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

Przemyslaw Jeziorski† Ilya Segal‡

July 11, 2014

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 introduce 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 with 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 as 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. Specifically, 51% more clicks would occur in a hypothetical world in which each ad faces no competition. Moreover, our simulations indicate that clickthrough-optimal matching of ads to positions increases clickthrough rate by 23%, and user-optimal matching increases user welfare by 27%. Targeting ad placement to specific users could raise user welfare by 69%. Finally, user welfare could be raised by 1.6% if users had full information about the relevance of ads before they click on the link.

*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 grew 26% to reach $21.2 billion, according to the *Internet Advertising Revenue Report*, published by the Interactive Advertising Bureau and PricewaterhouseCoopers LLP.[1]

In order 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 takes a step in this direction, focusing on search advertising. Search advertising is performed by displaying a set of sponsored links accompanying results of consumers' Internet search queries. This form of advertising accounts for 41% of total Internet advertising revenues and is viewed as the most effective because of its precise targeting. In particular, a consumer's search string reveals a great deal about the products that the consumer is likely to be interested in; therefore, he can be shown only relevant ads, which in turn induce him to click. Although search advertising has recently received a lot of attention, researchers know little about consumer behavior in this market. This study contributes to the literature by analyzing consumer demand for sponsored ads.

Existing papers on search advertising postulate simple and restrictive models of user behavior. For example, Edelman, Ostrovsky, and Schwarz (2007) (henceforth OES) propose a model that assumes that the clickthrough rate (henceforth CTR) for a given ad in a given position is a product of an ad and position-specific effect, and does not depend on other ads that are displayed in the other positions. The EOS and its extensions were used in numerous studies, for example, Varian (2006), Edelman and Ostrovsky (2007), Yenmez (2009), Gomes and Sweeney (2009), Edelman and Schwarz (2010), Katona and Sarvary (2010) and Yao and Mela (2011). Another frequently used framework is a "cascade model" (Craswell, Zoeter, Taylor, and Ramsey (2008), Papadimitriou and Zhang (2008)). In this model 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.

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

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 (see Dupret and Piwowarski (2008)) is that once the model's parameters are estimated 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 the user's expected utility. Importantly, this utility is estimated directly from the preferences of actual users as revealed by their clicking behavior, rather than from the judgments of disinterested experts (see Carterette and Bennett (2008)). 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 in a form of sponsored links. An advertising impression is an ordered list of sponsored links. 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
```

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 describe the ads the user clicked and the times at which the clicks occurred.[2]

Our estimation strategy is based on the fact that searches using the same search strings 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 users' known characteristics or browsing histories. 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

[2] 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.

consider: "games," "weather," "white pages," and "sex".[3][4] In fact, it is easy to convince oneself of 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 CTR, and explicit experimentation by the engine. We believe that, at least for 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; that is, 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, they 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 positions 1 and 2, depends on the order in which these latter two ads are presented.

Next, we formulate and estimate a structural model of rational user behavior that nests the

[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 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.

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. That study assumes full satiation, that is, consumer gets utility from at most one purchase. In order to test for satiation our model instead parametrizes 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 in this case, no externalities are present 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, so he derives utility from at most one ad, and 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 to 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 such a Bayesian model, users have priors about the quality of ads on each position which can be different across users to control for the fact that different people might have had different ad experiences in the past. The users update the priors with signals about the quality of a particular ad contained in an ad description. We argue identification on the model's parameters, and jointly estimate these two mechanisms under the assumption that long-run user learning about quality of ads on each position 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 while others click 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 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 51% higher had satiation been absent. Moreover, we find evidence of user uncertainty: if this uncertainty were resolved prior to clicking, consumer welfare would be 1% higher with 0.4% higher overall CTR.[6]

[6]Note that in our model of expected utility maximization, cardinal utility has empirical meaning: impression A

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 the long-run, users learn exactly the new search engine placement policy collapsing the priors to degenerate distributions. We find that in the short run, the assortative matching provides about 11% improvement in welfare and 8% improvement in the CTR. In comparison, welfare-optimal placement provides 33% welfare improvement and CTR-optimal matching provides 23% 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.

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 69%, whereas CTR-optimal first-best targeting would raise the total CTR by 58%. Moreover, we find that these gains become more significant in the long run. In particular, welfare can be raised by 70% and CTR can be raised by 61%.

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

being x% better than impression B means that the user is indifferent to receiving impression B for sure and receiving impression A with probability $\frac{x}{100+x}$ and no ads at all with the complementary probability.

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 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 (for privacy reasons) does not keep track of user histories.

The paper is organized as follows. Section 2 describes the dataset and examines some reduced-form evidence. Section 3 presents 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 out of concerns for compromising user privacy. However, in 2008, Microsoft created a DVD with a sample of user search and advertising data, cleaned up to eliminate

[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 on 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.

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

privacy-compromising information. Microsoft distributed this DVD 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 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 dataset than shorter ones. Because the vast majority of the sessions contain only one impression, we believe sample selection issues are not of importance.

A Microsoft 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 we 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, applying 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 observed 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 website 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 that 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, in particular, 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 one another. Also, some of these externalities are inconsistent with the cascade model.

The externalities are evident by examining the dependence of CTR of a given ad in a given position on the relevance of the competing ads displayed in the same impression. For example, Table 1 presents the evidence for "externalities from above,"namely, we examine the dependence of CTR of *Domain 1*, displayed at position 2, conditional on different competitors displayed at position 1.[9] Comparison of the CTRs suggests negative externalities from the competitors. Namely, *Domain*

[9]We were able to conduct this analysis only for the most popular domain configurations that have enough observations to conduct statistical tests.

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: CTRs on *Domain 1* in search string "white pages" and *Domain 1* in search string "weather" at position 2 conditional on different domains at position 1. Standard errors are provided in parentheses. The estimates have asymptotic normal distributions.

1 in the "white pages" search string receives higher CTR if the competitor above has low relevance (e.g. *Domain 3*, which does not have any white-page information) rather than high relevance (e.g. *Domain 2* or *Domain 4*). We obtain the same conclusion when examining the "weather" search string. Specifically, *Domain 1* has higher CTR when facing *Domain 3* (which does not have any weather information) as the above competitor than facing *Domain 2* (which is a well known weather website). These differences in CTR are statistically significant (with a p-value of less than 1%) and 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, this time from *below*. Specifically, we compare CTRs of different domains, placed in position 1 of "white pages" search string, between those impressions that contain high relevance *Domain 1* as competitor in any of the lower positions and impressions that do not contain *Domain 1*. We find that the differences in these CTRs are statistically significant. Note that we do not control for the overall number of ads in the relevant impressions which can affect the CTR of position 1. Nevertheless, we show that our tests are a conservative assessment of the impact of *Domain 1* because the average number of competitors is lower in the impressions that have *Domain 1*. 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 choosing which ones to click.

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 at position 3. We were able to

	Domain 1 present as a competitor in lower positions?	CTR of position 1	Difference in CTRs of position 1	No. obs.	Average no. of ads in the impression
Domain 2	Yes	0.11	0.074*** (0.009)	5,061	6.1
in position 1	No	0.18		2,112	7.1
Domain 3	Yes	0.16	0.067*** (0.013)	1,560	7.1
in position 1	No	0.22		2,022	7.6
Domain 4	Yes	0.16	0.019 (0.032)	304	7.3
in position 1	No	0.17		253	7.3

Table 2: Comparison of CTRs of domains at position 1, in the "white pages" search string, between impressions that contain high relevance *Domain 1* as competitor in any of the lower positions and impressions that do not contain *Domain 1*. Statistically significant difference of CTRs between these two cases is evidence for the externalities "from below." Three stars means statistical significance at 1% level.

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 2 is 0.0434. When we switch the top two ads, the CTR drops to 0.0077; this drop is statistically significant at 0.05 level (an asymptotic Wald test statistic is 2.193 is distributed as standard normal). As mentioned earlier, we believe *Domain 3* is not a relevant domain for "weather," whereas *Domain 2* is. Thus, assortitatively matching the better competitor domains with the higher position in the top two positions negatively impacts the CTR of the lower ad. This externality can again be attributed to user satiation because assortative matching increases the likelihood of the user clicking on the better domain first, making him more satiated and less likely to subsequently click on an ad in a third position.

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 a more important source of externalities than learning. Identifying these two effects separately, given

	Search string and domain at position 1			
	games Domain 1	weather Domain 1	white pages Domain 1	sex Domain 1
Clicking on pos. 1	0.051	0.046	0.17	0.037
Not clicking on pos. 1	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 ads in positions 2-8 conditional on clicking and not clicking on position 1.

our dataset, 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 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 ad (the most popular) shown at 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 at position 1 on the "games" 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 user valuations of ads. For example, on the one hand some users may have higher utilities for all ads (e.g., due to higher beliefs about the relevance of sponsored-search advertising), and on the other hand, some users may have lower utility for all ads. Such heterogeneity has to be large enough to offset the negative correlation among clicks created by satiation. Beyond explaining a correlation in clicks, the vertical heterogeneity is also needed to explain disproportionate numbers of multiple-clicks observations (or "bundles"). Namely, our model without satiation, uncertainty, and with independent clicks (which is then equivalent to the EOS model) would predict only 911 bundles of 2 clicked ads, compared to 1,157 in the data, and only 20 bundles of 3 clicked ads, compared to 188 in the data. Note that introducing satiation into the model increases this discrepancy further as people click fewer bundles, thus, we are likely to find more vertical heterogeneity than one would find using a model without satiation.

3 Model

Consider a user i who faces an impression $[a(1), ..., a(N)]$, where N is a number of displayed advertisements, and $a(n) \in A$ is an ad displayed in slot n. The user's value of clicking on an ad a (or a quality of the link a) is given by

$$v_{ai} = v_a + \epsilon_{ai} + \delta_i,$$

where v_a is the mean quality of the link a common to all users, ϵ_{ai} is the idiosyncratic shock to quality of the link a, and δ_i is the user specific aggregate shock to value of any sponsored links.[10]

The user can click a subset $\mathcal{C}$ of ad slots and obtain a gross utility of

$$U_i(\mathcal{C}) = \left(\sum_{n \in \mathcal{C}} v_{a(n)i}^{1+R_i} \right)^{1/(1+R_i)} - \sum_{n \in \mathcal{C}} f_n, \tag{3.1}$$

where f_n is a cost of clicking on an ad in position n, and R_i is a parameter that captures the substitutability of different ads to the user. We allow for user heterogeneity in R_i and assume that $R_i = R + \sigma_R \eta_i$, where η_i is a standard normal random variable.

Users do not observe the quality v_{ai} prior to clicking, but have access to ad descriptions. Ad descriptions provide full information about the idiosyncratic parts ϵ_{ai} and the aggregate shock δ_i and convey unbiased signals x_{ai} about mean qualities v_a. Prior to reading the description users have priors about v_a at each slot, which are distributed as independent normal random variables with means $\bar{v}_n$ and variances $\bar{\sigma}_i^2$.[11] These priors are a result of a long-run cross-impression learning

[10] Note that the value of an ad does not depend on the position of an ad.

[11] We model position priors as independent random variable, but we allow for arbitrary correlation of actual qualities v_a across positions.

process.[12] We allow that the priors' means depend on the position to capture that users have different beliefs about the quality of advertising at each slot. However, we assume that prior variances $\bar{\sigma}_i^2$ are the same across positions. Theoretically, our model facilitates identification of the position-specific variance, but we failed to reject that these variances are the same with a 1%-size test. Further, we assume that position priors are unbiased, that is, the mean of quality of displayed ads in a particular position is the same as the mean of users' prior. To capture that some users might be more influenced by positions we allow the precision of the prior to vary across users. Specifically, we set $\bar{\sigma}_i = \bar{\sigma} + \sigma_{\bar{\sigma}}\zeta_i$, where ζ_i is a standard normal random variable. An extreme case, when $\bar{\sigma}_i$ is small, approximates cascading models in which the user clicks from top to bottom no matter which ads are shown. At the other extreme, when $\bar{\sigma}_i$ is large, the position of an ad does not influence its CTR.

In order to nest EOS model 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 $F_\epsilon(\varepsilon_{ai}) = 1 - e^{-\varepsilon_{ai}}$). As for δ_i, it is drawn from a normal distribution with mean zero and variance σ_δ^2.

The description signal about v_a is given by $x_{ai} = v_a + \sigma_\nu \nu_{ai}$, where ν_{ai} is a standard normal random variable.[13] The posterior beliefs about the $v_{a(n)i}$ are given by independent normal distributions with means

$$\hat{\mu}_{a(n)i} = \left(\frac{x_{a(n)i}}{\sigma_\nu^2} + \frac{\bar{v}_n}{\bar{\sigma}_i^2}\right) / \left(\frac{1}{\sigma_\nu^2} + \frac{1}{\bar{\sigma}_i^2}\right) + \epsilon_{a(n)i} + \delta_i,$$

and the variances

$$\hat{\sigma}^2_{a(n)i} = \left(\frac{1}{\sigma_\nu^2} + \frac{1}{\bar{\sigma}_i^2}\right)^{-1}.$$

Note that users know idiosyncratic components ϵ_{ai} and δ_i prior to clicking and learn the mean component v_a after clicking. An alternative choice would be to assume that users know the mean component and learn about the idiosyncratic components (in a similar way as Kim, Albuquerque, and Bronnenberg (2010)). In our case, such alternative assumptions would force the rational users to have homogeneous priors across positions and would rule out rational position clicking bias. Such homogeneity is a consequence of the search engine's inability to target advertising on individual characteristics, which implies that the position is independent of ϵ_{ai} and δ_i.[14] Consequently, in this

[12] In this study we do not have data on the behavior of users across impressions, so we are unable to describe how these priors are formed. However, under the assumptions outlined in this section, we are able to estimate the priors from the data.

[13] Similar to prior variances, we are technically able to allow σ_ν^2 to vary across ads, however, we failed to reject that these variances are the same with a 1%-size test.

[14] Note that the ad position is likely correlated with v_a.

study we use the simplest version of the rational model that supports different position priors in which users learn about v_a and have full information about ϵ_{ai} and δ_i. An important byproduct of this setup is maintaining conjugate priors while nesting the EOS model (such nesting requires ϵ_{ai}s to be exponential random variables). To test if such an informational assumption is important, we estimated a version of the model in which users are uncertain about both v_a and ϵ_{ai} and find quantitatively similar results.

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 $((\bar{v}_1, \bar{\sigma}_i), ..., (\bar{v}_N, \bar{\sigma}_i))$.

(ii) The user observes the impression $[a(1), ..., a(N)]$ and reads descriptions of all ads in the impression. The user learns ϵ_{ai} and δ_i and forms posterior beliefs about mean qualities v_a using description signals $(x_{1i}, \ldots, x_{Ni})$.

(iii) The user either clicks on an ad in a chosen position c or stops clicking (exits).

(iv) The user observes the true quality v_a of a clicked ad $a(c)$.

(v) Go to (iii).

We assume the user is a forward-looking expected-utility maximizer and knows all the parameters. His decision process can be modeled as a dynamic programming problem whose payoff-relevant state can be summarized with a set $\mathcal{C} \subset \{1, ..., N\}$ of clicked positions and a sufficient statistic about the utility of $\mathcal{C}$, given by $S = \sum_{n \in C} v_{a(n)i}^{1+R_i}$. The optimal continuation value of user i in state $(\mathcal{C}, S)$, which we denote by $V_i(\mathcal{C}, S)$, is governed by the following Bellman equation:

$$V_i(\mathcal{C}, S) = \max \left\{ S^{1/(1+R_i)} - \sum_{n \in \mathcal{C}} f_n, \max_{c \in \{1,..,N\} \backslash \mathcal{C}} EV_i \left(\mathcal{C} \cup c, S + v_{a(c)i}^{1+R_i} \right) \right\} \tag{3.2}$$

The expectation is taken with respect to the posterior of $v_{a(c)i}$ described earlier. This model is rich enough to nest the following special cases:

- $R_i = 0$ (additively separable utility), $\sigma_\nu = 0$ (no uncertainty): The user's clicking decisions on different ads are then independent, and no externalities are present across ads. Additionally, if user random effects are absent (i.e., $\sigma_\delta^2 = 0$), the clicks on the different positions are statistically independent, and the CTR on ad a at position n is $\Pr\{v_a + \varepsilon_{ai} - f_n \geq 0\} = F(f_n - v_a) = \min\{e^{v_a} e^{-f_n}, 1\}$. 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^{-v_a}) and the position fixed effect (e^{-f_n}).[15] This nesting is the key motivation for our adoption the exponential distribution of errors ε_{ai}, and allows a simple quantitative interpretation of the estimated fixed effects on the CTR. In the EOS case, a consistent estimate of the fixed effects 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, v_a can be identified. Note also that because only the differences $v_a - f_n$ are identified in the EOS model, the fixed effects f_n and v_a are identified only up to a constant.

- Perfect substitutability: $R_i = \infty$. In this case, the user's utility asymptotes to

$$U_i(\mathcal{C}, \mathbf{v}) = \max_{n \in \mathcal{C}} v_{a(n)i} - \sum_{n \in \mathcal{C}} f_n;$$

that is, the user derives utility from at most one ad (e.g., he derives no benefit from viewing a second weather forecast). This is reminiscent of the classical consumer search model (e.g., Weitzman (1979) and Kim, Albuquerque, and Bronnenberg (2010)). We can also approximate "cascade models" in which users click positions top to bottom by assuming that position clicking costs f_n increase sharply at position n relative to any variation in ad quality or that the precision of the position priors is low.

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 $[(\bar{v}_1, \bar{\sigma}_{ni}), ..., (\bar{v}_N, \bar{\sigma}_{Ni})]$ 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 (3.2) by backward

[15] If $\sigma_\delta^2 > 0$ and is not too large, 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.

induction. The solution produces an optimal policy as a function of a set of clicked ads $\mathcal{C}$ as well as the sufficient statistic, S, about the accumulated utility. We compute the user-level moments by running this policy forward 100 times and taking the average of user-level moments. In each iteration of the estimation algorithm we use 500 draws of the random components per observation, which amounts to solving about 50 million dynamic programming problems.[16] Note that the model cannot be solved using the reservation value approach described in Weitzman (1979), because Weitzman's model assumes extreme degree of satiation, that is $R_i = \infty$. One can show that for intermediate values of R_i the reservation value for the next click depends on the state S, whereas in Weitzman (1979) it was independent of the state. In practice, the reservation utilities have to be recomputed for each S, which is equivalent to fully solving the Bellman equation. Further details of the computational procedure are presented in Appendix A.

We do not explicitly model cross-visit user learning. Instead, we infer users' position priors using the assumption that the cross-impression learning generates unbiased priors and that ad placement is stationary. Under the first assumption, $\bar{v}_n = E[v_a|n]$. Under the second assumption, we can estimate the mean of the position prior by using the sample equivalent; that is, we set

$$\bar{v}_n = \sum_a \phi_{na} v_a, \tag{4.1}$$

where ϕ_{na} is the observed frequency of placing an ad a on the position n. We preestimate ϕ_{na} from the data and use it during the GMM estimation procedure to construct an estimate of $\bar{v}_n$ using parameter values of of v_a and equation (4.1).

Our model has 49 unknown parameters and identifies them using 78 conditional micro moments, which are described at the end of this section. The moments are conditional on the search string and ad placement in an impression. Additionally, micro moments are clustered by time stamp using $T = 88$ equally sized groups of $M = 1047$ observations which correspond to daily clustering.[17] Such a procedure allows for an arbitrary correlation of random effects within the day. Formally, the estimator minimizes the objective function

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

[16] Initial computations were possible thanks to supercomputer resources provided by Microsoft Corporation. The most recent results were obtained using a 112-core Berkeley computer cluster.

[17] We tried many different levels of clustering and find no significant differences between asymptotic distributions of the estimates. For this reason, we decided to report the clustering level that makes the least assumptions about the distribution of the random effects.

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

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

where $\bar{g}_{tm}$ are the micro moments observed in the data for the impression m in cluster t, $g_{tm}(\theta|\epsilon,\delta)$ are the micro moments predicted by the model conditional on (ϵ,δ), and $d(\epsilon,\delta)$ is the distribution of the random components. The integral of the random components is approximated using a frequency estimator with $L = 500$ IID draws of (ϵ,δ) for each observation.

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

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

and

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

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

To obtain an estimator of Ω we need to make further assumptions about the correlation of G_t across clusters; that is, we assume that G_t follows an MA(1) process. An important consequence of this assumption is that ad placement and random effects can be serially correlated across days. Consequently, we use Newey and West (1987) estimator of Ω given by

$$\hat{\Omega}_T = \left(1 + \frac{1}{L}\right)\left[\hat{\Omega}_0 + \frac{1}{2}(\hat{\Omega}_1 + \hat{\Omega}_1')\right],$$

where

$$\hat{\Omega}_0 = \frac{1}{T}\sum_{t=1}^{T} G_t G_t' \qquad \hat{\Omega}_1 = \frac{1}{T-1}\sum_{t=2}^{T} G_t G_{t-1}'$$

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

In practice we perform the estimation in several steps. We start with the OLS estimates of the OES model, $\hat{\theta}^{(0)}$, and compute the corresponding estimate of the optimal weighting matrix $\hat{W}^{(0)}$. Next, we obtain $\hat{\theta}^{(1)}$ using GMM and repeat the steps until convergence. To minimize the computation time we gradually increase the number of draws, L, until we reach 500.

In the reminder of this section we discuss the particular choice of moment and identification. Our parameters are divided into four groups:

(i) domain mean qualities v_a,

(ii) position fixed effects f_n, variance of the priors $\bar{\sigma}_n^2$ for each keyword and heterogeneity of this variance $\sigma_{\bar{\sigma}}^2$,

(iii) variance, σ_δ^2, of the user random effect,

(iv) the satiation parameters (R, σ_R^2), the domain/position normalizing constant, and the precision of the description signals σ.

We discuss the identification of all groups separately.

We can identify effects of parameters in the group (ii) separately from 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 search-string-specific probabilities of clicking on each position from 1 to 5 and clicking probabilities for each domain (we dropped the moments proven to have 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 can identify ad qualities v_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.[18]

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. On the one hand, if the drop of CTRs on lower positions is caused by position priors, good ads should benefit from top positions less than bad ads and good ads should lose CTR from lower positions more than bad ads. On the other hand, if the CTR drop is caused by clicking cost (position fixed effects), we should not see interactions between positions and ads. To capture these effects, we include domain-position-specific CTRs for the top two advertisers placed at each of the top three positions. Including such

[18] Because the number of clicks on positions 6 and 7 is very small, we assume the cost of clicking on those are 10% and 30% higher, respectively, than on position 5; these numbers don't affect the estimation.

moments is similar to supplementing an OLS regression of log-CTR on position and ad dummies, that is used to estimate an EOS model, with interaction terms.[19]

To identify the standard deviation σ_δ of the user random effect, we include the unconditional probabilities of bundles of two and three clicks. Increasing σ_δ boosts the correlation of clicks on different ads in the same impression, and increases the probabilities of clicking bundles. These moments also identify the variance of R_i, because the higher variance generates more bundle clicks.

One of the main contributions of this paper is identifying the user's 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 few 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 three such moments per search string for the three most popular domains. The satiation parameter is identified from these moments, as 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. Similarly, we can identify the precision of the description signal, which determines the amount of user uncertainty prior to clicking. In particular, if we observe domains with a large CTR but also with a large continuation probability, we infer that precision of the description is low.

[19]We find the interaction terms in the log-CTR regressions to be statistically different from zero, which is consistent with different position priors.

We argued in Section 2 that cascade model and AE model are not very realistic because they cannot explain 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 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-Marquard gradient method[20] 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, 6, 7, and 8 present the estimates of the model. Table 4 presents the estimated position clicking costs for each search string. Table 6 presents the estimated quality measures of selected domains, organized by search string. Table 7 contains the estimates of the satiation parameter R and the user heterogeneity parameter σ_δ^2. Finally, Table 8 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. To interpret the magnitude of numbers presented in Table 4, 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 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,

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

	Search string			
	games	weather	white pages	sex
Position 1	−1.66 (0.031)	−1.10 (0.028)	−1.94 (0.025)	−2.01 (0.053)
Position 2	−1.64 (0.023)	−1.71 (0.032)	−2.59 (0.024)	−2.20 (0.051)
Position 3	−2.07 (0.020)	−2.12 (0.030)	−3.25 (0.031)	−2.54 (0.065)
Position 4	−3.68 (0.075)	−3.84 (0.048)	−5.30 (0.091)	−3.18 (0.206)
Position 5	−3.98 (0.140)	−4.28 (0.104)	−5.52 (0.201)	−4.38 (0.611)

Table 4: Estimates of clicking cost in the baseline model.

	Search string			
	games	weather	white pages	sex
Position 1	−2.16 (0.049)	−3.37 (0.018)	−1.59 (0.036)	−1.74 (0.037)
Position 2	−2.46 (0.042)	−3.89 (0.018)	−1.46 (0.037)	−2.62 (0.037)
Position 3	−2.48 (0.042)	−3.61 (0.012)	−1.87 (0.038)	−2.89 (0.043)
Position 4	−2.65 (0.039)	−3.72 (0.016)	−2.31 (0.044)	−3.05 (0.057)
Position 5	−2.74 (0.038)	−4.01 (0.023)	−2.49 (0.047)	−3.09 (0.062)
Position 6	−2.79 (0.038)	−4.22 (0.029)	−2.62 (0.050)	−3.12 (0.066)
Position 7	−2.82 (0.038)	−4.38 (0.033)	−2.68 (0.051)	−3.13 (0.066)
Position 8	−2.83 (0.038)	−4.46 (0.035)	−2.69 (0.051)	−3.13 (0.067)

Table 5: Estimates of the mean of position priors.

	Search string			
	games	weather	white pages	sex
Domain 1	-2.04 (0.054)	-2.70 (0.040)	-0.35 (0.057)	-0.38 (0.038)
Domain 2	-2.06 (0.051)	-3.66 (0.021)	-1.08 (0.040)	-2.11 (0.046)
Domain 3	-1.27 (0.055)	-3.82 (0.046)	-1.26 (0.033)	-2.70 (0.051)
Domain 4	-2.21 (0.058)	-2.70 (0.040)	-0.80 (0.081)	-1.36 (0.097)
Domain 5	-2.83 (0.036)	-4.53 (0.034)	-2.71 (0.044)	-3.13 (0.073)

Table 6: Estimates of domain quality and probabilities of relevance.

Satiation parameter		Preference shock
Mean (R)	Std. dev. (σ_R)	Std. dev. (σ_δ)
0.57 (0.026)	0.18 (0.028)	2.15 (0.027)

Table 7: Estimates of the user specific random effects: satiation parameters and preference shocks.

Std. dev. of the prior ($\bar{\sigma}$)				User random effect	Description noise
games	weather	white pages	sex	std. dev. ($\sigma_{\bar{\sigma}}$)	std. dev. (σ)
0.86 (0.210)	0.94 (0.127)	0.83 (0.144)	1.20 (0.107)	0.08 (0.064)	0.25 (0.023)

Table 8: Estimates of the learning process: standard deviations of the position priors for each keyword, user level random effect, and standard deviation of the noise in a description signal.

in the “games” search string, the CTR of a given ad at position 1 is $\exp(3.98 - 1.66) \approx 10$ times higher at position 1 than position 5. In the “weather” search string, the ratio is $\exp(4.28 - 1.10) \approx 24$.

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 does not come 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 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 inexpensive, 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 positions 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. Finally, we find no statistically significant differences in clicking cost of top positions in the “games” keyword, which suggests no additional value of winning slot 1 instead of 2.

Means of position priors are reported in Table 5 and are computed using the equation (4.1). The expectations about quality are monotonic, that is, people expect better quality ads on higher positions. We note that the quality drops sharply when moving from position 1 to 4 and remains roughly constant across positions 4 to 8. Moreover, we note that impact of the position prior on clicking bias depends on the search string. On the one hand, in “games” search string the drop in expectations between positions 1 and 4 is negligible, while the increase of the clicking cost is substantial. It suggests that most of the position bias in this keyword is driven by the psychic cost. On the other hand, in “sex” search string the expectation difference between positions 1 and 4 is substantial, while the difference in cost is negligible, which suggests that most of the clicking bias is driven by rational expectations.

Table 6 presents the estimates of the mean qualities 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 for 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 much lower quality than *Domain 3*. The structural model attributes this 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. The lower quality of this domain might be attributed to the domain's description as a service with maps; therefore, users might correctly think the domain does not contain weather information.

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 website, *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, *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 7 presents our estimates of satiation and user heterogeneity. 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 $\exp \sigma_\delta \simeq 9$.

Table 8 contains estimates of the Bayesian signaling model. Overall, we find statistically significant evidence for uncertainty about ad quality. Moreover, the users employ 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 small but statistically significant differences between search strings. Likewise, we find relatively small and insignificant heterogeneity across users. Additionally, we find the standard deviation of the prior is about 4-5 times higher than the standard deviation of the description signal. In other words, the description provides about 4-5 times more precise signaling of the quality than position.[21]

[21] In an alternative model in which users learn about $v_a + \epsilon_{ai}$ we find statistically significant heterogeneity in position

To interpret the quantitative significance of the externality parameters R and σ^R, we perform two counterfactual exercises. In the first, we consider a hypothetical impression with only two advertisements, and 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 9 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 three 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 40%-50% higher had it faced no competition from position 1.

priors across users.

Games

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.057 (0.004)	0.053 (0.003)	0.063 (0.004)	0.070 (0.005)	0.121 (0.003)
(d2)	0.112 (0.003)	-	0.084 (0.003)	0.096 (0.004)	0.112 (0.004)	0.159 (0.002)
(d3)	0.138 (0.011)	0.114 (0.012)	-	0.126 (0.011)	0.141 (0.011)	0.199 (0.007)
(d4)	0.094 (0.004)	0.076 (0.004)	0.068 (0.003)	-	0.094 (0.004)	0.144 (0.003)
(d5)	0.064 (0.002)	0.056 (0.001)	0.053 (0.002)	0.059 (0.001)	-	0.126 (0.003)

Weather

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.049 (0.001)	0.053 (0.001)	0.045 (0.001)	0.054 (0.001)	0.080 (0.003)
(d2)	0.041 (0.001)	-	0.044 (0.001)	0.037 (0.001)	0.045 (0.001)	0.062 (0.001)
(d3)	0.027 (0.001)	0.028 (0.002)	-	0.027 (0.001)	0.034 (0.001)	0.045 (0.002)
(d4)	0.054 (0.001)	0.058 (0.002)	0.064 (0.003)	-	0.065 (0.002)	0.097 (0.004)
(d5)	0.027 (0.001)	0.027 (0.001)	0.032 (0.001)	0.026 (0.001)	-	0.048 (0.002)

White pages

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.136 (0.004)	0.156 (0.004)	0.128 (0.003)	0.167 (0.005)	0.285 (0.004)
(d2)	0.068 (0.001)	-	0.097 (0.001)	0.078 (0.001)	0.110 (0.002)	0.223 (0.005)
(d3)	0.056 (0.001)	0.064 (0.001)	-	0.059 (0.002)	0.080 (0.004)	0.190 (0.005)
(d4)	0.083 (0.002)	0.103 (0.003)	0.116 (0.004)	-	0.128 (0.003)	0.250 (0.006)
(d5)	0.047 (0.001)	0.050 (0.002)	0.059 (0.001)	0.049 (0.001)	-	0.165 (0.002)

Sex

	(d1)	(d2)	(d3)	(d4)	(d5)	R=0
(d1)	-	0.169 (0.005)	0.184 (0.004)	0.158 (0.010)	0.187 (0.004)	0.258 (0.008)
(d2)	0.056 (0.002)	-	0.089 (0.002)	0.071 (0.003)	0.094 (0.002)	0.150 (0.003)
(d3)	0.039 (0.001)	0.045 (0.002)	-	0.042 (0.001)	0.056 (0.001)	0.113 (0.002)
(d4)	0.062 (0.005)	0.093 (0.010)	0.112 (0.007)	-	0.115 (0.008)	0.179 (0.013)
(d5)	0.032 (0.001)	0.038 (0.001)	0.044 (0.001)	0.036 (0.002)	-	0.105 (0.003)

Table 9: CTRs of the counterfactual impressions with only two advertisers. Each cell contains a CTR of the advertiser in position 2 for different ad configurations. Specifically, each row represents a given domain in position 2 and each column prescribes a different competitor domain in position 1. The last column with $R = 0$ presents CTR of an ad in position 2 conditional on a "dummy competitor" in position 1 that creates no satiation.

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 10 presents the simulation results. Unlike 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 better domains generally tend to lose less CRT 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 51% had satiation been absent.[22]

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. 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 that 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 fit of our model we investigate how well our model explains non-sequential clicks. Our model must be able to predict these events because they provide 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.

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

		Search string			
		games	weather	white pages	sex
Domain 1	True data	0.093 (0.0011)	0.075 (0.0012)	0.131 (0.0012)	0.251 (0.0038)
	R=0	0.142 (0.0021)	0.097 (0.0019)	0.255 (0.0043)	0.329 (0.0049)
	No uncertainty	0.092 (0.0011)	0.077 (0.0016)	0.139 (0.0027)	0.257 (0.0038)
Domain 2	True data	0.042 (0.0013)	0.021 (0.0007)	0.086 (0.0007)	0.067 (0.0018)
	R=0	0.094 (0.0022)	0.039 (0.0016)	0.187 (0.0035)	0.125 (0.0030)
	No uncertainty	0.042 (0.0019)	0.021 (0.0007)	0.087 (0.0014)	0.066 (0.0017)
Domain 3	True data	0.133 (0.0056)	0.023 (0.0010)	0.068 (0.0010)	0.029 (0.0012)
	R=0	0.203 (0.0073)	0.037 (0.0018)	0.157 (0.0031)	0.073 (0.0021)
	No uncertainty	0.141 (0.0040)	0.022 (0.0010)	0.068 (0.0013)	0.028 (0.0012)
Domain 4	True data	0.045 (0.0014)	0.012 (0.0004)	0.091 (0.0004)	0.129 (0.0084)
	R=0	0.091 (0.0027)	0.045 (0.0019)	0.206 (0.0052)	0.194 (0.0099)
	No uncertainty	0.045 (0.0018)	0.012 (0.0005)	0.094 (0.0038)	0.130 (0.0089)
Domain 5	True data	0.011 (0.0003)	0.003 (0.0002)	0.006 (0.0002)	0.013 (0.0009)
	R=0	0.033 (0.0012)	0.009 (0.0006)	0.025 (0.0009)	0.040 (0.0020)
	No uncertainty	0.011 (0.0003)	0.003 (0.0001)	0.005 (0.0002)	0.012 (0.0009)

Table 10: Counterfactual domain CTRs if there are no externalities, i.e., $R = 0$, and if there is no uncertainty, i.e., $\sigma = 0$.

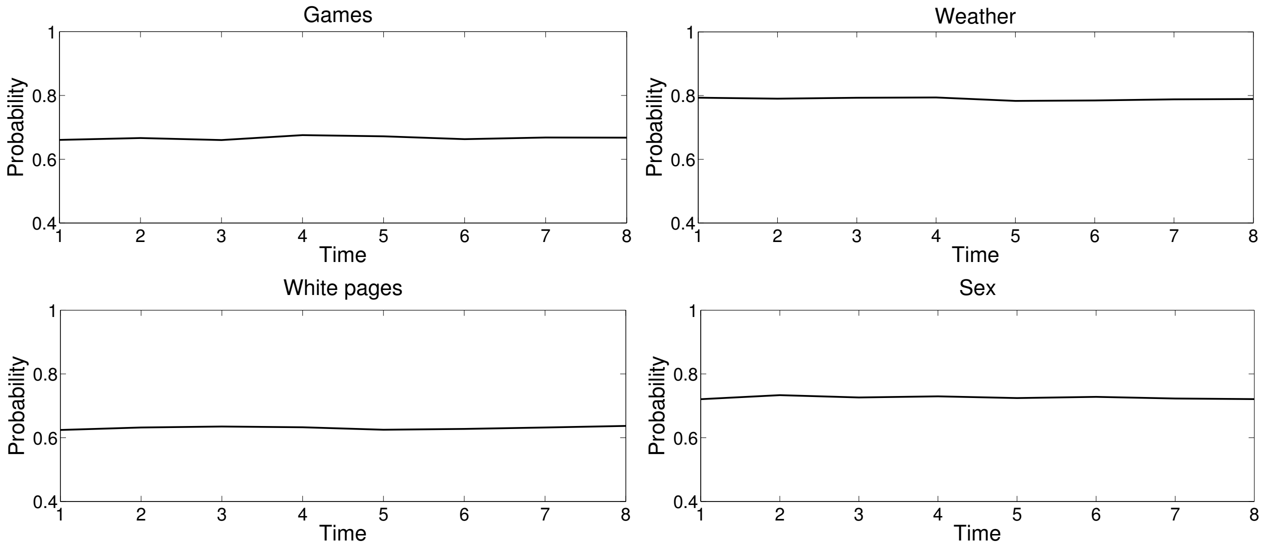

Figure 1: Average probability of choosing a correct bundle sliced by time and keyword.

To test the stationarity assumption, we investigate whether the goodness of our model's fit depends on time. We start by computing the predicted probability of choosing the observed bundle for each impression in the data. Next, we slice the data into eight equal subsamples by time and compute the average of the probabilities for each subsample. The goal is to investigate if the prediction error of the model does not fluctuate too much over time, which could be indicative of a non-stationary environment. Figure 1 presents the results. We find that the probabilities of choosing observed bundle are very stable. This stability suggests that 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 that we do not overfit or underfit any of the search strings.

Another common test of specification and fit is the test of overidentified restrictions. Under the null hypothesis of the correctly specified model and correct distributional assumptions of the moment conditions, namely clustering and serial correlation, the following statistic

$$\mathcal{H}_T = \frac{1}{T}\left(\sum_{t=1}^{T} G_t'\right) W_T \left(\sum_{t=1}^{T} G_t\right)$$

is asymptotically distributed as χ^2 with 29 degrees of freedom (number of moment conditions minus number of parameters). We obtain that $\mathcal{H}_T = 44.2$; we cannot reject the null hypothesis using 2.5%-level test.

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.[23] In particular, we are interested in considering the matching policy that maximizes the users' expected utility and a 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 v_a. Assortative matching is feasible for Microsoft, provided that the company knows 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 and uncertainty:

Proposition 6.1. *If $R_i = 0$, $\sigma = 0$ and each ad receives a CTR of 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 implies that the same is true on expectation over δ_i.

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

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

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

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 is 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 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 lesser extent.

[23] Because we do not observe the advertisers' bids, we use the total CTR as our proxy for the search engine's revenue.

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, 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, according to the above proposition. However, the payoff from strategy (d) is maximized on impression $(1, 2)$, because with some probability, the user does not click on the high-cost slot. Thus, for parameter values for which strategy (d) 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 ad has higher expected quality.*

Similarly, if no externalities are present, that is, if $R_i = 0$, 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 ad 1 is placed on position 2, the posterior means are 7.5 and 5.5, 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 would raise user welfare by 11% and the total CTR by 8%. This matching policy does not coincide with either the user optimal policy and CTR optimal policy. The former raises welfare by 33%, whereas the latter raises CTR by 23%. Both of these policies are Pareto improvements over assortative matching.

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.

	Utility		CTR	
Data	0.25 (0.004)	-	0.18 (0.001)	-
No uncertainty	0.25 (0.003)	+1.62% (0.28)	0.18 (0.001)	+0.38% (0.53)
Assortative	0.28 (0.007)	+11.24% (2.15)	0.20 (0.003)	+8.39% (1.32)
Max U	0.33 (0.004)	+32.78% (0.99)	0.22 (0.002)	+21.18% (0.78)
Max CTR	0.32 (0.006)	+27.41% (1.29)	0.22 (0.002)	+23.20% (0.86)
First best U	0.42 (0.007)	+69.13% (2.08)	0.26 (0.002)	+42.13% (1.21)
First best CTR	0.33 (0.006)	+32.01% (1.73)	0.29 (0.002)	+57.89% (1.40)

Table 11: Short run counterfactuals

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 7% 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 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 shows that moving toward first-best welfare-maximizing raises the users' expected utility by 69% from the actual data, and raises the total CTR by 42%. If we instead implement CTR-maximizing, first-best

	Utility		CTR	
Data	0.25 (0.003)	-	0.18 (0.001)	-
No uncertainty	0.25 (0.003)	+1.62% (0.36)	0.18 (0.002)	+0.38% (0.56)
Assortative	0.30 (0.010)	+22.59% (4.15)	0.21 (0.005)	+16.82% (2.70)
Max U	0.33 (0.003)	+34.49% (1.10)	0.22 (0.003)	+21.98% (1.41)
Max CTR	0.32 (0.004)	+28.49% (0.98)	0.23 (0.002)	+25.17% (1.07)
First best U	0.42 (0.006)	+70.16% (2.86)	0.27 (0.003)	+51.45% (1.70)
First best CTR	0.33 (0.006)	+34.97% (2.34)	0.29 (0.004)	+61.64% (1.99)

Table 12: Long run counterfactuals

targeting, we increase the CTR by 58%, 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; 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 assumption that position priors are correct and have zero variance. In such a case, users would have no uncertainty about the quality of advertising. Table 12 reports the results. Additionally, this 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 34.5%, 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 51% more clicks in the long run, compared to 42% in the short run (and 58% vs 62% for "Max CTR" policy). Therefore, we conclude that forward-looking companies should be more willing to invest in learning users' tastes. However, we also find significant complementarities when removing uncertainty and conducting user-level targeting together. Companies might therefore be willing to invest in both removing uncertainty and targeting, even if investing in removing uncertainty alone is not profitable.

7 Conclusion

This paper provides empirical evidence of externalities among ads, user heterogeneity and user uncertainty regarding the quality of ads in the sponsored search market. We provide this evidence, using both reduced-form tests and a structural model 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 few percent, with a maximal increase of a few percent of CTR for some advertisers) 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 33%, and the increase could go up to 69% 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 CTR differences between positions, we can evaluate long-run counterfactuals. 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 ads to specific consumers, it would achieve larger CTR gains (62% in long-run versus 58% 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 us to do so, because it does not track users beyond the single search session. This direction is important 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.

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

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

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

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

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.*

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

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.

Yao, S., and C. F. Mela (2011): "A dynamic model of sponsored search advertising," *Marketing Science*, 30(3), 447–468.

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

A Numerical appendix

As mentioned in Section 4, Bellman equation (3.2) has to be solved for every S. There two complications when solving this equation: (i) S is continuous state variable, (ii) integration on RHS of the Bellman equation is burdensome. Note that the Bellman equation has to be solved millions of times in order to estimate the model, thus every extra millisecond of computation of a standalone DP problem counts. For this reason, we employ a series of efficient numerical approximations that jointly interpolate the integral and the state S.

The integration of the right-hand-side of the Bellman equation is performed using Gauss-Hermite quadrature with five nodes. Let y_k be the normalized Gauss-Hermite grid points, then the quadrature for integrating the posterior distribution of $v_{a(c)i}$ has the following nodes

$$\hat{v}_{a(c)i,k} = \hat{\mu}_{a(c)i} + \sqrt{2}\hat{\sigma}_{a(c)i}y_k.$$

Consequently, we replace the Bellman equation (3.2) with its discretized version

$$V_i\left(\mathcal{C}, S\right) = \max\left\{S^{1/(1+R_i)} - \sum_{n\in\mathcal{C}} f_n, \max_{c\in\{1,..,N\}\setminus\mathcal{C}} \sum_{k=1}^{5} w_k V_i\left(\mathcal{C}\cup c, S + \hat{v}_{a(c)i,k}^{1+R_i}\right)\right\},$$

where w_k are Gauss-Hermite weights. Under this discretization it suffices to consider a finite number of values for S for each $\mathcal{C}$, namely, the values of S from the set

$$\left\{\sum_{c\in\mathcal{C}} \hat{v}_{a(c)i,k_c)}^{1+R_i} : k_c \in \{1,\ldots,6\}\right\},$$

where $\hat{v}_{a(c)i,6}$ is equal to the true value of $v_{a(c)i}$. We need this extra point to evaluate policies at the true values of $v_{a(c)i}$ during the forward simulations. The above observations significantly reduce the state space without introducing additional approximation errors beyond the Gauss-Hermite integration. For example, if $\mathcal{C}$ has three clicked ads, there are 6^3 relevant values for S which gives $\binom{8}{3}6^3$ state points. Additionally, we find it is enough to consider bundles of a maximum 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 13,152 state points. [24]

[24] The number of state points is computed as $\binom{8}{3}3^6 + \binom{8}{2}2^6 + \binom{8}{1}1^6 + 1$.