

Optimal Bidding in Multi-Item Multi-Slot Sponsored Search Auctions

Vibhanshu Abhishek and Kartik Hosanagar

The Wharton School, University of Pennsylvania, Philadelphia, PA-19104, USA
{vabhi, kartikh}@wharton.upenn.edu

DRAFT: PLEASE DO NOT SHARE

Abstract

We study optimal bidding strategies for advertisers in sponsored search auctions. In general, these auctions are run as variants of second-price auctions but have been shown to be incentive incompatible. Thus, advertisers have to be strategic about bidding. Uncertainty in the decision-making environment, budget constraints and the presence of a large portfolio of keywords makes the bid optimization problem non-trivial. We present an analytical model to compute the optimal bids for keywords in an advertiser’s portfolio. To illustrate the model, we calibrate the parameters to data from an advertiser’s sponsored search campaign. The results of a field implementation show that the proposed bidding technique is very effective in practice. We extend our model to account for interaction between keywords, which is measured by the spillovers from generic keywords into branded keywords. The spillovers are estimated using a dynamic linear model framework. Subsequently, we use the estimates of the spillovers to jointly optimize the bids of the keywords using an approximate dynamic programming approach.

Keywords: Sponsored search, search engine marketing, bid optimization, stochastic optimization, stochastic modeling

Subject Specification: Marketing: Advertising and media, Estimation/statistical techniques. Games/group decisions: Bidding/auctions, Stochastic. Information systems: Decision support systems. Industries: Computer/electronic.

Area of Review: Marketing

Acknowledgement: *The authors thanks Vadim Cherepanov for valuable assistance with prior ver-*

sions of this work and Peter Fader for several helpful comments and suggestions. The paper is an extended version of conference papers from the ACM Conference on Electronic Commerce (EC'08) and the Conference on Information Systems and Technology (CIST'08), and we thank their anonymous reviewers for several helpful suggestions.

1 Introduction

With the growing popularity of search engines among consumers, advertising on search engines has also grown considerably. Search engine advertising or sponsored search has several unique characteristics relative to traditional advertising and other forms of online advertising. Compared to traditional advertising in print/television, sponsored search is highly measurable allowing advertisers to identify which keywords are generating clicks and which clicks are getting converted to purchases. Compared to other forms of online advertising such as banner ads, search advertising enjoys much higher click-through and conversion rates. Search queries entered by users convey significant information about the users' current need and context and allow search engines to better target ads to the users than is possible in other forms of online advertising. Further, search engine users, unlike users on another website, primarily use the search engine to reach some other website. Advertising is an effective way to enable that process.

Search engines commonly use Pay Per Click (PPC) auctions to sell their available inventory of ad positions for any search query. The auction mechanism is referred to as the Generalized Second Price (GSP) auction. In these auctions, advertisers select keywords of interest, create brief text ads for the keywords and submit a bid for each keyword which indicates their willingness to pay for every click. For example, a travel agent may submit the following set of two tuples $\{(miami\ vacation, \$10), (hawaii\ vacation, \$15), (best\ travel\ deals, \$5), \dots\}$ where the first element in any two-tuple is the keyword and the second element is the advertiser's bid. Large advertisers typically bid on hundreds of thousands of keywords at any instant. When a user types a query, the search engine identifies all advertisers bidding on that (or a closely related) keyword and displays their ads in an ordered list. The search engine uses the advertisers' bids along with measures of ad relevance to rank order the submitted ads. Whenever a consumer clicks on an ad in a given position, the search engine charges the corresponding advertiser a cost per click (CPC) which is the minimum bid needed to secure that position. The auctions are continuous sealed bid auctions. That is, advertisers can change their bids at any time and cannot observe the bids of their competitors. Typically advertisers are only given summary reports with details such as the total number of impressions, clicks and conversions, average rank and average CPC for each keyword on a given

day. Several of these auctions are very competitive. For example, in March 2006, there were 177 advertisers bidding in the Yahoo auction for the keyword “car insurance” with the maximum bid being \$14.01 per click. The average CPC on search engines has been continually rising over the last couple of years and search advertising is increasingly becoming a major advertising channel for several firms.

The GSP auction described above differs from traditional auctions in a number of ways. First, search engines display multiple ads in response to a user query. However, the auction cannot be treated as a multi-unit auction because each ad position is different in the sense that top positions generate more clicks for the same number of ad impressions. Further, the CPC increases as the rank of an ad decreases (i.e. the CPC is higher for top ranked ad than a lower ranked ad). Thus, the advertiser has to trade-off a higher number of clicks attained at a top position against the lower margin per click. Due to this trade-off, it may sometimes be better for an advertiser to underbid and sacrifice a few clicks in order to get a higher margin per click. Indeed, several authors have demonstrated that popular second-price search auctions such as those used by Google and Yahoo are not incentive compatible (Aggarwal, Hosanagar and Smith 2008; Edelman, Ostrovsky and Schwarz 2007). Thus, bidding one’s true valuation is often suboptimal. Further, advertisers have short-term budget constraints which imply that bids cannot be submitted independently for keywords. For example, if the advertiser submits a very high bid for the keyword “hawaii vacation” then it is likely that the keyword consumes a significant portion of the advertiser’s budget leaving a very limited budget for another keyword. Therefore the bids for the thousands of keywords are inextricably linked. Finally, considerable uncertainty exists in the sponsored search environment. For example, the number of queries for “miami vacation” on any given day is stochastic and is a function of the weather, special events in Miami and a variety of other unknown factors. Similarly, consumer click behavior cannot be precisely predicted and the bids of competitors are also unknown due to the sealed bid nature of the auction. The stochasticity in query arrival, consumer click behavior and competitors’ bids imply that the number of clicks and total cost associated with any bid are all stochastic. All these factors - namely the incentive incompatibility of the auction, budget constraints, large portfolio of keywords and uncertainty in the decision environment - make the advertiser’s problem of bidding in sponsored search a non-trivial optimization problem. In this

paper, we formulate and solve the advertiser’s decision problem.

The paper makes three main contributions. The first is a contribution towards improving managerial practice. Advertisers spend billions of dollars on sponsored search. An entire industry of Search Engine Marketing (SEM) firms have emerged that provide bid management services. The techniques described in the paper can help increase the Return on Investment (RoI) for advertisers and SEM firms. For example, in a field test of the approach, we observe a 26.4 % increase in revenues relative to the approach used by an advertiser. The second key contribution is that our approach represents a significant step forward for the academic literature on bidding in multi-slot auctions. All the papers to date have studied the problem either in a deterministic setting or in a single-slot setting and have relied on heuristic solution techniques due to the complexity of the optimization problem. In contrast, we compute optimal bids in the more realistic stochastic multi-slot setting. The third contribution of this paper is to incorporate the inter-dependencies between the keyword into a multi-period bidding problem.

The rest of the paper is organized as follows. In Section 2, we review the relevant literature and position our work within the literature on sponsored search. In Section 3, we formulate the problem, derive the optimality condition and discuss how it may be used to compute the optimal bids. In Section 4, we demonstrate the application of our model to a real-world dataset on clicks, conversions and bids for an advertiser. We contrast the optimal bids computed by our model with those used by the firm and present results from a field implementation of the bids. Finally, we discuss the limitations of the work and conclude in Section 6.

2 Literature review

Below, we review three streams of active research within the field of sponsored search with a particular emphasis on prior work on bidding in sponsored search.

Mechanism Design: Search engines run Pay Per Click auctions in which they charge advertisers whenever a consumer clicks on an ad.¹ A primary area of focus in sponsored search research has

¹Other payment rules are also feasible. These include Pay Per Action (PPA) auctions in which advertisers are charged only if the consumer performs a valid action such as a purchase. Hybrid schemes are also feasible. For example, in the context of banner ads, Kumar et al. (2007) propose a hybrid pricing model based on a combination of ad impressions and clicks.

been the design of the auction mechanism. Two important questions from a mechanism design perspective are the rules used to rank order the ads and the rules used to determine the amount paid by advertisers. Feng, Bhargava and Pennock (2007) compare the performance of various ad ranking mechanisms and find that a yield-optimized auction, that ranks ads based on the product of the submitted bid and ad relevance, provides the highest revenue to the search engine. Further, they find that ranking based on bid alone fares nearly as well when bids and relevance are positively correlated and editorial filters are used to eliminate ads of low relevance. Weber and Zheng (2007) study competition between two vertically differentiated firms that advertise through a search intermediary and also find that the intermediary's revenue maximizing design ranks the ads based on a weighted average of their bid and product qualities. Liu, Chen and Whinston (2008) also find that efficient auction design is achieved by weighting bids by click-through rate and further report that minimum bids should be higher for advertisers with lower click-through rate. In terms of payment rules, Edelman and Ostrovsky (2007) study first price sponsored search auctions in which advertisers pay the amount they bid and find empirical evidence of bidding cycles in such auctions. The authors indicate that a VCG-based mechanism eliminates such bidding cycles and generates higher revenues for the search engine compared to the first price auction. Zhang and Feng (2005) and Asdemir (2006) present dynamic bidding models that also predict cyclical patterns in bidding. The payment rule commonly used by search engines currently is a second price auction commonly referred to as the GSP auction. Edelman and Ostrovsky (2007) demonstrate that the GSP auction, unlike Vickrey-Clarke-Groves (VCG) mechanism, is not incentive compatible. Thus, advertisers have to bid strategically even in the absence of budget constraints. Varian (2007) also characterizes the Nash equilibrium for a second price multi-slot auction and demonstrates the auction is not incentive compatible. Aggarwal, Goel and Motwani (2006) propose a "laddered" auction mechanism that is incentive compatible but the mechanism has not been adopted possibly due to the complexity of the payment rules.

Consumer behavior in sponsored search: The sponsored search environment presents rich data on consumer behavior. Modeling users' propensity to click on ads and to purchase upon clicking is an important area of recent focus. Several approaches have been proposed to model clicks for individual keywords and ads (Ali and Scarr 2007; Craswell et al. 2008). However the

sparseness of clicks and purchase data makes it hard to estimate individual keyword-level models. Rutz and Bucklin (2008) present a hierarchical Bayesian model of individual keyword conversions and demonstrate that the model effectively deals with the challenges posed by data sparseness and keyword heterogeneity. Ghose and Yang (2008) and Agarwal, Hosanagar and Smith (2008) apply similar models.

Optimal Bidding in Sponsored Search: The stream of work closely related to our paper is that on budget constrained bidding in sponsored search. Rusmevichientong and Williamson (2006) propose a model for selecting keywords from a large pool of candidates. Their model does not however address optimal bidding for these keywords and ignores the multi-slot context. Feldman et al (2007) study the bid optimization problem and indicate that randomizing between two uniform strategies that bid equally on all keywords works well. The authors assume that all clicks have the same value independent of the keyword. Further, their results are derived in a deterministic setting where the advertisers position, clicks and cost associated with a bid are known precisely. Borgs et al (2007) propose a bidding heuristic that sets the same average Return on Investment (RoI) across all keywords. Their model is also derived for a deterministic setting. Finally, Muthukrishnan, Pal and Svitkina (2007) study bidding in a stochastic environment where there is uncertainty in the number of queries for any keyword. The authors focus on a single slot auction and find that prefix bidding strategies that bid on the cheapest keywords work well in many cases. However, they find that the strategies for single slot auctions do not always extend to multi-slot auctions and that many cases are NP hard .

The prior work reveals three themes. One, the literature on sponsored search mechanism design has established that GSP auctions are not incentive compatible. This feature combined with the advertiser budget constraints suggest a need to develop bidding policies. Two, the empirical work in sponsored search provide a variety of useful models that can be applied towards modeling consumer click behavior and bidding behavior of advertisers. These can ultimately be used to develop data-driven optimization strategies. Three, the issue of budget constrained bidding has received some attention. While these early papers on bid optimization have helped advance the literature, they have tended to focus on deterministic settings or single slot auctions, both of which are restrictive assumptions in the sponsored search context. Further, these papers develop heuristic strategies due

to the complexity of the optimization problem. In contrast, we determine optimal bids in a budget-constrained multi-unit multi-slot auction under uncertainty in the decision-making environment.

3 Model

Advertisers usually maintain a portfolio of thousands of keywords. They submit bids for each keyword on a regular basis during a billing cycle (e.g., several firms submit bids on a daily basis). During each time period when bids need to be computed, the bid management system accepts a budget for that time period as an input and computes the bids for all keywords. We adopt the same framework and focus on the bid optimization problem during a specific time period in which the budget and the set of keywords have been specified.²

The effects of sponsored ads are two folds – (i) awareness and (ii) lead generation. Consumers usually start their search process by using generic terms e.g. “hawaii vacation”. Bidding on these generic keyword might help the advertiser generate brand-specific (or retailer-specific) exposure in the form of impressions and clicks. These brand related exposures might enhance the awareness of a particular brand and it might enter the consideration set of a consumer. The increased awareness can subsequently lead to increased subsequent branded search activity (“spillover”). There is strong evidence to suggest that spillovers from generic to branded advertising activity is present in the context of sponsored search ads (Rutz and Bucklin 2010). More directly, sponsored ad are responsible for lead generation and eventual conversion. An advertiser must weigh in these two effect while making his ad decisions. For example, there are several generic keywords that lead to several impressions and clicks but do not directly lead to conversions. The increase the brand awareness which subsequently results in increase search activity for brand the consumers were exposed to.

3.1 Notation and Setup

In this section we introduce our notation and the general framework we use to study the advertiser’s decision problem.

²A common practice in the SEM industry is to use $Daily\ Budget = (Remaining\ Balance)/(Number\ of\ days\ left\ in\ cycle)$, where remaining balance is the initial monthly budget less the amount spent thus far. We do not focus on how the budget for a given time period is computed and treat it as an exogenous parameter in our formulation.

During a given timeperiod (say a day) a keyword k is searched S_k times, where S_k is a random variable. S_k also represents the total number of impressions, i.e. the number of times the advertiser's ad is displayed by the search engine. The expected number of impressions is defined as $\mu_k = \mathbb{E}[S_k]$. The superscript (s) is used to denote the s^{th} search. We denote the bid of the advertiser for the keyword as b_k , and assume that the advertiser does not change the bid during the timeperiod. Every time the keyphrase is searched, the advertiser's ad is placed at some position in the list of all sponsored results. Let $pos_k^{(s)}$ be the position at which the ad was shown in the s^{th} search, with the topmost position denoted position 0. Let $\delta_k^{(s)}$ be an indicator of whether a person who was searching for the keyword clicked on the advertiser's link, or not: $\delta_k^{(s)} = I(\text{click}_k^{(s)})$. An analysis of click-through data by Feng et al. (2007) suggests that the probability that a user clicks on an ad in position $pos_k^{(s)} = i$ is

$$Pr\{\delta_k^{(s)} = 1 | pos_k^{(s)} = i\} = \frac{\alpha_k}{(\gamma_k)^i}$$

where α_k and γ_k are keyword specific constants indicating the probability of a click at the topmost position and the rate at which the click probability decays with the rank of the ad respectively.

The advertiser's value from a click is denoted by an independent random variable w_k . $v_k^{(s)}$ denotes the advertiser's value from the s^{th} impression. It is equal to 0 if the user does not click and equal to w_k if the user clicks ($v_k^{(s)} = \delta_k^{(s)} w_k$). Let $\underline{b}_k^{(s)}$ be the advertiser's cost per click i.e. the bid of the advertiser at the next position $pos_k^{(s)} + 1$. The cost associated with impression s may then be expressed as $c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$.³ Because consumers do not know any advertiser's bid, it seems reasonable to assume that given an ad's position in the list, the probability that a person clicks on the ad does not depend on the bid of the next advertiser. That is, conditional on the position $pos_k^{(i)}$, the vector $(\underline{b}_k^{(i)}, \delta_k^{(i)})$ has independent components. We also assume that S_k is independent of other variables.

Besides the advertiser there are N_k other advertisers who place their bids for keyword k . Initially, we assume that N_k is known to the advertiser. It can be observed, for example, by submitting sample

³The discussion assumes that ads are ordered by bid and that the advertiser pays the bid of the next advertiser. A common practice is to use the product of bid and a quality score to rank order the advertisers, and the payment is the minimum bid needed to secure the position (e.g. the payment per click for an advertiser in position i is $bid(i+1) * Quality(i+1) / Quality(i)$). This does not affect our model. If we normalize the bid of all competitors by the ratio of their quality score relative to our advertiser ($NormalizedBid = bid * Quality_{Competitor} / Quality_{Advertiser}$), our analysis can be interpreted as based on this normalized bid.

queries to the search engine and observing the number of ads displayed. Subsequently, we relax this assumption in Section 3.4 and derive our results for a case in which N_k is a random variable. The bids of the competitors cannot be observed because the auction is a sealed bid auction. The key assumption we make is that the competitors place their bids according to some known distribution $F_k(b)$. This assumption can be interpreted in multiple ways. One way is to assume that each competitor submits his bid according to this distribution. Alternatively, they determine their bids using some process that is unknown but the resulting final distribution of all their bids is $F_k(b)$. Finally, D denotes the advertiser's budget in a given timeperiod of interest. Table 1 summarizes our notation.

3.2 Model Formulation

The advertiser faces the following decision problem:

$$\begin{aligned} \max_{\{b_k\}} E \left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)} | b_k \right] \\ \text{s.t. } E \left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} | b_k \right] \leq D. \end{aligned} \quad (1)$$

The objective is to determine bids b_k in order to maximize the advertiser's expected revenues. The constraint implies that the expected cost should be less than or equal to a budget D . Note that the budget is not modeled as a hard constraint. This is a common format in which budget constraint is specified by advertisers in the SEM industry, and reflects an objective function of the form $\max_{\{b_k\}} E \left[\sum_k \sum_{s=1}^{S_k} (v_k^{(s)} - \lambda c_k^{(s)}) | b_k \right]$. Thus, the objective is to maximize expected profit but the shadow price of ad dollars is specified in the form of a constraint on the expected cost. The optimization problem in Equation (1) always has a solution as shown in Appendix A1 (All important proofs appear in the Appendix). Solving the problem gives the following optimality condition

$$\forall k : \frac{d}{db_k} E \left[\sum_{s=1}^{S_k} v_k^{(s)} | b_k \right] = \lambda \frac{d}{db_k} E \left[\sum_{s=1}^{S_k} c_k^{(s)} | b_k \right]. \quad (2)$$

where λ is the Lagrange multiplier. The optimality condition states that for the optimal bids the ratio of the marginal change in the advertiser's expected revenues over the marginal change

in the advertiser's expected cost should be constant across keywords. An alternative way to interpret it is as follows. If we decrease the bid for keyphrase k' by ε , then the expected cost will decrease by $\varepsilon \frac{d}{db_{k'}} E \sum_{s=1}^{S_{k'}} [c_{k'}^{(s)} | b_{k'}]$ and, hence, we may increase the bid for another keyphrase k by $\varepsilon \frac{d}{db_{k'}} E \sum_{s=1}^{S_{k'}} [c_{k'}^{(s)} | b_{k'}] / \frac{d}{db_k} E \sum_{s=1}^{S_k} [c_k^{(s)} | b_k]$. In this case the expected increase in profits will be

$$\varepsilon \frac{\frac{d}{db_k} E \sum_{s=1}^{S_k} [v_k^{(s)} | b_k]}{\frac{d}{db_k} E \sum_{s=1}^{S_k} [c_k^{(s)} | b_k]} \frac{d}{db_{k'}} E \sum_{s=1}^{S_{k'}} [c_{k'}^{(s)} | b_{k'}] - \varepsilon \frac{d}{db_{k'}} E \sum_{s=1}^{S_{k'}} [v_{k'}^{(s)} | b_{k'}] = 0.$$

We assume that consumer click behavior and competitor bidding behavior is i.i.d across ad impressions during the given time period. Hence, in Expression (2) we may cancel the sums over s . Therefore, the optimal vector of bids should satisfy the following condition:

$$\forall k : \frac{d}{db_k} E [v_k | b_k] = \lambda \frac{d}{db_k} E [c_k | b_k]. \quad (3)$$

3.3 Optimality Condition

It is hard to use the optimality condition (3) to compute the optimal bids. In order to apply (3), the advertiser has to be able to compute $E [v_k | b_k]$ and $E [c_k | b_k]$ accounting for the uncertainty in competing bids and consumer query and click behavior. In this Section, we express the optimality condition in terms of parameters that can be estimated. We assume that the number of competitors N_k is known and is constant. We can identify the number of competitors by performing a search on keyword k at a search engine.

Consider a specific keyword k . We tentatively drop the subscript k as we focus on an individual keyword. In order to compute $E [v | b]$, we need to identify the probability of a click given the bid b , which in turn depends on the probability distribution of the ad position. The conditional probability of being at position i conditional on a bid b is

$$\Pr \{pos = i | b\} = \binom{N}{i} (1 - F(b))^i F(b)^{N-i}. \quad (4)$$

The probability that a competitor bids more than b is equal to $1 - F(b)$, and the probability of

obtaining a position i given bid b is determined by a Bernoulli process. Recollect that the positions start from 0, i.e., the topmost ad has position $pos = 0$. Thus the position i indicates that there are i other advertisers ranked above. Feng, Bhargava and Pennock's (2007) analysis of click-through data suggests that the probability that a user clicks on an ad in position pos is

$$\Pr \{ \delta = 1 | pos = i \} = \frac{\alpha}{\gamma^i}, \quad (5)$$

where α and γ are keyword specific constants. It follows that

$$\begin{aligned} \Pr \{ \delta = 1 | b \} &= \sum_i \Pr \{ \delta = 1 | pos = i \} \Pr \{ pos = i | b \} \\ &= \sum_i \frac{\alpha}{\gamma^i} \binom{N}{i} (1 - F(b))^i F(b)^{N-i} \\ &= \alpha \gamma^{-N} (1 + (\gamma - 1) F(b))^N. \end{aligned} \quad (6)$$

Proposition 1: The expected value of an impression is given by

$$E[v|b] = E[\delta w|b] = \Pr \{ \delta = 1 | b \} Ew = \alpha \gamma^{-N} (1 + (\gamma - 1) F(b))^N Ew. \quad (7)$$

It follows from Proposition 1 that

$$\frac{d}{db} E[v|b] = \alpha N \gamma^{-N} (\gamma - 1) f(b) (1 + (\gamma - 1) F(b))^{N-1}. \quad (8)$$

We now derive an expression for $E[c|b]$. In order to do so, we need to characterize the probability distribution function of the bid of the next advertiser in the list of sponsored results. We first derive some intermediate results.

Lemma 1: The distribution function of the bid of the next advertiser in the list conditional on the bid and the position is given by

$$F(\underline{b}|b, pos = i) = \left(\frac{F(\underline{b})}{F(b)} \right)^{N-i}. \quad (9)$$

Applying,

$$F(\underline{b}|b, pos = i, \delta = 1) = F(\underline{b}|b, pos = i) = \left(\frac{F(\underline{b})}{F(b)}\right)^{N-i}, \quad (10)$$

we can derive the following lemma.

Lemma 2: The conditional distribution of the bid of the next advertiser conditional on the bid and the fact that the ad was clicked is

$$\begin{aligned} F(\underline{b}|b, \delta = 1) &= \sum_{i=0}^N F(\underline{b}|b, \delta = 1, pos = i) \times \Pr\{pos = i|b, \delta = 1\} \\ &= \left(\frac{1 - F(b) + \gamma F(\underline{b})}{1 + (\gamma - 1)F(b)}\right)^N. \end{aligned} \quad (11)$$

When a user clicks on an ad, the advertiser has to pay the bid of the next advertiser in the list.

Applying Lemma 2 and Equation (6) gives us

Proposition 2: The expected cost of an impression is given by

$$\begin{aligned} E[c|b] &= E[\delta \underline{b}|b] \\ &= E[\underline{b}|b, \delta = 1] \Pr\{\delta = 1|b\} \\ &= \alpha \gamma^{-N} \left(b[1 + (\gamma - 1)F(b)]^N - \int_0^b [1 - F(b) + \gamma F(\underline{b})]^N d\underline{b} \right). \end{aligned} \quad (12)$$

Using Proposition 2 we can derive that

$$\frac{dE[c|b]}{db} = \alpha N \gamma^{-N} f(b) \left((\gamma - 1)b[1 + (\gamma - 1)F(b)]^{N-1} + \int_0^b [1 - F(b) + \gamma F(\underline{b})]^{N-1} d\underline{b} \right). \quad (13)$$

Substituting Expressions (8) and (13) in Equation (3),

$$\begin{aligned} \frac{dE[v|b]}{db} &= \lambda \frac{dE[c|b]}{db}, \\ \frac{1}{\lambda} &= \frac{1}{Ew} \left(b + \frac{\int_0^b [1 - F(b) + \gamma F(\underline{b})]^{N-1} d\underline{b}}{(\gamma - 1)[1 + (\gamma - 1)F(b)]^{N-1}} \right). \end{aligned}$$

Proposition 3: The optimality condition (expressed in terms of estimable parameters) is

$$\forall k : \frac{1}{Ew_k} \left(b_k + \frac{\int_0^{b_k} [1 - F_k(b_k) + \gamma_k F_k(b)]^{N_k-1} db}{(\gamma_k - 1)[1 + (\gamma_k - 1)F(b_k)]^{N_k-1}} \right) = const. \quad (14)$$

Proposition 4: A unique bid b_k^* satisfies the optimality condition (Equation 14) for keyword k when

$$\gamma_k > 1 + \frac{1}{F_k(b)} \left[f_k(b)(N_k - 1) \frac{\int_0^b [1 - F_k(b) + \gamma_k F_k(x)]^{N_k-2} dx}{[1 + (\gamma_k - 1)F_k(b)]^{N_k-2}} - 1 \right].$$

The optimality condition can be used in conjunction with the budget constraint to compute the optimal bids. For several common distribution and a wide range of parameters, we show in the appendix that the conditions for a unique bid (Proposition 4) are satisfied. In order to compute the optimal bids, the following keyword-specific constants need to be known: α_k (the click-through rate at the top position), γ_k (rate at which CTR decays with position), Ew_k (expected revenue from a click), N_k (number of competing bidders), and $F_k()$ (distribution of competing bids). We estimate these parameters using a real-world dataset and illustrate how bids may be computed in Section 4.

3.4 Bid Computations

In this section we use the estimated parameters to compute the optimal bids. The optimal bids should satisfy equation (14)

$$\frac{1}{Ew_k} \left(b_k + \frac{\int_0^{b_k} [1 - F_k(b_k) + \gamma_k F_k(b)]^{N_k-1} db}{(\gamma_k - 1)[1 + (\gamma_k - 1)F(b_k)]^{N_k-1}} \right) = const, \quad (15)$$

and the budget constraint,

$$E \left[\sum_k \sum_s c_k^{(s)} \right] = D.$$

The parameters estimated in Section 4.1 satisfy the uniqueness condition specified in Proposition (4). The last constraint can be rewritten as

$$\sum_k \mu_k \alpha_k \gamma_k^{-N_k} \left(b_k [1 + (\gamma_k - 1) F_k(b_k)]^{N_k} - \int_0^{b_k} [1 - F_k(b_k) + \gamma_k F_k(\underline{b})]^{N_k} d\underline{b} \right) = D. \quad (16)$$

For a given *const* in Equation (15), we compute the bid that satisfies the equation for every keyword. Then we use Equation (16) to calculate the expected total cost for the computed bids. If the expected cost is lower than D , we increase the constant, otherwise we decrease it. The process repeats until the expected total cost is sufficiently close to the budget.

4 Data Description

Our dataset is from a leading meat distributor that sells through company owned retail stores as well as through mail-order catalogs and online. This firm bids on thousands of keywords across several search engines and has a substantial online presence. Our dataset consists of daily summary records for 247 keywords which this firm uses to advertise on Google. We divide these keywords in 29 unique product categories which span frozen meats, seafoods and desserts. A comprehensive list of these product categories appears in Table 1. The daily record for every keyword has the following fields,

$$(id, t, b, i, cl, avgcpc, avgpos)$$

where

- id* - Unique identifier for each keyword
- t* - date
- b* - bid submitted by advertiser
- i* - number of impressions during the day
- cl* - number of clicks during the day
- avgcpc* - average cost per click on the day
- avgpos* - average position during the day

This dataset is representative of the type of data available to advertisers in sponsored search. Advertisers only get summary reports from search engines and do not have information on clicks

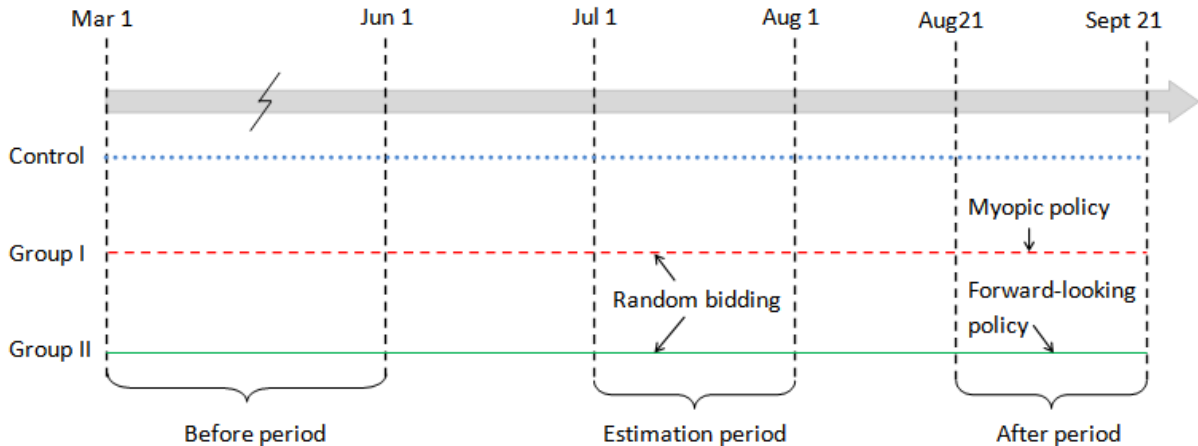


Figure 1: Illustration of the timeline for the various data collection periods.

and position for each individual ad impression.

Our dataset is divided into three distinct periods as shown in Figure 1. The first period runs from March 1-May 31, 2011. This period forms the “before” period for our analysis during which the bids for these 247 keywords were decided by the firm spanning 29 product categories show in Table 1. The summary statistics for the data during this period is shown in Table 2. In this duration, there were 1.36 million impressions of these ads and they received 11,651 clicks in total. The total weekly cost of these ads was \$964 and the weekly gross revenue generated from these keywords was \$1728. We use the data from this period to compute the expected value *per-click* (Ew) and the expected daily impressions (S) for each keyword. For the subsequent analysis we randomly divide these keywords by product categories into three distinct groups. The first group forms the control group for our experiment. We use the keywords in Group I to measure the effectiveness of the myopic bidding policy outlined in Section 3 to and the keywords in Group II to measure the effectiveness of the forward looking policy we will outline in Section 6. The three groups are fairly well matched in terms of the product categories included within each group as well as in terms of summary statistics across these groups which are presented in Table 3. The control group forms the baseline to eliminate any time trends that might enter the analysis due to changes in online activity or search engine strategy.

Table 1: Product Categories

Bacon	Flat Iron	Pork
Beef	Gift Basket	Porterhouse
Beef Jerky	Gifts	Prime Rib
Beef Sirloin	Halibut	Salmon
Burgers	Ham	Shrimp
Catfish	Hot Dogs	Sole
Cheesecake	Lobster	Surf and Turf
Corned Beef	Lobster Bisque	Swordfish
Crab	London Broil	Trout
Filet Mignon	Orange Roughy	

Table 2: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Bid	1.18	1.01	0.35	10.00
Avg CPC	0.73	0.59	0.00	4.42
Avg Pos	3.15	1.90	1.00	12.41
Impressions	5637.22	13106.39	1.00	98373.00
Clicks	48.37	86.76	0.00	593.00
CTR	0.03	0.07	0.00	0.60
Cost	45.95	95.11	0.00	747.43
Revenue	83.26	132.47	0.00	974.31
Gross Profit	37.31	140.78	-747.43	902.20
RPC	4.33	14.30	0.00	158.96

Table 3: Summary for the three groups of keywords.

	Control	Group I	Group II
Products Categories	8	10	10
Keywords	55	89	85
Impressions/Keyword	5637.80	6783.64	4989.04
Clicks/Keyword	50.53	64.26	35.81
CTR	0.0089	0.0095	0.0072
CPC	0.92	1.04	1.00
RPC	1.69	1.71	1.85
ROI	0.84	0.65	0.85

The second period spans from July 1-July 31, 2011 which we refer to as the “estimation period” for our analysis. We ignore the month of June from our analysis as there is a huge increase in online activity during this month due to Father’s Day. During this period bids for the control group are held constant but we submit random bids for the keywords in Groups I and II. The bids are uniformly drawn from $\$0.10 \times [1, 30]$ resulting in a minimum bid of 10¢ to a maximum bid of \$3.00. The upper limit of \$3.00 was prescribed by the advertiser. The bids are drawn weekly which leads to four unique bids per keyword in the estimation period. This variation in bids leads to a significant variation in the ad position and helps the identification of the parameters of the model. The exact identification strategy is discussed later on in Section 5.

Finally, optimal bids are computed based on estimated parameters and deployed by the firm between August 21-September 21, 2011. These bids are re-estimated every week incorporating recent data in the analysis. This period forms our “after” period. Data from the after period is used to assess the effectiveness of both the myopic policy that ignores interaction between keywords and the forward looking policy that incorporates the interaction between keywords. In Section 5 we discuss the estimation of parameters using data from the “estimation period”. Subsequently, we discuss the results from the field implementation of the myopic policy.

5 Empirical Analysis

In this section we apply our technique to a real-world dataset of clicks and costs for several keywords and derive the optimal bids for these keywords. We then describe the results from a field implementation of the suggested bids.

5.1 Estimation Approach

Our data provide daily summary measures (average position, average cost per click, total clicks) but not the outcome of each individual impression. Given just these daily summary measures, it is hard to apply regression or Maximum Likelihood Estimation techniques directly on the aggregated data, hence we use the Generalized Methods of Moment (GMM) approach to estimate these parameters. Following the idea of the method of moments, we derive analytical expressions for the

moments we observe empirically, namely, the expected position ($avgpos_t$), cost per click ($avgcpc_t$) and click-through rate ($ctr_t = cl_t/i_t$) for every keyword, where t is the time index. These moment are as follows,

$$E(pos_t|b_t) = N_t(1 - F(b_t)), \quad (17)$$

$$E(\underline{b}_t|b_t, \delta_t = 1) = \int_{x < b_t} x d \left(\frac{1 - F(b_t) + \gamma F(x)}{1 - (1 - \gamma) F(b_t)} \right)^{N_t}, \quad (18)$$

$$E(\delta_t|b_t) = \alpha \gamma^{-N_t} (1 - (1 - \gamma) F(b_t))^{N_t}. \quad (19)$$

The observed moments can be expressed in terms of the analytical moments in the following manner

$$\begin{aligned} avgpos_t &= E(pos_t|b_t) + \xi_{1t}, \\ avgcpc_t &= E(\underline{b}_t|b_t, \delta_t = 1) + \xi_{2t}, \\ ctr_t &= E(\delta_t|b_t) + \xi_{3t}, \end{aligned}$$

where $\xi_t = (\xi_{1t}, \xi_{2t}, \xi_{3t})'$ is the random shock. This is the most general formulation of our model. However, we make several assumptions in order to estimate this model from the available data – (i) As the data contains only the daily aggregates, we cannot directly estimate the distribution function $F(b)$ using nonparametric approaches as we have very few bids for every keywords. To address this constraint associated with our dataset, we enforce a parametric form on the distribution $F(b)$, and estimate its parameters using the first moments associated with the position, cost per click and click-through rate. For the parametric form of the distribution $F(\cdot)$ we choose the Weibull distribution. This choice is based on two factors. Firstly, the Weibull distribution can take on diverse shapes and offers a great deal of flexibility. Secondly, an analysis of a secondary dataset of bids submitted to a search engine for several keywords in the insurance sector (Abhishek, Hosanagar

and Fader, 2009) shows that the Weibull distribution is reasonably good for modeling the bids.⁴ Note that we are not assuming that the distribution of bids for keywords is the same across the two datasets, rather that the bids are from the same family (Weibull) and the parameters can vary across keywords. The Weibull distribution has the following cumulative distribution function

$$F(x; \theta, \lambda) = 1 - \exp\left\{-\left(\frac{x}{\lambda}\right)^\theta\right\}.$$

It is defined by two parameters θ and λ . Therefore, we have four unknowns parameters for any keyword $(\lambda, \theta, \alpha, \gamma)$ and 3 moment conditions for every unique bid. (ii) The other simplifying assumptions we make are that N_t is deterministic during the day. N_t on a particular day is estimated as the average number of competing ads in the in the estimation period. This assumption might be problematic when N_t is small as the variation in the ad position can be driven partly by a variation in N_t , however when N_t is sufficiently large as in our dataset, this does not effect the estimation procedure significantly.

The estimates of the parameter $\beta = (\alpha, \gamma, \lambda, \theta)$ is given by

$$\hat{\beta} = \arg \min_{\beta \in B} \boldsymbol{\xi}(\beta)' W \boldsymbol{\xi}(\beta),$$

where $\boldsymbol{\xi}(\beta)$ is vector of error between the observed and computed moments for a particular keyword during the observation period and W is a weighting matrix. The choice of W is critical as it determines the asymptotic properties of the estimator and in our analysis W is a consistent estimate of $\mathbb{E}[\boldsymbol{\xi}(\beta)' \boldsymbol{\xi}(\beta)]$. As we do not know $\mathbb{E}[\boldsymbol{\xi}(\beta)' \boldsymbol{\xi}(\beta)]$, an iterative-GMM estimator is used (see Hanson, Heaton and Yaron, 1996) where the weighting matrix is iteratively re-estimated till it converges.

In order to compute the optimal bids we also need to know Ew , the expected *revenue per-click*. The expected *revenue per-click* is provided to us by the firm and we assume that it does not change over the course of this field experiment. The firm uses a technique similar to the first-click attribution method that is commonly used in online advertising. More sophisticated techniques can be used to estimate Ew but this discussion is outside the scope of this current paper. It should be

⁴We tried several distributions like the Normal, Log-Normal, Gamma, Exponential and Logit but the Weibull distribution is the best representation for tha advertisers' bids. This is not a resrestrictive assumption and any parametric distribution can be plugged in the general estimation procedure outline here.

kept in mind that we do not assume that the value per-click is constant but assume that conditional on a click, the consumer’s subsequent behavior is i.i.d (i.e. independent of the position of the ad etc.). Although this might not true under all circumstances as pointed out by Ghose and Yang (2009) and Agarwal et. al. (2011), this assumption is commonly used in practice and academic literature (Rusmevichientong and Williamson, 2007).

5.2 Identification Strategy

The parameters of this model can be estimated if we have at least 2 unique bids per keyword in the data. However, there are two important reasons why historical data cannot be used to estimate the parameters of this model - (i) insufficient variation in bids (ii) potential endogeneity in advertiser’s bids.⁵

5.2.1 Limited Variation in Bids

In typical SSA campaigns advertisers change their bids very infrequently, usually once in several months, hence it is difficult to identify the parameters of the model. In our dataset, there are very few changes in the bids for the keywords and the average number of unique bids per keyword is 1.12. Discussions with the campaign manager revealed that the bids for these keywords are updated very infrequently, sometimes on a quarterly basis given the huge portfolio of keywords. This phenomenon is not specific to our firm but is widely applicable in the industry and has been pointed out earlier by Rutz and Bucklin (2011) and Ghose and Yang (2009). As our model is underidentified with less than two unique bids, we use the period of random bidding to generate random bids which would lead to sufficient variability in the bids drawn for a particular keyword across days. This in turn leads to a significant variation in the ad position and helps the identification of the parameters of the model.

5.2.2 Endogeneity of Bids

The second concern with using historical data is the potential endogeneity of bids. In order for the GMM to provide consistent estimate we require that $E[b\xi] = 0$ or the bids and the random

⁵We thanks the reviewers for the suggestion to discuss the identification strategy.

shock are independent of each other. However, as the firm is bidding strategically it might increase the bid for a particular keyword if there is a random increase in demand, for example on a sunny weekend. Hence it is very likely that the bids for a particular keyword are correlated with these random shock in the before period. Since we as researchers are unaware of these random shocks which the advertiser might be aware of, we need to correct for this potential endogeneity in bids. One way to address this endogeneity can be using an instrumental variable (IV) approach. Another way to address this issue can be using random bids. Given the lack of variation in the bids in the before period, we resort to the latter technique.

5.3 Estimation Details and Results

In order to estimate the parameters, a nonlinear solver is used to estimate the parameters in our implementation. ⁶ The parameter estimates for a few representative keywords are shown in Table 4. \bar{N} represents the mean number of daily competing ads in the observation period. For brevity, we plot the distribution of the estimated parameters for all keywords in Groups I and II in Figure XXX. A complete table is available from the authors upon request.

Table 4: Parameter estimates for a sample subset of keywords.

keyword	λ	θ	α	γ	\bar{N}	$Ew(\$)$
beef sirloin steak	1.7651	0.5351	0.0266	2.1237	9.5	0.00
Steak Burger	0.6697	2.1944	0.0069	1.2915	5.1	1.53
cheesecakes	0.9736	1.3265	0.0004	1.6091	7.0	1.08
Porterhouse Steak	1.1413	0.8639	0.008	1.1661	4.6	0.43
smoke salmon	1.3414	1.1752	0.0073	1.0255	10.1	6.62
corned beef	1.5368	0.7492	0.0018	1.0175	10.7	3.80
hot dog order	1.0769	1.0869	0.0101	1.6486	7.3	3.00
birthday gifts	1.1756	0.842	0.0009	1.0659	40.2	5.74
birthday present	0.7524	1.3841	0.0122	1.0434	7.1	0.45
lobster bisque	1.311	1.0074	0.0145	1.9293	11.3	0.00

Although there is significant heterogeneity between the keywords, the estimated parameter values are fairly typical in a sponsored search context. The mean click-through rate (α) at the topmost position is 0.026 and the mean decay parameter (γ) is 1.68 which is similar to the values reported earlier (Feng et. al.,2007, Craswell et. al., 2008). Clearly, given the variation in the

⁶We use the Fletcher-Xu hybrid method provided as a part of the ClsSolve routine in TOMLAB.

revenue per-click (Ew) across keywords some keywords are more profitable than other, however the relative competitiveness of the keywords should be accounted for while deciding the optimal bids of these keywords.

5.4 Field Implementation

Once we estimate parameters α, γ, λ and θ for all keywords in Group I and II, we estimate the optimal bids for these keywords. In this section we will focus on the myopic policy outlined in Section 3 and hence we discuss the results of the field implementation only for Group I. In the subsequent section, once we outline the details of the forward looking policy, we will revisit the field implementation details for keywords in Group II.

For the keywords in Group I, we use a daily budget $D = \$72.00$ based on the mean weekly spending of around \$500 during the 3 month “before” period. The bids are recomputed every week after we estimate the parameters $(\alpha, \gamma, \lambda, \theta)$ based on new data, however there is not much change in the bids across weeks. The bids are recomputed to account for competitive reaction as well as changes in the parameters driving consumer click behavior. A more detailed discussion on competitive reaction follows in Section 7.

Table 5: Parameter estimates for a sample subset of keywords.

keyword	Old Bids (\$)	New Bids (\$)
beef sirloin steak	0.82	0.00
Steak Burger	2.19	0.95
cheesecakes	0.66	0.70
Porterhouse Steak	0.76	0.30
smoke salmon	1.16	2.55
corned beef	0.31	3.00
hot dogs order	0.76	1.85
birthday gifts	0.96	1.75
birthday present	1.61	0.20
lobster bisque	0.46	0.00

The rationale for these bids can be inferred from the parameters listed in Table 4. Consider, for example, bids for keywords smoke salmon, hot dogs order and birthday gifts. Our algorithm suggests increasing their bids. For keywords smoke salmon and hot dogs order, we observe that their expected value per click (Ew) is high and it makes sense that the algorithm is suggesting

that we increase their bids. The keyword birthday gifts has a substantially Ew , yet its bid is not raised by a significant amount. This is because the keyword is very expensive (low θ) and it is very difficult to attain the top position. There are other keywords where it is worthwhile to spend the advertising dollars.

The suggested bids were deployed in the field by the advertiser for a period of 4 weeks.⁷ During the 12 weeks in the “before” period, the firm spent a total of \$5937.58 on the keywords in Group II and obtained revenues of \$9776.10. In the “after” period, the total cost and total revenues associated with the keywords were \$3178.82 and \$4594.43 respectively. In the same periods the total cost and total revenues associated with the control keywords were \$1667.54 and \$1480.80 respectively. We use a Difference-in-Difference approach to compute the effect of our algorithm. The improvement in performance due to the algorithm is given by

$$\begin{aligned}
 \tau_{Algo} &= \Delta ROI_{\text{Group I}} - \Delta ROI_{\text{Control}} \\
 &= (44.53\% - 64.65\%) - (-11.20\% - 84.30\%) \\
 &= 75.38\%
 \end{aligned}$$

The performance of the advertising campaign increases by 75.38% as a result of the proposed algorithm. We however notice that there is an absolute decrease in the ROI in the campaign compared to the “before” time period and this decrease is very noticeable for the Control group. After several discussions with the advertiser, we concluded that this decrease is due to some changes in the manner that the search engine displays search results. From July onwards, the search engine started showing a yellow background color for some of the top ads which reported to a significant increase in the CTR of the ads. This phenomenon has been reported by several blogs. A more careful analysis of the advertiser’s portfolio revealed that not only had the CTR increased for similar keywords in the same duration, but also the performance of these keywords had fallen considerably during this time. For the Control group we see an increase in the CTR from 0.89% to 1.4% and for the keywords in Group II we see a change from 1.04% to 1.15%. The increase in CTR might account for the increased costs of these keywords.

⁷These parameters and the bids are re-estimated every week, but there is very little change in the parameters or the suggested bids during this period.

6 Incorporating Interdependence between Keywords

The preceding discussion assumes that keywords are independent of each other. In reality, consumers search across several keywords before making a purchase decision and this might lead to interaction between keywords. E.g. a consumer might start of his search with a generic keyword like “hawaii vacation” but eventually ends up searching for “Marriott hawaii vacations”. While searching for vacations in Hawaii, he could have been exposed to ads from Marriott which became a part of his consideration set. As a result, he searched for “Marriott hawaii vacations” which potentially lead to a purchase from Marriott. If the advertiser was not advertising on the generic keyword “hawaii vacations”, the consumer could never be exposed to the ad from Marriott and this would lead to a lost sale for Marriott. This example illustrates that although some ads (e.g. generic ads) do not lead to direct conversions, they help the advertiser in creating awareness about the brand/product which can increase the probability that it becomes a part of the consumer’s consideration set. More specifically in our context, these brand related exposures can lead to enhanced awareness of a particular brand which might subsequently lead to increased branded search activity (“spillover”). Rutz and Bucklin (2010) show that there are considerable spillovers from generic to branded search activity in the context of sponsored search ads. This suggests that there are complementarity effects between keywords and advertisers should incorporate the effects of spillovers while making their bidding decisions.

6.1 Measuring Spillovers

In order to incorporate the spillover-effect in our decision model we first need to estimate the changes in awareness due to search activity and its affect on future search activity. We use the Nerlove-Arrow model (Rutz and Bucklin 2010, Naik et. al. 1998, Nerlove and Arrow 1962) to capture the evolution of awareness

$$\frac{dA_t}{dt} = -\tilde{\alpha}^A A_t + \beta \mathbf{X}_t, \quad (20)$$

where A_t refers to the awareness level at time t , $\tilde{\alpha}^A$ measures the decay of awareness with time, \mathbf{X}_t is a vector of covariates that capture the search activity at time t and β captures the extent

to which different kinds of search activity affect the level of awareness. Let the keywords belong to two groups: G - generic and B - branded. The total number of generic impressions at time t is defined as $IMP_t^G = \sum_{k \in G} i_{k,t}$ and the total number of branded impressions at time t is defined as $IMP_t^B = \sum_{k \in B} i_{k,t}$, where $i_{k,t}$ are the impressions for keyword k at time t . Similarly, the total number of generic and branded clicks at time t are defined as $CLK_t^G = \sum_{k \in G} cl_{k,t}$ and $CLK_t^B = \sum_{k \in B} cl_{k,t}$, respectively. As we only observe daily data, we use a discrete time analogue of the model presented in Equation (20),

$$A_t = \alpha^A A_{t-1} + \beta_G^{CLK} CLK_{t-1}^G + \beta_B^{CLK} CLK_{t-1}^B + \varepsilon_t^A, \quad (21)$$

where $\alpha^A = (1 - \tilde{\alpha}^A)$ captures the carry-over rate of awareness and ε_t^A is the idiosyncratic error term. Like Rutz and Bucklin (2010) we assume that the awareness at time t is affected by the generic search activity at time t but in addition we also assume that branded search activity might also have an impact on awareness. Awareness is not observed in the data and is latent in this state-space model.

Equation (21) specifies the manner in which search activity affects awareness. Next, we outline how awareness affects both generic and branded search activity. In our model, we assume that awareness only affects the consumer's propensity to search but it has no effect on the consumer behavior after the search is executed. This implies that awareness affects the number of impressions (queries) but has no impact on the *click-through* or *conversion rates*. This assumption is in keeping with the findings of Rutz and Bucklin (2010) who show that awareness does not have a statistically significant impact on *click-through* and *conversion* rates. We assume that awareness leads to change in the search activity for the generic and branded ads in the following manner,

$$IMP_t^G = \alpha_G^{IMP} IMP_{t-1}^G + \gamma_G^{IMP} A_{t-1} + \varepsilon_t^{G,IMP}, \quad (22)$$

$$IMP_t^B = \alpha_B^{IMP} IMP_{t-1}^B + \gamma_B^{IMP} A_{t-1} + \varepsilon_t^{B,IMP}. \quad (23)$$

It should be noted that the affect of awareness is computed from aggregate data (not individually for each keyword). In order to have a parsimonious model we assume that the spillover effect is

constant across all branded keywords and is given by γ_B^{IMP} . Similarly, the spillover effect across generic keywords is constant and is given by γ_G^{IMP} .

We use a Dynamic Linear Model (DLM) to estimate the parameters in Equations (21)-(23). Combining these equation we get

$$\begin{bmatrix} IMP_t^G \\ IMP_t^B \\ A_t \end{bmatrix} = \begin{bmatrix} \alpha_G^{IMP} & & \gamma_G^{IMP} \\ & \alpha_B^{IMP} & \gamma_B^{IMP} \\ \beta_G^{IMP} & \beta_B^{IMP} & \alpha^A \end{bmatrix} \begin{bmatrix} IMP_{t-1}^G \\ IMP_{t-1}^B \\ A_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \sum_{i=G,B} \beta_i^{CLK} CLK_{t-1}^i \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{G,IMP} \\ \varepsilon_t^{B,IMP} \\ \varepsilon_t^A \end{bmatrix}$$

where the correlated error terms ε_t^i account for random shocks and $\varepsilon \sim N(\mathbf{0}, V_\varepsilon)$. DLM have been used in several situations where an important component of the model is unobserved (Bass et. al. 2007, Naik et. al. 2009, Rutz and Bucklin 2010). We estimate this model a Markov Chain Monte Carlo (MCMC) approach as proposed by West and Harrison (1997). The variation in the number of impressions and clicks for generic and branded keywords help us identify the parameters of the model.

The estimated parameters of the model are as follows

[Table containing the estimation results]

We find that there is a significant spillover from generic search activity into branded search. First, we find that there is a strong positive impact of generic clicks on awareness ($\beta_G^{CLK} > 0$). Second, increased awareness leads to increased branded search activity ($\gamma_B^{IMP} > 0$). We can combine these results to say that every click on a generic ad increases the number of branded impressions by $\gamma_B^{IMP} \beta_G^{CLK}$. We also observe that the effects of branded click on awareness and awareness on generic search activity are insignificant ($\beta_B^{CLK} \approx 0, \gamma_B^{IMP} \approx 0$). These finding are consistent with the results reported by Rutz and Bucklin (2010). It is reasonable to believe that awareness about a particular brand does not affect consumers' generic search behavior. Similarly, if a consumer is already aware of a brand, clicking on a branded ad is less likely to change his awareness about the brand. In the subsequent discussion we assume that only generic clicks affect future search behavior and this effect is limited to branded searches. We also assume that all generic

clicks are the same and have exactly the same impact on all branded keywords. More formally,

$$\begin{aligned}\mu_{k,t+1} &= \alpha_B^{IMP} \mu_{k,t} + \gamma_B^{IMP} \beta_G^{CLK} CLK_{G,t-1}^{CLK} \quad \forall k \in B, \\ \mu_{k,t+1} &= \alpha_G^{IMP} \mu_{k,t} \quad \forall k \in G.\end{aligned}\tag{24}$$

Equation (24) shows that every click on a generic ad on average leads to $\gamma_B^{IMP} \beta_G^{CLK} = [InsertValue]$ more searches (and hence impressions) of all branded ads. The increased impressions for branded keywords, which are usually more profitable, leads to higher revenues in future periods.

6.2 Bid Computation with the Awareness Effect

Bidding on keywords has two effects - current period revenues and future awareness as discussed earlier. We see in the preceding analysis, awareness is generated primarily through clicks on generic ads which usually have a low expected value per click. As a result, the advertiser faces a trade-off between maximizing current period revenues or increasing awareness (through more generic clicks) to increase revenues in the future. We consider the advertiser's problem of deciding the bids for the keywords in each time period so as to maximize the total profits. Most sponsored search campaign are active for a fixed amount of time hence we assume that the campaign runs for T time periods. We also assume that the ad spend in every time period should be less than D . Advertisers usually decide their daily or weekly advertising budget which do not change frequently. In fact, several search engines allow advertisers to set a limit on the daily expenditure related to an ad and once the advertising expenditure goes above this amount, the search engines stop displaying this ad. Keeping these assumptions in mind, the multi-period bidding problem can be stated as follows

$$\max_{\{\bar{\mathbf{b}}_t\}} \sum_{t=1}^T r(\bar{\boldsymbol{\mu}}_t, \bar{\mathbf{b}}_t) \quad \text{s.t.} \quad \sum_k \mu_{k,t} c_k(b_{k,t}) \leq D, \quad t = 1, \dots, T$$

where $\bar{\boldsymbol{\mu}}_t = (\mu_{1,t}, \dots, \mu_{K,t})^T$ is a vector of the expected number of impressions and $\bar{\mathbf{b}}_t$ is a vector of bids for each keyword in period t . The current period profit is computed using the formula in Equation (1). For the sake of exposition, we define the ad spend in time periods t as $C(\bar{\boldsymbol{\mu}}_t, \bar{\mathbf{b}}_t) = \sum_k \mu_{k,t} c_k(b_{k,t})$. We formulate a finite horizon dynamic program with T periods to

solve this problem.

$$\begin{aligned}
V(1, \bar{\mu}_1) &= \max_{\{\bar{\mathbf{b}}_t\}} \sum_{t=1}^T r(\bar{\mu}_t, \bar{\mathbf{b}}_t) \quad \text{s.t. } C(\bar{\mu}_t, \bar{\mathbf{b}}_t) \leq D, \quad t = 1, \dots, T, \\
&= \max_{\bar{\mathbf{b}}_1 \text{ s.t. } C(\bar{\mu}_1, \bar{\mathbf{b}}_1) \leq D} \left\{ r(\bar{\mu}_1, \bar{\mathbf{b}}_1) + \left(\max_{\{\bar{\mathbf{b}}_t \text{ s.t. } C(\bar{\mu}_t, \bar{\mathbf{b}}_t) \leq D\}} \sum_{t=2}^T r(\bar{\mu}_t, \bar{\mathbf{b}}_t) \right) \right\}, \\
&= \max_{\bar{\mathbf{b}}_1 \text{ s.t. } C(\bar{\mu}_1, \bar{\mathbf{b}}_1) \leq D} \left\{ r(\bar{\mu}_1, \bar{\mathbf{b}}_1) + V(2, \bar{\mu}_2) \right\},
\end{aligned}$$

where $V(t, \bar{\mu}_t)$ is the value function at time t . More generally, the Bellman equation for this problem is as follows

$$V(t, \bar{\mu}_t) = \max_{\bar{\mathbf{b}}_t \text{ s.t. } C(\bar{\mu}_t, \bar{\mathbf{b}}_t) \leq D} \left\{ r(\bar{\mu}_t, \bar{\mathbf{b}}_t) + V(t+1, \bar{\mu}_{t+1}) \right\}.$$

$\bar{\mu}$, the vectors of mean impressions, constitute the state-space and the bids, $\bar{\mathbf{b}}$, are the control variable. The state evolves in a manner shown earlier in Equation (24). As this is a finite horizon problem, we use backward induction to solve for the optimal bids. At $t = T$, the advertiser does not care about awareness and the optimal policy in the last stage is to bid according to the policy outlined in Proposition 3. In order to find the optimal bids for $t < T$, we use approximate dynamic programming. Lets assume that the expected number of generic clicks at time t are $CLK = \{0, 1, \dots, M\}$. For every $CLK \in \{0, 1, \dots, M\}$, we evaluate the subsequent state and optimal revenues in period $t + 1$. We now solve the problem in Equation (1) with the additional constraint that there are exactly CLK generic clicks in period t . The optimal policy in this period is to choose a CLK (and the associated bids, $\bar{\mathbf{b}}_t$) that maximize the sum of current reward and the optimal future rewards. For the field experiment, one time period spans an entire week and there are $T = 4$ time periods in total. The optimal bids computed for every period are as follows

[Optimal Bids]

[Results of this forward looking policy]

6.3 Comparison of the Myopic and Forward-Looking Policies

In this section we compare the model with independent keywords with the model that subsumes the effect of awareness. We observe that....**[Analysis]**

7 Discussions

8 Conclusions

The presence of a large portfolio of keywords, multiple slots for each keyword and significant uncertainty in the decision environment make an advertiser’s problem of bidding in sponsored search a challenging optimization problem. In this paper, we formulated the advertiser’s decision problem and analytically derived the optimality condition. Our bid optimization model addresses a major gap in prior work related to incorporating multiple slots per item and the uncertainty in competitor bidding behavior and consumer query and click behavior. We illustrated the technique using a real-world dataset. A field test suggests that the approach can substantially boost advertisers’ RoI.

[Write about the awareness effect, trade-off between generic and branded keywords. Discuss parameter estimates. Discuss comparison of the two methods.]

There are a number of interesting avenues along which our work can be extended. We discuss these below.

Exploration and Learning: Our analysis assumes that keyword-specific parameters are known or can be easily estimated based on recent historical data. If there has been sufficient bid exploration in the recent history, these parameters can be estimated as demonstrated in our empirical study. However new keywords and keywords for which bids have settled down into a relatively narrow range present a challenge. Thus an important area of opportunity to further extend our work is to combine optimization with a suitable exploration technique. Exploration is clearly expensive but facilitates more accurate estimation of parameters. Heuristics proposed for Multi-armed Bandit and budget constrained Multi-armed Bandit problems are particularly relevant for balancing exploration and exploitation.

Modeling Advertiser Heterogeneity: The key assumption we make in this paper is that competitor bids are drawn from the same distribution. This allows us to keep the model tractable and solve the complex stochastic optimization problem faced by an advertiser but ignores heterogeneity among competitors. Modeling heterogeneity among advertisers and strategic play is an important next step for our research. Additionally, our focus in this paper, like that of the stream of work on optimal bidding, is the operational bid determination problem faced by an advertiser at any given

instant rather than an economic analysis of the long-term equilibrium that results from the bidding strategies of advertisers in a market. Equilibrium analysis is another interesting direction, albeit a complex one in this setting due to the presence of multiple keywords and a budget constraint.

Appendix A1: Proofs

Solution of Equation (1)

The constrained optimization problem is as follows

$$\max_{\{b_k\}} E \left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)} \right], \quad \text{s.t.} \quad D - E \left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \right] \geq 0.$$

The Lagrangian can be written as:

$$\mathcal{L} = E \left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)} \right] + \lambda \left\{ D - E \left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \right] \right\}.$$

KKT Conditions

$$\begin{aligned} \forall k : \frac{d\mathcal{L}}{db_k} &= \frac{d}{db_k} E \left[\sum_{s=1}^{S_k} v_k^{(s)} | b_k \right] - \lambda \frac{d}{db_k} E \left[\sum_{s=1}^{S_k} c_k^{(s)} | b_k \right] = 0 \\ \lambda &\geq 0, \\ D - E \left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \right] &\geq 0. \end{aligned}$$

Assuming the budget constraint is binding (i.e. $\lambda > 0$), then there exists an extremum s.t.

$$\forall k : \frac{d}{db_k} E \left[\sum_{s=1}^{S_k} v_k^{(s)} | b_k \right] = \lambda \frac{d}{db_k} E \left[\sum_{s=1}^{S_k} c_k^{(s)} | b_k \right]$$

As $\text{rank} \left(\frac{d(D - E[\sum_k \sum_{s=1}^{S_k} c_k^{(s)}])}{db} \right) > 0$, there exists at least one local maxima, and it maximizes the objective function if it is unique.

Assuming $v_k^{(s)}, c_k^{(s)}$ are i.i.d., the optimality condition reduces to

$$E[S_k] \frac{dE[v_k | b_k]}{db_k} = \lambda E[S_k] \frac{dE[c_k | b_k]}{db_k},$$

$$\text{or } \frac{dE[v_k|b_k]}{db_k} = \lambda \frac{dE[c_k|b_k]}{db_k}.$$

Proof of Lemma 1

The probability that the bid of the next advertiser is less than x for some $x < b$ conditional on the bid b and the position i is equal to the probability that exactly i advertisers bid more than b and exactly $N - i$ advertisers bid less than x divided by the probability that the position is i . That is,

$$\begin{aligned} F(\underline{b} = x|b, pos = i) &= \Pr\{\underline{b} < x|b, pos = i\}, \\ &= \frac{\Pr\{\underline{b} < x, pos = i|b\}}{\Pr\{pos = i|b\}}, \\ &= \frac{\binom{N}{i} (1 - F(b))^i F(x)^{N-i}}{\binom{N}{i} (1 - F(b))^i F(b)^{N-i}}, \\ &= \frac{F(x)^{N-i}}{F(b)^{N-i}}. \end{aligned}$$

Proof of Lemma 2

$$\begin{aligned} F(\underline{b} = x|b, \delta = 1) &= \Pr\{\underline{b} < x|b, \delta = 1\}, \\ &= \sum_{i=0}^N \Pr\{\underline{b} < x|b, \delta = 1, pos = i\} \times \\ &\quad \Pr\{pos = i|b, \delta = 1\}, \\ &= \sum_{i=0}^N F(x|b, pos = i) \times \\ &\quad \frac{\Pr\{\delta = 1|b, pos = i\} \Pr\{pos = i|b\}}{\Pr\{\delta = 1|b\}}, \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=0}^N \left(\frac{F(x)}{F(b)} \right)^{N-i} \times \\
&\quad \frac{\frac{\alpha}{\gamma^i} \binom{N}{i} (1-F(b))^i F(b)^{N-i}}{\alpha \gamma^{-N} (1+(\gamma-1)F(b))^N}, \\
&= \sum_{i=0}^N \binom{N}{i} (\gamma F(x))^{N-i} \frac{(1-F(b))^i}{(1+(\gamma-1)F(b))^N}, \\
&= \frac{(1-F(b) + \gamma F(x))^N}{(1+(\gamma-1)F(b))^N}.
\end{aligned}$$

Proof of Proposition 2

$$\begin{aligned}
E[c|b] &= E[\delta \underline{b}|b], \\
&= \Pr\{\delta = 1|b\} E[\underline{b}|b, \delta = 1], \\
&= \alpha \gamma^{-N} [1 + (\gamma - 1)F(b)]^N \int_0^b \underline{b} d \left(\frac{1 - F(b) + \gamma F(\underline{b})}{1 + (\gamma - 1)F(b)} \right)^N, \\
&= \alpha \gamma^{-N} \int_0^b \underline{b} d[1 - F(b) + \gamma F(\underline{b})]^N, \\
&= \alpha \gamma^{-N} \left(b[1 + (\gamma - 1)F(b)]^N - \int_0^b [1 - F(b) + \gamma F(\underline{b})]^N d\underline{b} \right). \quad (\text{Integrating by parts})
\end{aligned}$$

Proof of Equation 13

$$\begin{aligned}
\frac{dE[c|b]}{db} &= \alpha \gamma^{-N} \left([1 + (\gamma - 1)F(b)]^N + N(\gamma - 1)bf(b)[1 + (\gamma - 1)F(b)]^{N-1} \right. \\
&\quad \left. - [1 - F(b) + \gamma F(b)]^N \cdot 1 + Nf(b) \int_0^b [1 - F(b) + \gamma F(x)]^{N-1} dx \right), \\
&= \alpha N \gamma^{-N} f(b) \left((\gamma - 1)b[1 + (\gamma - 1)F(b)]^{N-1} + \int_0^b [1 - F(b) + \gamma F(x)]^{N-1} dx \right).
\end{aligned}$$

Proof of Proposition 4

Let $h_N(b) = \int_0^b [1 - F(b) + \gamma F(x)]^N dx$, $g_N(b) = [1 + (\gamma - 1)F(b)]^N$ and $\Psi(b) = b + h_{N-1}(b)/((\gamma - 1)g_{N-1}(b))$. If $\Psi(b)$ is monotonically increasing then there is a unique b^* that satisfies the optimality

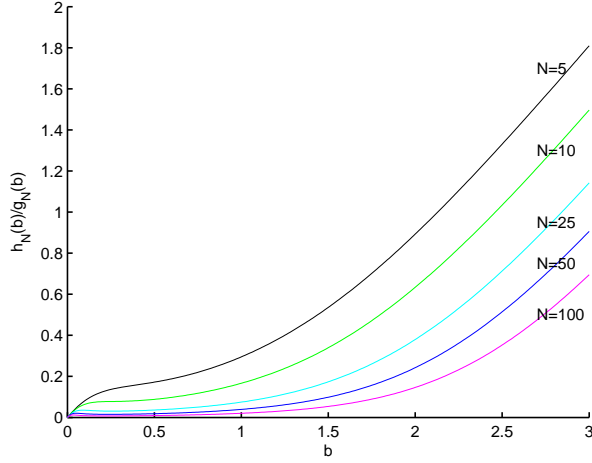


Figure 2: $h_N(b)/g_N(b)$ v/s b assuming the competitors bids are Weibull($\lambda = 1.59, \theta = 1.37, \gamma = 1.42$). The ratio $h_N(b)/g_N(b)$ decreases as N increases.

condition (Equation 14).

$$\begin{aligned}\Psi(b) &= b + \frac{h_{N-1}(b)}{(\gamma - 1)g_{N-1}(b)} \\ \Psi'(b) &= 1 + \frac{h'_{N-1}(b)}{(\gamma - 1)g_{N-1}(b)} - \frac{h_{N-1}(b)g'_{N-1}(b)}{(\gamma - 1)g_{N-1}^2(b)}, \\ &= \frac{g_{N-1}(b)[(\gamma - 1)g_{N-1}(b) + h'_{N-1}(b)] - h_{N-1}(b)g'_{N-1}(b)}{(\gamma - 1)g_{N-1}^2(b)}.\end{aligned}$$

$\Psi'(b) > 0$ if $g_{N-1}(b)[(\gamma - 1)g_{N-1}(b) + h'_{N-1}(b)] - h_{N-1}(b)g'_{N-1}(b) > 0$, or

$$\begin{aligned}\gamma[1 + (\gamma - 1)F(b)] &> (N - 1)f(b) \times \\ &\left[(\gamma - 1) \frac{\int_0^b [1 - F(b) + \gamma F(x)]^{N-1} dx}{[1 + (\gamma - 1)F(b)]^{N-1}} + \frac{\int_0^b [1 - F(b) + \gamma F(x)]^{N-2} dx}{[1 + (\gamma - 1)F(b)]^{N-2}} \right], \\ &= (N - 1)f(b) \left[(\gamma - 1) \frac{h_{N-1}(b)}{g_{N-1}(b)} + \frac{h_{N-2}(b)}{g_{N-2}(b)} \right].\end{aligned}$$

We can show that the ratio $h_N(b)/g_N(b)$ is decreasing in N implying $h_{N-2}(b)/g_{N-2}(b) \geq h_{N-1}(b)/g_{N-1}(b)$ for all $N \geq 2$. This intuition is illustrated in Figure (2) for a sample distribution. It can be seen that $h_N(b)/g_N(b)$ decreases as N is increased.

This implies that $\Psi'(b) > 0$ if (write substituting)

$$\gamma[1 + (\gamma - 1)F(b)] > \gamma f(b)(N - 1) \frac{h_{N-2}(b)}{g_{N-2}(b)},$$

$$\text{or } \gamma > 1 + \frac{1}{F(b)} \left[f(b)(N - 1) \frac{h_{N-2}(b)}{g_{N-2}(b)} - 1 \right]$$

If the rate of decay of the *ctr* with respect to position (γ) is high enough, then there exists a unique b^* that satisfies the optimality condition. For some common distributions like the Weibull, Gamma and Log-Normal we numerically find that $\Psi(b)$ is always increasing in b and there exists a unique bid for every keyword k that satisfies the optimality condition. This is illustrated in Figure (3) for some sample parameters.

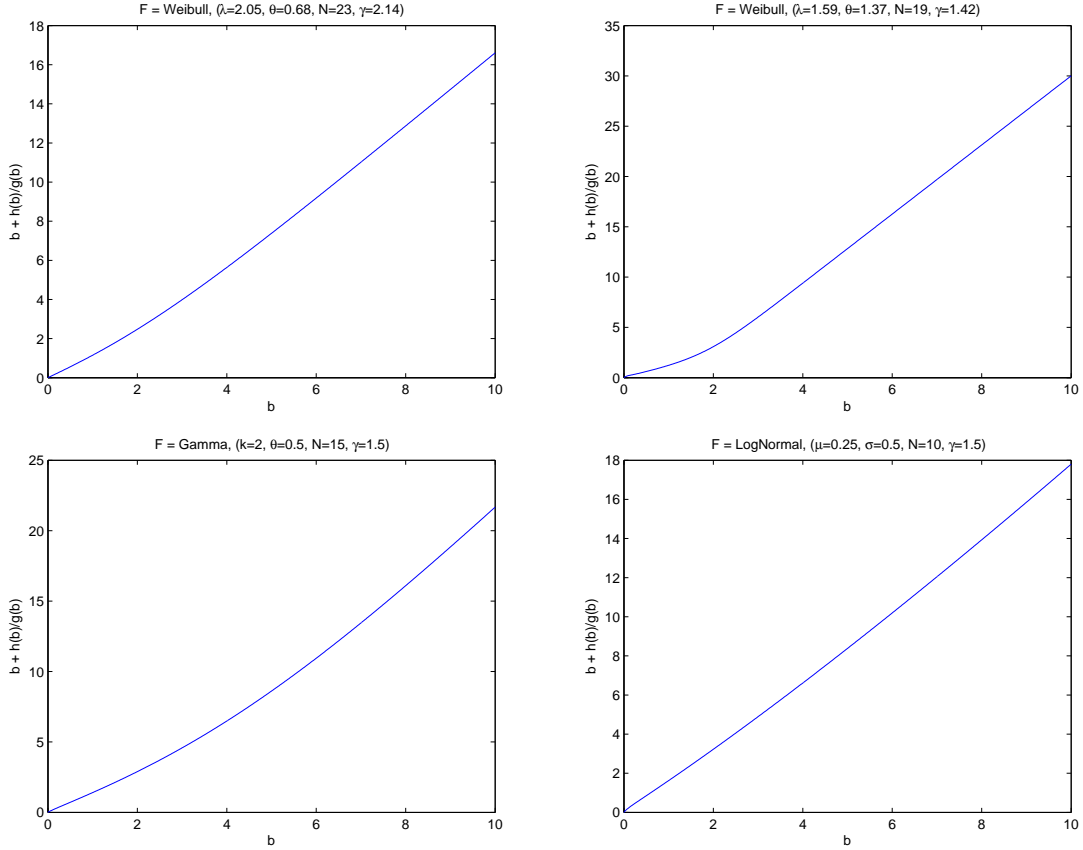


Figure 3: $\Psi(b)$ for various distributions.

Appendix A2: GMM estimation in the presence of three or more unique bids

As pointed out earlier, there are 4 parameters α, γ, λ and θ in this model and 3 moment conditions per bid. Thus the system can be identified with data from just two unique bids. A typical dataset however contains several distinct bids and the system of equations is over-identified. In order to use all the available data, we can use a generalized method of moments (GMM) procedure to estimate the parameters of the model.

Acknowledgment

The authors thanks Vadim Cherepanov for valuable assistance with prior versions of this work and Peter Fader for several helpful comments and suggestions. The paper is an extended version of the conference papers from the ACM Conference on Electronic Commerce (EC'08) and the Conference on Information Systems and Technology (CIST'08), and we thank their anonymous reviewers for several helpful suggestions. This research was funded by the Mack Center for Technological Innovation.

References

- Abhishek, V., Hosanagar, K. and Fader, P., Identifying and Resolving the Aggregation Bias in Sponsored Search Data, SSRN Working Paper, 2009.
- Agarwal, A., Hosanagar, K. and Smith, M. Location, Location and Location: An analysis of Position and Profitability in Sponsored Search, Working Paper, 2008.
- Aggarwal, G., Goel, A. and Motwani, R. Truthful Auctions for Pricing Search Keywords. Proceedings of the 7th ACM conference on Electronic commerce, Ann Arbor, Michigan, 2006.
- Ali, K. and Scarr, M. Robust methodologies for modeling web click distributions. In Proceedings of the 16th international Conference on World Wide Web (WWW'07), Banff, Canada, May 2007: 511-520.
- Asdemir, K. Bidding Patterns in Search Engine Auctions, Working Paper, University of Alberta, 2006.

- Borgs, C., Chayes, J., Immorlica, N., Jain, K., Etesami, O., and Mahdian, M. Dynamics of bid optimization in online advertisement auctions. In Proceedings of the 16th international Conference on World Wide Web (WWW'07), Banff, Canada, May 2007, 531-540.
- Craswell N., Zoeter, O., Taylor, M., and Ramsey, B. An experimental comparison of click position-bias models. Proceedings of WSDM'08, 2008.
- Edelman, B., Ostrovsky, M. and Schwarz, M. Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords. American Economic Review, 97(1), March 2007, pp.242-259.
- Edelman, B. and Ostrovsky, M. Strategic Bidder Behavior in Sponsored Search Auctions. Decision Support Systems, v.43(1), February 2007, pp.192-198.
- Feldman J., Muthukrishnan S., Pal M., and Stein C. Budget optimization in search-based advertising auctions. In Proceedings of the 8th ACM Conference on Electronic Commerce, 2007.
- Feng, F., Bhargava, H., and Pennock, D. 2007. Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms. Informs Journal on Computing, 19(1),137-148.
- Ghose, A. and S. Yang. 2008. Analyzing Search Engine Advertising: Firm Behavior and Cross-Selling in Electronic Markets, Proceedings of the World Wide Web Conference (WWW '08), Beijing, May 2008.
- Hansen, L. P., Heaton, J. and Yaron A. Finite-sample Properties of Some Alternative GMM Estimators. Journal of Business and Economic Statistics, 14(3):262-280, 1996.
- Liu, D., J. Chen and A.B. Whinston. Ex-Ante Information and Design of Keyword Auctions. Information Systems Research, Forthcoming, 2008.
- Kumar, S., M. Dawande and V. Mookerjee. Optimal Scheduling and Placement of Internet Banner Advertisements. IEEE Transactions on Knowledge and Data Engineering, Volume 19, Number 11, November 2007, pp. 1571-1584.
- Muthukrishnan, S., Pal, M., and Svitkina, Z. Stochastic models for budget optimization

in search-based. Proceedings of the 16th international Conference on World Wide Web (WWW'07), Banff, Canada, May, 2007.

Rusmevichientong, P. and Williamson, D. An Adaptive Algorithm for Selecting Profitable Keywords for Search Based Advertising Services. Cornell University working paper, 2006.

Rutz, O. and Bucklin, R. A Model of individual Keyword Performance in Paid Search Advertising. Working Paper, 2008.

Varian, H. Position Auctions. International Journal of Industrial Organization, 25, 2007.

Weber, T. A. and Z. E. Zheng. A model of search intermediaries and paid referrals. Information Systems Research, Vol. 18, No. 4, December 2007, pp. 414-436.

Zhang, M. and Feng, J. Price Cycles in Online Advertising Auctions. Proceedings of the 26th International Conference on Information Systems (ICIS), Dec. 2005, Las Vegas, NV.

Table 6: **Summary of notation**

k	Variable that indexes keywords
S_k	Random variable denoting number of searches for keyword k
μ_k	Expected number of search for keyword k ($\mathbb{E}[S_k]$)
(s)	Superscript to denote s^{th} search
b_k	Bid for keyword k
$pos_k^{(s)}$	Position for keyword k in s^{th} search. $pos_k^{(s)} = 0$ denotes the top position.
$\delta_k^{(s)}$	Indicator variable for click on s^{th} search. $\Pr \left\{ \delta_k^{(s)} = 1 pos_k^{(s)} = i \right\} = \frac{\alpha_k}{(\gamma_k)^i}$
w_k	Random variable indicating value of a click
$v_k^{(s)}$	Value of the s^{th} search ($v_k^{(s)} = \delta_k^{(s)} w_k$)
$\underline{b}_k^{(s)}$	The bid of the next advertiser
$c_k^{(s)}$	The cost of the s^{th} search ($c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$)
N_k	Number of competitors
$F_k(\cdot)$	Distribution of bids of competitors
D	Advertiser's budget