position 1), and compare it with the CTRs with satiation for different actual competitors placed in slot 1. Table 8 presents the results. The biggest losses due to satiation occur in the “sex” string, on ads that compete with *Domain 1* at position 1. For example, the CTR of Domain 3 at position 2 would be almost four times higher if it did not compete with *Domain 1* at position 1. On the other hand, *Domain 1* itself, being a high-quality ad, does not suffer much from externalities: its CTR at position 2 would have been only 30%-40% higher had it faced no competition from position 1.

We perform the second counterfactual exercise 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 $R_i = 0$) and compare the results with the actual empirical CTRs. Table 9 presents the simulation results. 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 domains have similar CTRs; however, *Domain 4* gains much more in the counterfactual. Although,

		Search string			
		games	weather	white pages	sex
Domain 1	True data	0.104	0.090	0.171	0.268
	No uncertainty	0.104	0.095	0.189	0.287
Domain 2	True data	0.061	0.040	0.112	0.107
	No uncertainty	0.068	0.043	0.117	0.108
Domain 3	True data	0.163	0.030	0.071	0.039
	No uncertainty	0.181	0.030	0.072	0.041
Domain 4	True data	0.055	0.015	0.124	0.156
	No uncertainty	0.059	0.019	0.134	0.163
Domain 5	True data	0.014	0.006	0.013	0.019
	No uncertainty	0.015	0.007	0.013	0.020

Table 10: Counterfactual domain CTRs if there is no uncertanity, i.e, $\sigma = 0$

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 81% had satiation been absent.[18]

Table 7 contains estimates of the Bayesian signaling model. Overall, we find statistically significant evidence for uncertainty about ad quality. Moreover, the users use both position and description to form their beliefs prior to clicking. We allow the precision of the priors to vary across search strings and users. Consequently, we find statistically significant but small differences between search strings. However, we find relatively large and significant heterogeneity across users. One explanation is that some users are more experienced than others, but the experience does not depend on the search string. Additionally, we find the standard deviation of the prior is about 1.5 times higher than the standard deviation of the description signal. In other words, the description provides about 6 times more precise signaling of the quality than position.

We can also quantify the effects of user uncertainty about relevance by considering the counterfactual 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.) It is straightforward that eliminating user uncertainty will raise user welfare.

[18]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.

However, because of satiation, we are a priori unclear how doing so would affect CTRs of ads. Table 10 presents the CTR effects on each domain of removing uncertainty about the relevance of ads. We note the ads benefit in a heterogeneous way, which depends on their quality, composition of competitors in the data as well as positions they are usually presented on. In general, if the mean of the position prior is close to the true quality of an ad, the ad benefits less from removing the uncertainty. A good example is Domain 1 in the "games" search string. Its quality is close to the average quality of ads usually presented in the same position. The same is true for all aggregated domains marked as 5, which have low quality and are usually presented in low positions that have pessimistic priors.

5.1 Goodness of fit

To assess the our model's fit we investigated how well our model explains non-sequential clicks. Our model must be able to predict these events because they are our main motivation for developing a new structural model. We predict that among impressions that contain any clicks, users skip positions 44% of the time. In the data, we computed this number to be 46%, which means we can accurately predict the frequency of these events.

To test stationarity assumption, we investigate whether the goodness of our model's fit depends on time. We start by computing the probability of choosing the correct-observed bundle for each impression. Next, we slice the data into eight equal subsamples by time and compute the average of the simulated probabilities for each subsample. The goal is check the make sure if the prediction does not fluctuate too much over time, which could be indicative of a non-stationary environment. Figure 1 presents the results. We find the average subsample correct-choice probability is very stable. This stability suggests the true ad-quality measures as well as other parameters of the model do not change over time. Moreover, we observe more or less the same fit across search strings, which reassures us we do not overfit or underfit any of the search strings.

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. (Because we do not observe the advertisers' bids, we use the total CTR as our proxy for the search engine's revenue.) In particular, we are interested in considering the matching policy that maximizes the users' expected utility and a

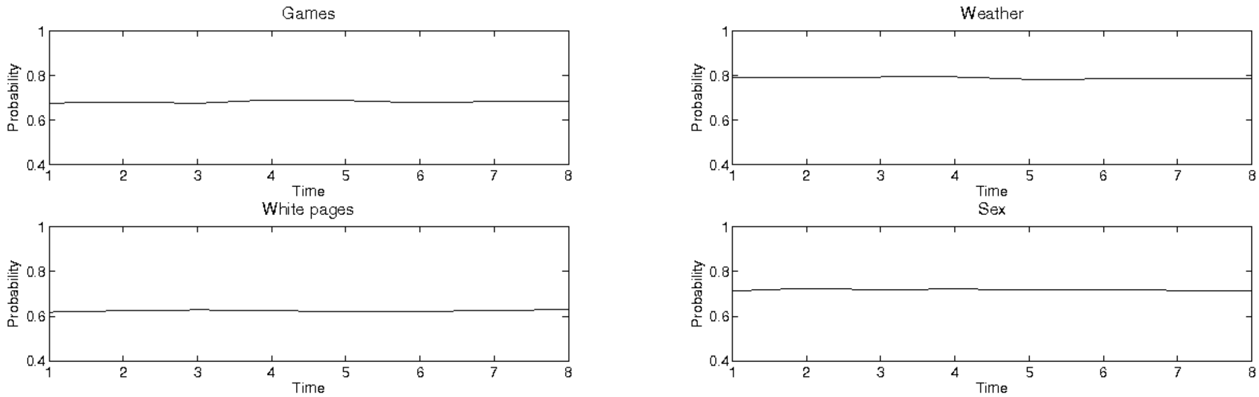

Figure 1: Average probability of choosing a correct bundle sliced by time and keyword.

potentially different policy that maximizes the total CTR.

A natural candidate for an optimal matching policy is *Assortative Matching (AM)*, in which the ads are displayed in decreasing order of their quality $\bar{v}_a$. (This policy is feasible for Microsoft, provided that they know the qualities of the different ads. We suspect 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 and uncertainty:

Proposition 6.1. *If $R_i = 0$, $\sigma = 0$ 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 the proposition 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 quality q in a position with cost f is $\pi(q, f) = \Pr\{q + \varepsilon > f\} = \max\left\{e^{q-f}, 1\right\} = e^q \cdot e^{-f}$. Because this function is supermodular in $\left(e^q, e^{-f}\right)$, a well-known result implies that assortative matching maximizes the total CTR.

The user's expected utility from having an ad with quality q at position with cost f, it can be computed as

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

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

When $R_i > 0$, the conclusion no longer holds, and we can find examples in which the total CTR or expected user utility i not maximized by AM. The intuition for how assortative matching can be

improved upon for users is that for two ads with the same quality $\bar{v}_a$ putting the ad with the higher posterior variance in a higher position might be optimal, so as to reduce the user's cost of learning its quality. (Similar changes raise the CTR but to a smaller extent.)

Example 6.1. *Suppose $R_i > 0$. There are two ads: $A = \{1, 2\}$, with ad 2 having no (or very negligible) posterior uncertainty. 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 much worse than expected. (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. Because strategy (c) yields the same payoff as if $R_i = 0$, 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 $(1, 2)$, because with some probability, user does not click on the high-cost slot. Thus, for parameter values for which strategy (b) is sufficiently likely to be optimal to the user on both impressions, displaying the uncertain ad above the certain ad is optimal, even if the certain as has higher expected quality.*

Similarly, if there no externalities are present, that is, $R_i = 0$, but uncertainty does exist, the assortative matching might be sub-optimal both for maximizing consumer surplus and for CTR.

Example 6.2. *Suppose there are two ads with description signals $x_1 = 10$ and $x_2 = 5$, which are equal to true qualities, and no user heterogeneity exists. The position priors have means 10 and 1, and the variance of the description noise and position priors is 1. The clicking costs are $f_1 = 0$ and $f_2 = 4$. When the ads are matched assortitatively, that is, ad 1 is placed on position 1, the posterior means are 10 and 3, CTR is 1, and the utility is 10. When the ad 1 is placed on position 2, the posterior means are 7.5 and 5.5, and the CTR is 2 and utility is 11.*

We simulate both user- and CTR-optimal matching policies on our data, and compare them to both assortative matching (ad qualities are obtained by estimating EOS model using OLS) and the actual data. As Table 11 shows, we find assortative matching that does not take into account uncertainty would raise user welfare by 10.62% and the total CTR by 6.76%. This matching policy does not coincide with either the user optimal policy and CTR optimal policy. The former raises welfare by 16.71%, whereas the latter raises CTR 15.69%. Both of these policies are Pareto improvements over assortative matching.

	Utility		CTR	
Data	0.35	-	0.24	-
No uncertainty	0.37	+4.47%	0.26	+6.09%
Assortative	0.39	+10.62%	0.26	+6.76%
Max U	0.411	+16.71%	0.265	+9.40%
Max CTR	0.407	+15.69%	0.268	+10.82%
First best U	0.56	+59.86%	0.322	+33.00%
First best CTR	0.44	+24.19%	0.37	+53.46%

Table 11: Short-run counterfactuals

The above assertions treat the average CTR as a proxy for total revenue. In reality, however, the revenue is a sum of clicks weighted by costs per click. When the costs per click are heterogeneous, one can give examples under which assortative matching gives suboptimal results. Consider the following example from the data.

Example 6.3. *Take Domains 2 and 3 from the "games" search string, and ignore all the other domains. In the data, we observe impressions when those domains switch places with each other. One way to rationalize this fact is that the search engine is indifferent between both placements. Moreover, suppose that the search engine is doing a bid-weighted assortative matching using the OLS estimates of EOS model. We can therefore infer that the bids for those ads have to be proportional to the inverse of the exponent of the OLS quality estimates, that is.* $\exp(-0.65)$ *and* $\exp(-0.92)$*. One can take those bids, and compute the search engine revenue under different matching policies of those ads to first and second position. OLS assortative matching gives about 6% less revenue than non-assortative matching.*

Unfortunately, given the available dataset, separately identifying costs per click is impossible without having data on advertisers' bids or valuations.

6.1 First-best targeting

Next we examine the improvements that "first-best" targeting could achieve, that is, conditioning the impressions on the user's utility characteristics ϵ_i, δ_i, R_i, user-specific priors as well as signals x_i. Such matching approximates the situation in which the search engine uses information about the consumers, such as the search history or demographics, to tailor the impression. Table 11

	Utility		CTR	
Data	0.35	-	0.24	-
No uncertainty	0.37	+4.47%	0.26	+6.09%
Max U	0.43	+21.30%	0.29	+19.53%
Max CTR	0.42	+20.75%	0.29	+19.71%
First best U	0.58	+63.72%	0.36	+50.40%
First best CTR	0.44	+26.03%	0.41	+68.04%

Table 12: Long-run counterfactuals

shows that moving toward first-best welfare-maximizing raises the users' expected utility by 60% from the actual data, and raises the total CTR by 33%. If we instead implement CTR-maximizing first-best targeting, we increase the CTR by 53%, however, with smaller gain to the utility (about 24%). The fact that CTR-optimal matching raises consumer surplus suggests that extra profit opportunities from exploring user-level targeting are also beneficial for the consumers; however, the welfare- and profit-maximizing incentives are not perfectly aligned. 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 did not target sponsored search results to individual users at the point of our writing a first draft of this paper. However, targeting display ads within web pages is now common (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 targeted advertising can achieve.

6.2 Long-run counterfactuals

Because we estimate position priors separately from position cost, we can investigate the extent to which users adapt their behavior to the changing matching policies. We computed utility- and CTR-maximizing second- and first-best matchings under the assumptions that position priors have zero variance and are correct. In such case, users would have no uncertainty about the quality of advertising. Table 12 reports the results. Additionally, the table includes a baseline case without uncertainty in order to quantify additional gain from counterfactual matching policies. We find the gain from second-best matching policies does not change much in the long run. For example, the

additional gain from “Max U” policy is about 21%, which is basically the same as the short-run gain combined with the no-uncertainty gain reported in Table 11. The same seems to be true for the long-run welfare impact in first-best matching. However, the long-run CTR impact of first-best matching is significantly higher. For example, “Max U” brings about 50% more clicks in the long run, compared to 33% in the short run (and 68% vs 53% for “Max CTR” policy). Therefore, we conclude that forward-looking companies should be more willing to invest in learning users’ tastes. Another take away is that significant complementarities exist when removing uncertainty and conducting user-level targeting. Companies might therefore be willing to invest in both removing uncertainty and targeting, even if investing in removing uncertainty alone is not profitable to .

7 Conclusion

This paper provides empirical evidence of externalities among ads in sponsored-search advertising, user heterogeneity, and user uncertainty regarding the quality of ads. We provide evidence, using both reduced-form tests and a structural models of expected utility-maximizing users.

The advantage of the structural model is that we can estimate the impact of externalities and uncertainty on CTRs of advertisers, the social welfare of consumers, and total CTR that is a proxy for profits for the search engine. We find a significant impact of both uncertainty (usually in the range of a couple percent, with a maximal 26% increase of CTR for one advertiser) and externalities (usually around a 50% drop in CTR) on advertisers’ CTR.

We also make counterfactual predictions for different ad-placement regimes and quantify “user experience” as the average user’s expected utility. We find that an alternative ad-placement policy could raise user welfare by 16.7%, and the increase could go up to 60% if information is available to target the placement to specific consumers. This finding suggests a large potential for ad targeting based on user level covariates, such as demographics or previous search history.

Because we separately identify the contribution of users’ expectations and position (behavioral) cost to differences CTR between positions, we can evaluate a long-run impact counterfactual policies. We find that short-run and long-run gains from counterfactual matchings are similar if the search engine cannot target ads using users’ characteristics. If it could target to specific consumers, it would achieve larger CTR gains (68% in long-run versus 53% in the short-run).

In this paper, we are unable to model long-run user learning explicitly. For example, we cannot evaluate how quickly users learn about the new ad-placement regimes. We believe studying long-run

user learning in more detail is important; however, our dataset did not allow is to do so, because it does not track users beyond the single search session. This direction is an important one 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 (2011): "Position Auctions with Consumer Search," *The Quarterly Journal of Economics*, 126(3), 1213–1270.

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.

Edelman, B., and M. Schwarz (2010): "Optimal Auction Design and Equilibrium Selection in Sponsored Search Auctions," *American Economic Review*, 100(2), 597–602.

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. D., and K. S. Sweeney (2009): "Bayes-nash equilibria of the generalized second price auction," in *Proceedings of the 10th ACM conference on Electronic commerce*, EC '09, pp. 107–108, New York, NY, USA. ACM.

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

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.*

Yang, S., and A. Ghose (2010): "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?," *Marketing Science*, 29(4), 602–623.

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