Multi-Score Position Auctions

[Extended Abstract]

Denis Charles
Microsoft
1 Microsoft Way, Redmond, WA, USA
cdx@microsoft.com

Nikhil R. Devanur
Microsoft Research
1 Microsoft Way, Redmond, WA, USA
nikdev@microsoft.com

Balasubramanian Sivan
Microsoft Research
1 Microsoft Way, Redmond, WA, USA
balu2901@gmail.com

ABSTRACT

In this paper we propose a general family of position auctions used in paid search, which we call multi-score position auctions. These auctions contain the GSP auction and the GSP auction with squashing as special cases. We show experimentally that these auctions contain special cases that perform better than the GSP auction with squashing, in terms of revenue, and the number of clicks on ads. In particular, we study in detail the special case that squashes the first slot alone and show that this beats pure squashing (which squashes all slots uniformly). We study the equilibria that arise in this special case to examine both the first order and the second order effect of moving from the squashing-all-slots auction to the squash-only-the-top-slot auction. For studying the second order effect, we simulate auctions using the value-relevance correlated distribution suggested in Lahaie and Pennock [2007]. Since this distribution is derived from a study of value and relevance distributions in Yahoo! we believe the insights derived from this simulation to be valuable. For measuring the first order effect, in addition to the said simulation, we also conduct experiments using auction data from Bing over several weeks that includes a random sample of all auctions.

Categories and Subject Descriptors

F.0 [**Theory of Computation**]: General; J.4 [**Social and Behavioral Sciences**]: Economics

Keywords

Generalized Second Price Auction, Squashing

1. INTRODUCTION

Sponsored search auctions have been the "killer app" for algorithmic game theory, due to the enormity of the scale (with 10s of billions of dollars in annual revenue) and the automated nature of these auctions. On the one hand, the basic design of this auction, referred to as the Generalized Second Price (GSP) auction has been the gold standard and has remained constant in the academic community, while on the other hand several extensions have been incorporated into this basic design in practice. These extensions seek better trade-offs between various objectives and are held as trade secrets. The aim of this paper is to systematically consider a general family of auctions and identify specific instances that perform better than GSP and its variants that are used in practice.

One such widely adopted extension of the GSP auction is *squashing*, and was introduced in the widely cited paper of Lahaie and Pennock [13]. While the usual GSP auction ranks the ads by the product of their bids and click probabilities, squashing allows the click probability (or more generally a *relevance* score, denoted by e) to be raised to some fixed exponent (i.e., the ranking is by $b \times e^{\alpha}$, $\alpha \geq 0$, here b is the bid and e the click probability). This generalizes GSP to a single parameter family of auctions, thus giving more freedom in auction design.

In this paper we propose a general family of auctions, which we call multi-score position auctions (MSPA), and experimentally evaluate a particular sub-family of these auctions, which we term dual score auctions (DSA). The basic idea of an MSPA is that each slot has its own scoring function, along with an initial score to select the unordered set of ads to be shown. The payment rule is in the spirit of GSP: each ad must pay the minimum bid required to retain its slot. This generalizes GSP with squashing (and hence GSP), where the scoring function is the same for all slots, namely, the squashed score. There is a particularly attractive special case of DSA that only applies squashing to the first slot and does not squash the other slots. This special case seems appealing since it does not increase the number of auction parameters over the squashed-GSP auction and seems to be the simplest non-trivial example of a DSA auction. To our knowledge, all previous modifications to GSP consist of a single scoring function for the auction. We study this special case of DSA empirically in detail and also study the nature of its equilibria.

We examine both the first-order and second-order effect of going from the auction where we squash all slots uniformly with exponent α to the one where we squash only the top slot at α. In this context, first-order effect refers to a comparison of the two auctions' revenue and clicks based on existing bids. Second-order effect refers to making the same comparison, but using the respective equilibrium bids for each auction.

WSDM'16, February 22–25, 2016, San Francisco, CA, USA.

DOI: http://dx.doi.org/10.1145/2835776.2835822

Our experimental setup to measure first-order effect.

Our experimental setup uses actual search queries and click responses to compute revenue and clicks, so it captures all the complex dependence of clicks on the entire context of the search results page. This avoids pitfalls of assumptions such as that the click-through rates are a product of a slot dependent factor and an advertisement dependent factor. We achieve this by using certain experimental traffic on Bing for which the order of the ads was randomized. Due to this randomness the order of the ads for some of these search results pages coincides with the order given by any specific DSA auction which we would like to measure. Furthermore, the randomization ensures that this happens sufficiently often and is a representative sample of all search queries. We use data from several weeks so that we have a sufficient volume of instances. Thus, we can evaluate the performance of a DSA auction for any choice of exponents.

Our simulation setup to measure second order effect.

With experiments on actual search engine traffic as described in the previous paragraph, it is very difficult to measure second order effects. This is because a typical A/B test is allotted a small fraction of *search engine traffic*. On the other hand, what we need to measure second order effect is a small set of *advertisers* to respond to the new auction format. This requires changing the auction format for all the queries relevant to all the advertisers in the concerned set, which often demands a large portion of traffic to be segregated for this experiment since each advertiser bids on numerous keywords. Given this difficulty, we measure second order effects using simulations à la Lahaie and Pennock [13]. They generate values and click-probabilities from a particular (joint) distribution which fits well a popular keyword in Yahoo! search. We use the same distribution as theirs. For any auction, given the values, we compute the bids that form an equilibrium and evaluate the auction at these bids to measure second-order effect. (Further, we use the same distribution to measure first-order effect too, for comparison with the experimental results from Bing traffic.)

A random sample of all queries.

This combination of simulations to measure advertiser reaction and real data to only measure first order effects has also been used in other recent papers such as Bachrach et al. [6] and Roberts et al. [17]. However, there is a significant difference in the nature of real data used by them and by us. The real data used by these papers, as well as the data used by Lahaie and Pennock [13] to derive their distributions are for a few popular keywords, keywords that are searched many times. However, it is known that a significant fraction of the revenue comes from the "tail queries", queries that are seen only a few times. Thus any insight obtained from such data may not necessarily be broadly applicable. The way we get our auction data is truly a random sample of all queries over several weeks and is therefore representative of the entire market. We provide more details of our experimental design in Section 3.

Findings from our first-order experiments.

Our main finding regarding first-order-effect based on Bing traffic is that *squashing only the top slot is better than squashing all the slots, both in terms of revenue, and the number of clicks (which is a proxy for user relevance/user satisfaction*, i.e., the DSA auction with exponent α for the top slot and exponent 1 for the remaining slots does better than the squashed GSP auction with exponent α. More details on the exact numbers are in Section 3. Furthermore, the first-order results measured from our experiments with actual Bing traffic are broadly in agreement with the first-order effect measured from simulations using the value and click-probabilities distribution from [13].

Findings from our second-order simulations.

The main task in measuring second order effects is computing the equilibrium of the squash-all-slots auction and squash-only-the-top-slot aucton. For computing the equilibrium of the squash-all-slots auction, we use standard, "lowest revenue fetching envy-free equilibrium" suggested independently by Edelman et al. [9] and Varian [19]. In his paper, Varian [19] gives empirical evidence for advertiser bids coinciding with the predictions of this envy-free equilibrium. For the squash-only-the-top-slot auction, equilibrium computation is more intricate, since there could be multiple equilibria (even after placing the envy-free restriction) and the equilibrium selection has a significant impact. But the overall trend is still positive, with a strong trend that shows that if the equilibrium prefers bidders with higher relevance, then the improvements are significant, and somewhat surprisingly, even more than what we get for the first order! This suggests that the search engine should try to steer the bids towards an equilibrium that favors more relevant advertisers, which is a good idea anyway. The search engine could do this, for instance, via bidding agents that can adjust bids, (under some constraints provided by the advertiser) or by influencing the matching of candidates to auctions. Equilibrium derivation is described in Section 4 and the results from simulations using these equilibria are described in Section 5.

Intuitive explanation.

Both the first-order and second-order effects point towards squash-only-the-top-slot auction being superior to squash-all-slots auction. While this fact acts as an empirical justification for using the former, here we provide an intuitive explanation for this phenomenon. The effect of squashing, which changes the rank-score from $b \cdot e$ to $b \cdot e^{\alpha}$ for $0 \leq \alpha \leq 1$, is to undermine the contribution of click probability in deciding the ordering of ads. (To see this, consider the extreme case of $\alpha = 0$, where the bids alone decide the ordering of ads.) When does squashing have a positive effect on revenue? Consider an arbitrary slot, and 2 competing ads for that slot. Fixing the ads in the other slots, let e_1, e_2 be the click probabilities of ads 1 and 2 for this slot under consideration; let b_1 and b_2 be the bids of these two ads. There are three natural rankings possible here: GSP ranks these ads according to their rank score of $b \cdot e$, squashed-GSP ranks these ads according to their rank-score $b \cdot e^{\alpha}$, and finally there is also the natural ranking of ads according to relevance (click probabilities). We analyze three different cases that come out of how these three different rankings compare, and argue that *only in one case can squashing improve revenue. Further, we argue that this case is likely to occur only for the first slot.* As a convention, let 1 be the ad that wins the slot under consideration when using GSP ranking, i.e., let $b_1 e_1 \geq b_2 e_2$. Note that GSP's revenue from this slot is $\frac{b_2 e_2}{e_1} e_1 = b_2 p_2$. Also, recall that $0 \leq \alpha \leq 1$.

1. *GSP order = Squashed-GSP order = Click-probability order.* Mathematically, this case translates to $\{b_1e_1 \geq b_2e_2;\ b_1e_1^\alpha \geq b_2e_2^\alpha; e_1 \geq e_2\}$. This is the only case where squashing improves revenue. GSP's revenue is b_2e_2 from this slot, while squashed-GSP's revenue is $\frac{b_2e_2^\alpha}{e_1^\alpha}e_1 = b_2e_2^\alpha e_1^{1-\alpha}$. The latter is at least b_2e_2 whenever $e_1 \geq e_2$ which is true in this case.

2. *GSP order = Squashed-GSP order $\neq$ Click-probability order.* Mathematically, this is $\{b_1e_1 \geq b_2e_2;\ b_1e_1^\alpha \geq b_2e_2^\alpha; e_1 \leq e_2\}$. In this case, squashed GSP's revenue is $\frac{b_2e_2^\alpha}{e_1^\alpha}e_1 = b_2e_2^\alpha e_1^{1-\alpha}$. The latter is at most b_2e_2 because we have $e_1 \leq e_2$ in this case. Thus squashing hurts revenue.

3. *GSP order = Click-probability order $\neq$ Squashed-GSP order.* Mathematically, this is $\{b_1e_1 \geq b_2e_2;\ b_1e_1^\alpha \leq b_2e_2^\alpha; e_1 \geq e_2\}$. In this case, squashed GSP's revenue is $\frac{b_1e_1^\alpha}{e_2^\alpha}e_2 = b_1e_1^\alpha e_2^{1-\alpha}$. The latter is at most b_2e_2 because we have $b_1e_1^\alpha \leq b_2e_2^\alpha$. Thus in this case squashing hurts revenue. Further, squashing also hurts the number of clicks, because in GSP the number of clicks is proportional to e_1, whrere as in squashed-GSP, due to a change in the ordering, the number of clicks is proportional to e_2, which is at most e_1.

The question now is, given that squashing hurts in 2 out of 3 cases, whether one should squash or not. Given that the above analysis is on a per-slot basis, a more refined question would be to identify which slots are likely to fall in case 1 and therefore should be squashed (as squashing increases revenue in case 1), and which slots are likely to fall under cases 2, 3 and therefore should not be squashed (as squashing decreases revenue in cases 2, 3, and sometimes also decreases clicks). We claim that slot 1 is more likely to fall in case 1 than any other slot. The reason is that often the GSP winner of the first slot wins not only in terms of $b \times e$ but also in the individual dimensions of b and e. In this case, it is immediate that we fall in case 1. Even if the first slot winner wins only in terms of e, but not in b, often the margin of victory in e is high enough that even if one alters the GSP rank-score a bit by introducing a squashing exponent, it is not very likely to alter GSP's order, and therefore GSP order = Squashed-GSP order = Click-probability order, putting us in case 1 where one gets higher revenue. But as one goes down to larger slot numbers (lower positions), the GSP winner is not so pronounced as the first slot winner: i.e., either the GSP winner of that slot doesn't always win in terms of click-probability e (putting us in case 2), or, the margin of victory in terms of click-probability e is not high enough that a tiny change in rank-score by squashing changes the squahsed-GSP order from GSP order puttting us in case 3. Thus the likelihood of being in case 1 decreases as we increase the slot numbers. Intuitively this explains why it is better not to squash the lower slots, and squash only the top slot.

Another implication of our results is that the revenue benefits of squashing accrue mostly from the top slot.

Generality of the results.

A natural question at this point is to assess the generality of our findings. For instance, will squash-only-the-top-slot be found superior in other search engines' data too? What are some properties in the dataset to look for to get a sense of whether this result will hold? The 3-case analysis in the previous paragraphs shows that for any dataset where the first slot winner wins "more comfortably" (i.e., the victory margin in terms of click-probability is sufficiently high) while the click-probability victory margin is not that large for the remaining slots, squash-only-the-top-slot auction is very likely to do better. This property is true in Bing data set, even after one excludes navigational queries and other "dominant advertiser queries" like Verizon bidding on "Verizon" keyword (although we are not allowed to release specific numbers like the bid decay rate, click probability decay rate, etc.). We believe that while the ratios may be different for different search engines, this general fact that the first slot winner wins "more" than the other slot winners should hold in most datasets.

Related Work: Apart from the foundational work of Aggarwal et al. [3], Edelman et al. [9], Varian [19], the work that is most closely related to ours is that of Lahaie and Pennock [13]. They propose a variant of the GSP auction, namely squashed GSP auctions, where the ranking uses a squashed score and contains as special cases both the rank-by-bid and usual GSP. They show that GSP with squashing allows the auctioneer to optimize revenue at equilibrium and justify their results with simulations. In later work Lahaie and McAfee [12] show that the squashed auction can, under certain circumstances, also produce an efficient ranking. Our work presents a further generalization of the auction mechanisms of these works. Our simulations closely match the ones done in Lahaie and Pennock [13]. We use counterfactual analysis to study the first-order effects and to analyze the revenue-relevance trade-offs for the auction on actual ad auction data. This methodology is expounded in Bottou et al. [7], however, in a machine learning rather than an auction design setting.

Other papers have proposed modifications to the basic mechanism. Aggarwal et al. [4] propose a truthful auction based on Gale-Shapley stable matching theory that can incorporate bidder and position specific minimum and maximum prices. Roberts et al. [17] propose an auction that ranks using the difference between the bid and the reserve price. In other words, the reserve price affects the ranking beyond just filtering out advertisers with low bids. They too use simulations and real data to show this could raise more revenue. Another example of an empirical evaluation of variants of the GSP auction is by Thompson and Leyton-Brown [18]. They too run simulations, similar to Lahaie and Pennock [13], using data generated from distributions believed to mimic the actual bid distribution. They compare the revenue guarantee at equilibrium with different quality-weighted reserve prices and un-weighted reserve prices, and conclude that un-weighted prices perform consistently better. They also show that squashing improves the revenue of quality-weighted reserve prices, which otherwise fall much behind un-weighted reserves. Bachrach et al. [6] also evaluate different ways of setting reserve prices, with an emphasis on the tradeoffs obtainable between different objectives, once again using a combination of simulations and real data.

GSP Auctions have been studied from many different angles [5, 1, 14, 16, 10], such as questioning the separability assumption about the click-through rates, which are observed to not hold in practice [8, 2, 11]. Ostrovsky and Schwarz [15] analyze how using reserve prices derived from Myerson's theory can help increase revenue. A more complete summary of these results is beyond the scope of this paper. However, we are not aware of any other papers suggesting modifications of the basic auction format itself.

2. MODEL

2.1 Position Auctions

Position Auctions refer to auctioning of ad slots alongside "organic search" results in response to a user query in a search engine. Abstractly, a set of ads, A, compete for a set of k slots numbered 1 through k, with $k < |A|$. The slots at the top of the page (which correspond to lower numbered slots in our notation) are more desirable to all the advertisers. The slots are allocated through an auction. The advertisers place bids that are used in the auction.

An important feature of this mechanism is that it is pay-per-click: while the search engine assigns slots, it gets paid only if the user clicks on the ad. Whether an ad gets clicked or not depends on the allocation of ads to slots, the most general view being that the click probability of any ad depends on the entire slate of ads being shown. A common assumption is the, so-called, rank-1 assumption: that the click probability is a product of an advertiser factor and the slot-factor. Opinions about the accuracy of this assumption are mostly divided. In any case, an important consideration in the auction is the inherent clickability of an ad, which we simply call as the click probability. This is estimated by the search engine/auctioneer using sophisticated machine learning algorithms. For the purpose of auction design these can be thought of as the input.

To complete the abstract auction design problem, the auction assigns each slot to an ad, and charges payments. We represent the assignment by $\sigma : [k] \to A$, $\sigma(j)$ is the identity of the ad that is assigned slot j.[1] Let v_j be the value of the ad in slot j for getting a click and let π_j be the payment made by the ad in slot j, on getting a click. Let $c(j, \sigma)$ be the probability that the ad in slot j is clicked, given the allocation σ.

Important objectives in these auctions are

1. Revenue: the expected revenue of an allocation σ and payment π is

$$\sum_{j=1}^{k} \pi_j c(j, \sigma).$$

   Revenue is clearly of interest to the search engine.

2. Number of clicks: The expected number of clicks given an allocation σ is

$$\sum_{j=1}^{k} c(j, \sigma).$$

   The number of clicks is an indication of ad relevance and engagement with the user, and is considered very important. More relevant ads and better engagement with the user bring back more users and, in the long run, benefit the search engine and the advertisers.

[1] In practice k is not fixed, but this is a minor and distracting feature that we ignore for the sake of discussion here.

2.2 Multi-score auctions

A multi-score auction is given by a set of $k+1$ functions, s_0 and s_j for $j = 1, \cdots, k$, from a pair of bid and click probability, (b, c), to a non-negative real number. Each s_j for $j = 0, \cdots, k$ is strictly monotonically increasing in each of its arguments. Let s_j^{-1} denote the inverse of s_j in the first argument, i.e., $s_j^{-1}(b', p)$ is the unique number b such that $s_j(b, p) = b'$.

The auction proceeds as follows, it first selects the top k ads according to s_0. Then it goes down the slots from 1 to k and for each $j = 1, \cdots, k$ picks the ad with the highest score according to s_j (from the ads not already picked for a higher slot) to be assigned slot j. The payments are set similar to the GSP rule: each advertiser pays the lowest amount he would have to bid in order to retain his slot. The auction is summarized in Algorithm 1.

All of our auctions can be extended to include reserve prices, but we ignore reserve prices to keep the exposition simple.

ALGORITHM 1: Multi-Score Position Auctions

Input: A set of candidate ads, A, and for each ad $i \in A$, its bid b_i and its click probability c_i.
Output: A selection and ranking of top k ads given by $\sigma : [k] \to A$. Their payments given by π_j, for $j = 1, \cdots, k$.
$S \leftarrow \arg k \max\{s_0(b_i, c_i) : i \in A\}$, the top k ads from A, ranked by $s_0(b_i, c_i), i \in A$;
$r_0 \leftarrow \arg\max\{s_0(b_i, c_i) : i \in A \setminus S\}$;
for $j = 1, \cdots, k$ **do**
    $\sigma(j) \leftarrow \arg\max\{s_j(b_i, c_i) : i \in S\}$;
    $S \leftarrow S \setminus \sigma(j)$;
    $\pi_j \leftarrow$
    $\max\left(\{s_j^{-1}(s_j(b_i, c_i), c_{\sigma(j)}) : i \in S\} \cup \{s_0^{-1}(r_0, c_{\sigma(j)})\}\right)$;
end

The multi-score auction is a generalization of standard auctions used in practice: the GSP auction corresponds to the scoring functions $s_j = bc$ for all j. Another common variant is GSP with *squashing*, which corresponds to $s_j = bc^\alpha$ for some α (we restrict to $\alpha \in [0, 1]$). Variants considered previously fall into the "single score" framework, where s_j is the same for all j. Extending it to possibly different scores for each slot gives more freedom to the auction designer.

2.3 Dual score auctions (DSA)

We consider a special case of the above general family of auctions. The class is parameterized by a pair of real numbers, (α, β). In the first step, the top k ads are picked according to bc^β. From among them, the ad in the first slot is picked according to bc^α, and the rest of the slots are filled with the remaining ads ranked by bc^β. In other words, these are multi-score auctions with $s_1(b, c) = bc^\alpha$ and $s_j(b, c) = bc^\beta$ for all $j \neq 1$.

We rename the ads so that σ is the identity function. Then advertiser 1 is charged

$$\max\left(\left\{\frac{b_j c_j^\alpha}{c_1^\alpha} : j \geq 2\right\} \cup \left\{\frac{b_j c_j^\beta}{c_1^\beta} : j > k\right\}\right).$$

Advertiser j for $j = 2, \cdots, k$ is charged $b_{j+1} c_{j+1}^\beta / c_j^\beta$. We refer to this auction as $\mathrm{DSA}(\alpha, \beta)$.

There are also other ways to get sub-families of MSPAs with few parameters. For instance, one can consider a separate scoring function for candidate *selection* (s_0) and a different one for *ranking* (s_j for all $j \geq 1$). The space of MSPAs is quite huge and it is possible that there several other interesting auctions in this framework.

3. EXPERIMENTS

We run experiments on actual data from Bing. These experiments measure first order effects, which measures the change in the performance metrics keeping the bids unchanged. It is almost impossible to effectively measure second order effects from real data. The reason is that such experiments are required to run only on a small sample of auctions. However, in order to measure the second order effects, all of the auctions an advertiser participates in must be included in the experiment. This creates a dependency which often forces to include a large part, if not all, of the auctions, contradicting the requirement that the experiment be only performed on a small fraction of auctions.

3.1 Experimental Design

For our empirical evaluation, we use data from existing randomized experiments in Bing. One of the existing experiments in Bing places the ads in an order that is chosen *uniformly at random from all permutations* from the set of auction candidates. This experiment itself runs on a random sample of actual search traffic. Given any value of (α, β), there is a non-trivial fraction of auctions for which the (random) order of the ads in this experiment matches the order of ads according to the auction $\textsc{dsa}(\alpha, \beta)$. This is a random sample of all auctions (since the randomization of the ad slate is independent of the auction) and given sufficiently many auctions, we get a representative sample of all search traffic. We use the bid data and the *actual user click reponses* from logged information for this experiment. We used logs from a duration of several weeks for the analysis. This lets us compare DSAs with different α, β parameters for a number of variants. As noted earlier, this is a significant departure from earlier papers which use data from a few popular keywords only. Moreover, our method allows us to evaluate the auctions with respect to actual clicks, which is also absent in previous work. Here are some key facts about our experimental design:

1. The data for the experiment was collected for a duration of 41 days.

2. The experiment performs a filtering that excludes navigational intent queries and other queries where there is a dominant advertiser, like Verizon bidding on the "Verizon" keyword.

3. Barring the dominant advertiser filter, and other sanitizations to remove nonbillable queries, etc., the experimental traffic represents a true random sample of the entire traffic to Bing.

4. For each value of α and β we report, the sample size, which is the number of samples in our data set that matched the order of $\textsc{dsa}(\alpha, \beta)$, was between 11 million and 13 million. Thus, there are no small sample issues.

Even though the order of the ads is randomized, it may not be perfect. A sanity check is to see whether for a given (α, β) there were sufficiently many queries for which the ads were ranked as per $\textsc{dsa}(\alpha, \beta)$. Due to confidentiality reasons, we present a slightly different statistic, for each β we present the average (over different αs) of the ratio of the number of queries matching $\textsc{dsa}(\alpha, \beta)$ to the number of queries matching $\textsc{dsa}(1, 1)$, i.e.,

$$\sum_{\alpha \in \Lambda} \frac{1}{|\Lambda|} \frac{\text{number of queries matching } \textsc{dsa}(\alpha, \beta)}{\text{number of queries matching } \textsc{dsa}(1, 1)}. \qquad (1)$$

One caveat of our setup is that we cannot change the criteria used to select the *set* of ads, alternatively s_0 is the criterion that Bing used for selecting auction candidates during the time period of our data collection. We only capture effects that are due to re-ranking of candidate ads.

3.2 Experimental Results

The main chart is in Figure 1, which compares $\textsc{dsa}(\alpha, 1)$ with $\textsc{dsa}(\alpha, \alpha)$. This shows that $\textsc{dsa}(\alpha, 1)$ always pareto-dominates $\textsc{dsa}(\alpha, \alpha)$, with substantial improvements in *both* RPS and CTR. This is in agreeement with the simulation results presented in Figure 3. This is the most interesting conclusion of these experiments, indicating that squashing only the top slot is better than complete squashing.

Why restrict α to $[0, 1]$?.

The optimal value of α is keyword specific, and depends on the correlation between bids and click-through-rates for that keyword. While theoretically the optimal α could lie anywhere in $(-\infty, +\infty)$, the kinds of correlation necessary to make the optimal α negative or too large are not realistic. Further, even if a negative α could afford the search engine large revenue in the short run, it results in very poor relevance, affecting user satisfaction. Consequently, we present results for $0 \leq \alpha \leq 1$ which seems to be the range of α's that give the right trade-off between revenue and relevance. Remarkably, Figure 1 shows that squashing only the top slot is better than squashing for *all* slots for all αs in $[0, 1]$. While we omit results for $\alpha \in [-1, 0]$, the trend of simultaneous increase in revenue and CTR continues to remain there too.

Next, we present a chart (Figure 2) showing that the number of queries is pretty evenly distributed, by giving for each β, the ratio of the number of queries matching $\textsc{dsa}(\alpha, \beta)$ to the number of queries matching $\textsc{dsa}(1, 1)$ as mentioned in Equation (1). Note that the value for $\beta = 1$ is less than 1, since this is an average over all αs.

4. EQUILIBRIA OF DSA

Unlike our experiments, where $c(j, \sigma)$ could be an arbitrary function of j and σ, measuring second-order effects of DSA requires computing DSA's equilibrium and hence some structure is necessary on $c(j, \sigma)$. We make the standard assumption that the ckick-through-rate of bidder s in position t is $e_s x_t$, for $e_s, x_t \in [0, 1]$. In our previous section's notation, $c(j, \sigma) = e_{\sigma(j)} x_j$. The quantity e_s is often referred to as the relevance of bidder s's ad and x_t is called as the position effect.

With the above separable click-through-rate assumption Varian [19] and Edelman et al. [9] characterize the set of envy-free equlibria or "symmetric Nash equilibria" (SNE) of the GSP auction and Lahaie and Pennock [13] do the same for GSP with squashing. These are a subset of pure Nash

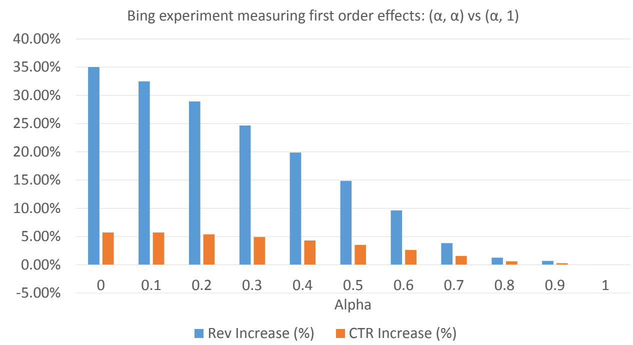

Figure 1: Impact on revenue and click-throughs for DSA(α, α) **vs.** DSA$(\alpha, 1)$**.**

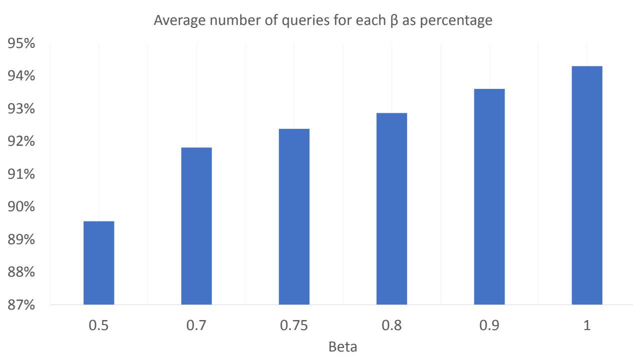

Figure 2: Number of queries as a function of β.

equilibria, obtained by strengthening some of the Nash equilibrium constraints to make them symmetric for deviations that take the bidder above and below his current slot. The characterization says that such an equilibrium always exists, there is only one order of bidders that can result in an SNE and gives a closed form formula for the bids at the lowest price equilibrium (which is uniquely defined).

Unfortunately, there isn't such a simple characterization for the SNE of DSA, and a SNE may not exist. When they exist, there may be multiple orderings of bidders which support equilibria. The first fact is not problematic: our simulations show that equilibrium exists for more than 91% of queries[2] for all $\alpha \geq 0$. The second poses an interesting conundrum, and our simulations show that the choice of equilibrium order results in significant difference to the performance metrics. Before we detail the results of our simulations, we detail the equilibrium constraints. Fix a particular ordering of bidders, say bidder s is in slot s for $s \in [n]$. Then the SNE conditions are as follows. For DSA(α, β), let $w_s = e_s^{\beta}$ and $\widehat{w}_s = e_s^{\alpha}$. Let b_s be the bid of advertiser s.

We first write envy-free constraints for bidders in slots $s > 1$ to not envy each other:

[2]Note that as α gets closer to 1, the percentage of queries that have a SNE increases and hits 100% at $\alpha = 1$ since that corresponds to pure GSP.

$$\forall s \neq 1, t \neq 1, \quad \left(v_s - \frac{b_{s+1} w_{s+1}}{w_s}\right) e_s x_s \geq \left(v_s - \frac{b_{t+1} w_{t+1}}{w_s}\right) e_s x_t \tag{2}$$

Let $\widehat{p}_1 = \max_{s \geq 2} b_s \widehat{w}_s$. We now write constraints for bidders in slots $s > 1$ to not envy bidder in slot 1.

$$\forall s \neq 1, \qquad \left(v_s - \frac{b_{s+1} w_{s+1}}{w_s}\right) e_s x_s \geq \left(v_s - \frac{\widehat{p}_1}{w_s}\right) e_s x_1 \tag{3}$$

Finally, we write the constraints for bidder in slot 1 not to envy any bidder in slot $s \geq 2$.

$$\forall s \neq 1, \qquad \left(v_1 - \frac{\widehat{p}_1}{\widehat{w}_1}\right) e_1 x_1 \geq \left(v_1 - \frac{b_{s+1} w_{s+1}}{w_1}\right) e_1 x_s \tag{4}$$

Given an ordering of the bidders, we can check if equilibrium exists using conditions (2),(3),(4). Once we know equilibrium exists for a given order, we can also find the *lowest* price equilibrium bids as follows. While all kinds of complicated allocations and bids could satisfy conditions (2),(3),(4), note that once the bidder in the first slot is fixed, the remaining $k-1$ slots should necessarily satisfy mutual envy-free conditions among themselves, and therefore, by Edelman et al. [9] and Varian [19], there is a unique ordering among them, namely, the order of $v_s w_s = v_s e_s^{\beta}$. The lowest SNE bids for slots $3, \ldots, k$ is uniquely determined, and is given by (see Varian [19] and Lahaie and Pennock [13]):

$$b_{s+1} w_{s+1} x_s = \sum_{t=s}^{k} (x_t - x_{t+1}) v_{t+1} w_{t+1}. \tag{5}$$

Fixing the above bids for slots $3, \ldots, k$ (bidders $s > k$ have $b_s = v_s$) determines the payments for all slots except 1 . The only quantity to be determined to compute revenue is $\widehat{p}_1$. We compute the smallest $\widehat{p}_1$ that satisfies conditions (3),(4). This determines the revenue to be $\frac{\widehat{p}_1}{\widehat{w}_1} x_1 e_1 + \sum_{s=2}^{k} \frac{b_{s+1} w_{s+1}}{w_s} x_s e_s$.

Remark 1. *It could be possible that an equilibrium for DSA(α, β) may exist satisfying conditions (2),(3),(4), but may not satisfy the lowest bids specified in (5). For our simulation purposes, we call such instances as no-equilibrium instances, and even with this restriction, the number of instances with equilbrium is at least 91% in the relevant regime of $\alpha \geq 0.4$. More precisely, we only look for equilibria that satisfy the lowest bid constraints for slots $3, \ldots, k$ in (5).*

Remark 2. *Note that since $\widehat{p}_1 = \max_{s \geq 2} b_s \widehat{w}_s$, bidder 2 is not necessarily the price setter for bidder 1.*

5. SIMULATIONS

We do simulations to compare the performance of DSA and GSP with squashing. The goal of doing simulations is to measure the second order effects, i.e., if we change the auction format, then we expect the advertisers to adjust their bids to the new format, with respect to which we should measure the performance metrics. How do advertisers bid? The standard assumption is that they bid at an SNE. Therefore we measure the percentage change in the performance metrics betweeen an SNE of GSP with squashing and an SNE of DSA. This is the same methodology used by Lahaie and Pennock [13] to compare GSP with squashing with pure GSP. For our simulations we use the same distribution as they did to generate the instances.

As in Lahaie and Pennock [13] we use a joint distribution for determining the bidders' value relevance pairs. The marginal distribution for values is a lognormal distribution with parameters $\mu = 0.35$ and $\sigma = 0.71$. The marginal distribution for relevance is a beta distribution with parameters $a = 2.71$ and $b = 25.43$. We use a Gaussian copula to create a joint distribution from these two marginal distributions, with differing levels of correlation. We report results for Spearman correlation levels that Lahaie and Pennock report fit well with the keyword for which they performed their simulations: namely, correlations of $0.3, 0.4, 0.5$.

We focus on the two most important objectives, revenue and click-throughs. We present relative impact on these two objectives as we move from $\textsc{dsa}(\alpha, \alpha)$ to $\textsc{dsa}(\alpha, 1)$. The percentage *increase* in revenue and the percent *increase* in clickthroughs are respectively

$$\left(\frac{\textsc{rev}(\alpha,1)}{\textsc{rev}(\alpha,\alpha)} - 1\right) \cdot 100 \text{ and } \left(\frac{\textsc{ctr}(\alpha,1)}{\textsc{ctr}(\alpha,\alpha)} - 1\right) \cdot 100.$$

Before we do the second order analysis, we do a first order analysis: how do the performance metrics change if we retain the SNE bids for the original auction. We do this as a sanity check to test whether this agrees with the first order analysis we perform on real data (see Section 3). These results are presented in Figure 3. The first order effects show a consistent improvement in both RPS and CTR for $\textsc{dsa}(\alpha, 1)$ vs. $\textsc{dsa}(\alpha, \alpha)$, in agreement with our analysis on real data. We emphasize that while simulation is for a single keyword's distribution, our real experiment is over a random selection of all queries for several weeks.

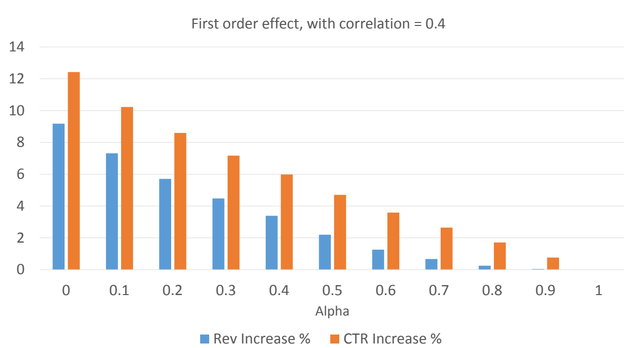

Figure 3: Impact on revenue and click-throughs for $\textsc{dsa}(\alpha, \alpha)$ vs. $\textsc{dsa}(\alpha, 1)$.

As mentioned before, there may not be an SNE at all in $\textsc{dsa}(\alpha, \beta)$ or there may be many orders of bidders for which equilibria exist. Observe that the possible number of such orders is at most n, since once you fix the bidder in the top slot, the rest of the bidders are ordered according to $b_j e_j^\beta$. Therefore it is computationally easy to just try all n choices for the first slot and check for equilibrium. This gives rise to the problem of equilibrium selection; if many of these choices lead to an equilibrium, which one do we choose?[3] We break ties according to the following 3 criteria, the relevance e_j, the α-score for true values $v_j e_j^\alpha$ and the β-score for true values $v_j e_j^\beta$. These orderings give successively lesser importance to relevance (the e_j's) and our simulations show that they perform successively worse.

The results of the second-order analysis are presented in Figure 4 for samples with a Spearman correlation of 0.4.

We also do simulations for samples with a Spearman correlation of 0.3 and 0.5. The results are in Tables 1, 2, 3 and 4, and the trends are consistent with what we get for correlation of 0.4.

The simulations suggest that the performance metrics depend significantly on the choice of equilibrium. As can be seen from Figure 4, there is a strong trend that shows that preferring advertisers with higher relevance can lead to a significant increase in all the desired peformance metrics. Perhaps surprisingly, when we simply use the relevance score (e_j) to break ties, this increase is even more than the first order increase. The search engine could steer the bids towards an equilibrium that prefers advertisers with higher relevance via, for example, bidding agents. Our simulations show that such a strategy could result in a significant improvement in the performance metrics.

References

[1] Abrams, Z. and Ghosh, A. 2007. Auctions with revenue guarantees for sponsored search. In *Internet and Network Economics, Third International Workshop, WINE 2007, San Diego, CA, USA, December 12-14, 2007, Proceedings*. 143–154.

[2] Aggarwal, G., Feldman, J., Muthukrishnan, S., and Pál, M. 2008. Sponsored search auctions with markovian users. In *Internet and Network Economics, 4th International Workshop, WINE 2008, Shanghai, China, December 17-20, 2008. Proceedings*. 621–628.

[3] Aggarwal, G., Goel, A., and Motwani, R. 2006. Truthful auctions for pricing search keywords. In *Proceedings 7th ACM Conference on Electronic Commerce (EC-2006), Ann Arbor, Michigan, USA, June 11-15, 2006*. 1–7.

[4] Aggarwal, G., Muthukrishnan, S., Pál, D., and Pál, M. 2009. General auction mechanism for search advertising. In *Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009*. 241–250.

[5] Athey, S. and Nekipelov, D. 2012. A structural model of sponsored search advertising auctions. Working paper.

[6] Bachrach, Y., Ceppi, S., Kash, I. A., Key, P., and Kurokawa, D. 2014. Optimising trade-offs among stakeholders in ad auctions. In *ACM Conference on Economics and Computation, EC '14, Stanford , CA, USA, June 8-12, 2014*. 75–92.

[7] Bottou, L., Peters, J., Quiñonero-Candela, J., Charles, D. X., Chickering, D. M., Portugaly, E.,

[3]Note that while there are multiple possible equilibria, all of these equilibria converge to the lowest revenue SNE when $\alpha = \beta$.

Figure 4: Impact on revenue and click-throughs for DSA(α, α) **vs.** DSA$(\alpha, 1)$ **for correlation** $= 0.4$**.**

α	0	0.2	0.4	0.6	0.8	1
First-order ΔRev %	7.43	4.55	2.36	0.97	0.17	0
Second-order (Relevance) ΔRev %	28.06	21.63	14.36	6.62	1.19	0
Second-order ($v_j e_j$) ΔRev %	13.43	8.43	3.03	-0.36	-1.16	0
Second-order ($v_j e_j^\alpha$) ΔRev %	-9.23	-6.58	-4.45	-3.31	-1.9	0

Table 1: Impact on revenue for $\textsc{dsa}(\alpha, \alpha)$ vs. $\textsc{dsa}(\alpha, 1)$ with a correlation of 0.5.

α	0	0.2	0.4	0.6	0.8	1
First-order ΔCTR %	9.82	6.68	4.48	2.70	1.29	0
Second-order (Relevance) ΔCTR %	16.34	12.38	8.73	5.74	3.14	0
Second-order ($v_j \cdot e_j$) ΔCTR %	12.94	9.25	5.73	3.37	1.71	0
Second-order ($v_j \cdot e_j^\alpha$) ΔCTR %	1.61	0.79	0.58	0.68	0.69	0

Table 2: Impact on click-throughs for $\textsc{dsa}(\alpha, \alpha)$ vs. $\textsc{dsa}(\alpha, 1)$ with a correlation of 0.5.

α	0	0.2	0.4	0.6	0.8	1
First-order ΔRev %	10.98	7.11	3.95	1.71	0.35	0
Second-order (Relevance) ΔRev %	38.78	29.17	19.71	9.39	2.21	0
Second-order ($v_j \cdot e_j$) ΔRev %	21.43	13.91	7.56	1.88	-0.69	0
Second-order ($v_j \cdot e_j^\alpha$) ΔRev %	-8.81	-5.15	-3.01	-2.19	-1.6	0

Table 3: Impact on revenue for $\textsc{dsa}(\alpha, \alpha)$ vs. $\textsc{dsa}(\alpha, 1)$ with a correlation of 0.3.

α	0	0.2	0.4	0.6	0.8	1
First-order ΔCTR %	15.01	10.59	7.16	4.49	2.06	0
Second-order (Relevance) ΔCTR %	22.42	16.8	11.77	7.55	4.1	0
Second-order ($v_j \cdot e_j$) ΔCTR %	17.79	12.4	7.9	4.6	2.3	0
Second-order ($v_j \cdot e_j^\alpha$) ΔCTR %	2.85	1.41	0.37	0.67	0.83	0

Table 4: Impact on click-throughs for $\textsc{dsa}(\alpha, \alpha)$ vs. $\textsc{dsa}(\alpha, 1)$ with a correlation of 0.3.

Ray, D., Simard, P., and Snelson, E. 2013. Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research 14*, 3207–3260.

[8] Cavallo, R. and Wilkens, C. A. 2014. GSP with general independent click-through-rates. In *Web and Internet Economics - 10th International Conference, WINE 2014, Beijing, China, December 14-17, 2014. Proceedings.* 400–416.

[9] Edelman, B., Ostrovsky, M., and Schwarz, M. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review 97,* 1, pp. 242–259.

[10] Gomes, R. D. and Sweeney, K. S. 2009. Bayes-nash equilibria of the generalized second price auction. In *Proceedings of the 10th ACM Conference on Electronic Commerce*. EC '09. ACM, New York, NY, USA, 107–108.

[11] Kempe, D. and Mahdian, M. 2008. A cascade model for externalities in sponsored search. In *Internet and Network Economics, 4th International Workshop, WINE 2008, Shanghai, China, December 17-20, 2008. Proceedings*. 585–596.

[12] Lahaie, S. and McAfee, R. P. 2011. Efficient ranking in sponsored search. In *WINE*. 254–265.

[13] Lahaie, S. and Pennock, D. 2007. Revenue analysis of a family of ranking rules for keyword auctions. In *ACM Conference on Electronic Commerce*. 13–16.

[14] Likhodedov, A. and Sandholm, T. 2003. Auction mechanism for optimally trading off revenue and efficiency. In *Proceedings of the 4th ACM Conference on Electronic Commerce*. EC '03. ACM, New York, NY, USA, 212–213.

[15] Ostrovsky, M. and Schwarz, M. 2011. Reserve prices in internet advertising auctions: A field experiment. In *Proceedings of the 12th ACM conference on Electronic commerce*. ACM, 59–60.

[16] Pin, F. and Key, P. 2011. Stochastic variability in sponsored search auctions: observations and models. In *Proceedings 12th ACM Conference on Electronic Commerce (EC-2011), San Jose, CA, USA, June 5-9, 2011*. 61–70.

[17] Roberts, B., Gunawardena, D., Kash, I. A., and Key, P. 2013. Ranking and tradeoffs in sponsored search auctions. In *ACM Conference on Electronic Commerce, EC '13, Philadelphia, PA, USA, June 16-20, 2013*. 751–766.

[18] Thompson, D. R. and Leyton-Brown, K. 2013. Revenue optimization in the generalized second-price auction. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*. EC '13. ACM, New York, NY, USA, 837–852.

[19] Varian, H. 2006. Position auctions. *International Journal of Industrial Organization 26,* 6, 1163–1178.