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

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 validate our approach, we estimate the parameters of the model using data from an advertiser's sponsored search campaign and use the bids proposed by the model in a field experiment. The results of the field implementation show that the proposed bidding technique is very effective in practice. We extend our model to account for interactions between keywords, in the form of positive spillovers from generic keywords into branded keywords. The spillovers are estimated using a dynamic linear model framework and used to jointly optimize the bids of the keywords using an approximate dynamic programming approach. Accounting for the interaction between keywords leads to an additional improvement in the campaign performance.

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

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 in contrast 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 (CTR) and conversion rates. Search queries entered by users convey significant information about users current need and context which allow search engines to better target ads to users than is possible in other forms of online advertising. Further, search engine users, unlike users on another websites, 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 meat seller may submit the following set of two tuples {(pork chop, $2), (fillet mignon, $5), (steak deals, $3),...} 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, it is not uncommon to have 100 or more advertisers bidding for the same keyword. 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 a large

1

number of 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 decreases as the rank of an ad increases (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 et al., 2006, Edelman et al., 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 "fillet mignon" then it may leave a very limited portion of the budget for another keyword. The performance of the keywords may also be interdependent, wherein clicks for one keyword may help generate more searches and clicks for another. 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 "fillet mignon" on any given day is stochastic and is a function of the weather, special events 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 with interdependent performance 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.

We propose two bidding policies in our paper. The first policy ignores the interaction between keywords and is referred to as the "myopic" policy in this paper. We extend this bidding policy to incorporate interaction between keywords, and refer to this policy as the "forward-looking" policy since it entails decision making over several time horizons. Depending on the advertiser's intent, level of sophistication and nature of the products being advertised, the advertiser might choose the

myopic or the forward-looking policy. This paper makes three main contributions. The first contribution is 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, as demonstrated in our field implementation. 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 that it is the first paper on bidding in sponsored search to incorporate the interdependence between keywords into a multi-period bidding problem. The interdependence in keyword performance, commonly referred to as spillovers, is a well-documented feature of sponsored search (Rutz and Bucklin, 2011) but has not been considered in the bidding literature.

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 for the myopic policy and discuss how it may be used to compute the optimal bids. In Section 4, we describe the dataset used for the analysis presented in this paper. In Section 5, we present the empirical analysis where we estimate the parameters of our model and run a field experiment with the bids suggested by the myopic policy. We compare the optimal bids computed by our model with those used by the firm and present results from a field implementation of the bids. We extend the myopic policy in Section 6 to incorporate interdependence between keywords. Finally, we discuss some limitations of our work in Section 7 and conclude in Section 8.

2   Literature review

In this section, 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 PPC auctions in which they charge advertisers whenever a consumer clicks on an ad.[1] A primary area of focus in sponsored search research has 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 et al. (2006) 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. 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. In a related paper, Edelman et al. (2007) demonstrate that the commonly used 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. Aggarwal et al. (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. Mehta et al. (2007) solve the problem of matching ad slots to advertisers using a generalization of the online bipartite matching problem. Given advertisers' bids and budget constraints, Mehta et al. (2007) provide a deterministic algorithm that achieves a competitive ratio of $1 - 1/e$ for this problem.[2] Mahdian et al. (2007) extend this work and provide a solution which is nearly optimal when the frequencies of keywords are accurately known and provides a good competitive ratio even when these estimates are completely inaccurate. Aggarwal and Hartline (2006), on the other hand, model this problem as a knapsack auction. However, they consider only truthful mechanism designs and analyze various pricing schemes and the payoffs under each of these pricing schemes. Most of the above referenced papers focus on the search engine's problem and analyze how different mechanisms affect the search engine's revenues.

---

[1] Other 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.

[2] Competitive ratio is the measure for comparing online algorithms to offline algorithms where all the information is known apriori.

Consumer behavior in sponsored search

The sponsored search environment presents rich data on consumer behavior. Modeling user's 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, Feng et al., 2006). Ali and Scarr (2007) compare several distributions to predict *click-through* rates and suggest that Pareto-Zipf distribution is the most appropriate for explaining CTR as a function of position. Feng et al. (2006) alternately assume an exponential decay in CTR with position and demonstrate that the model fits observed data well. Several other papers build richer models of consumer behavior incorporating the effect that ad attributes have on *click-through* and conversion (Ghose and Yang, 2009, Yang and Ghose, 2010, Rutz et al., 2012, Agarwal et al., 2011). Rutz and Bucklin (2011) propose a model that measures the interaction between keywords and show that there are significant positive spillovers from generic keywords to branded keywords in consumer search.

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 the 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 et al. (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. The first is that the literature on sponsored search mechanism design has established that GSP auctions are not incentive compatible. This feature combined

with the advertiser's budget constraint suggests a need to develop bidding policies. Secondly, the empirical work in sponsored search provides a variety of useful models that can be applied towards modeling consumer click behavior and the 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 tend to focus on deterministic settings or single slot auctions, both of which are restrictive assumptions in the sponsored search context. None of these papers account for any interdependence in keyword performance. 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. We also extend our basic model to incorporate interaction between keywords.

3 Analytical Model

Advertisers usually maintain a portfolio of thousands of keywords for a particular search engine. They submit bids for each keyword on a regular basis during a billing cycle. 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.[3] Although we consider the advertiser's problem of optimizing the bid for a particular search engine in this paper, our approach can be extended to multiple search engines by treating each keyword-search engine pair as a unique keyword.

Ads placed in response to consumer search queries can play two roles for advertisers. They can help generate purchases. Or they can help build awareness, which may translate into purchases in later periods. Consumers often start their search process with generic search terms e.g. "fillet mignon". Bidding on these generic keywords might help the advertiser generate brand-specific (or retailer-specific) exposure. This in turn might enhance the awareness of a particular brand and can lead to increased branded search activity ("spillover"). There is evidence of spillovers from generic to branded keywords in sponsored search ads (Rutz and Bucklin, 2011).

In the paper, we propose two policies that can be used an advertisers to optimize the bids for

---

[3] A 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.

keywords in an advertising campaign. We begin by proposing a "myopic" bidding policy that solves the one-shot decision problem of the advertiser which does not factor in indirect benefits from keywords such as awareness. We relax this assumption in Section 6 and incorporate interactions between keywords. The bidding policy that incorporates interactions between keywords is referred to as the "forward-looking" policy. The forward-looking policy solves the advertising problem in a multi-period context. The motivation for developing two bidding policies is twofold. First, the myopic policy is easier to implement and also relevant in the context of commoditized products where branding is not very relevant. The complexity of a forward-looking policy may be unnecessary for many advertisers. In addition, the forward-looking policy builds on the myopic policy and it is useful for the purpose of exposition to outline the myopic policy first. In this section, we ignore spillovers between keywords and assume that the keywords are independent.

3.1 Notation and Setup

We introduce our notation and the general framework used to study the advertiser's decision problem.

During a given time period (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 $\mu_k = \mathbb{E}[S_k]$. 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 day. 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 sth 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: $\delta_k^{(s)} = I\left(click_k^{(s)}\right)$.

The advertiser's value from a click is denoted by an independent random variable w_k. We assume that the precise value from a click is not known a priori but that it's expected value $\mathbb{E}[w_k]$ is known and equals Ew_k. In Section 5, we discuss how Ew_k is estimated from historical data. $v_k^{(s)}$ denotes the advertiser's value from the sth impression. $v_k^{(s)} = \delta_k^{(s)} w_k$, i.e. it equals w_k if the user clicks on the ad or 0 otherwise. 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 $c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$.[4] Because consumers do not know the bids placed by advertisers, it seems reasonable

---

[4]The discussion assumes that ads are ordered by bid and that the advertiser pays the bid of the next advertiser.

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 $\left(\underline{b}_k^{(i)}, \delta_k^{(i)}\right)$ 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. We assume that N_k is known to the advertiser. It can be observed, for example, by submitting sample queries to the search engine and observing the number of ads displayed. We note that the number of competitors may in reality vary a bit from one impression to another due to advertiser budget constraints, but we do not observe significant variation in this to warrant a random treatment for N_k.[5] The bids of the competitors cannot be directly observed because the auction is a sealed bid auction. The key assumption we make is that the competitors place their bids according to some distribution $F_k(.)$ and this does not change during the estimation period. The bids of competing advertisers are based on two factors - their intrinsic valuations for a click and their competitive responses in the GSP auction. We assume that there is an underlying valuation distribution (for clicks) which when combined with the advertisers' bidding strategies gives rise to the bid distribution $F_k(.).$[6] Finally, D denotes the advertiser's budget in a given time period of interest. Table 1 summarizes our notation.

3.2 Model Formulation

The advertiser faces the following decision problem:

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

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

---

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.

[5]The coefficient of intra-day variation in $N_k = 0.031$ and the coefficient of inter-day variation in $N_k = 0.102$. A model that incorporates the randomness in N_k is available from the authors upon request.

[6]The proposed bids might change $F_k(.)$, but for identification purposes we assume that this competitive reaction is minimal in the short-term. Later in the paper, we discuss how the competitive reaction can be factored in by re-estimating parameters periodically and updating bids.

Table 1: 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.
w_k	Random variable indicating value of a click
Ew_k	Expected value of a click on keyword k ($\mathbb{E}[w_k]$)
$v_k^{(s)}$	Value of the sth 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 sth search ($c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$)
N_k	Number of competitors
$F_k(.)$	Distribution of bids of competitors
D	Advertiser's budget

the form $\max_{\{b_k\}} \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} (v_k^{(s)} - \lambda c_k^{(s)})|b_k\right]$. 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

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

where λ is the Lagrange multiplier. The optimality condition states that at 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 $\varepsilon \frac{d}{db_{k'}}\mathbb{E}\sum_{s=1}^{S_{k'}}\left[c_{k'}^{(s)}|b_{k'}\right]$ and, hence, we may increase the bid for another keyphrase k by $\varepsilon \frac{d}{db_{k'}}\mathbb{E}\sum_{s=1}^{S_{k'}}\left[c_{k'}^{(s)}|b_{k'}\right] / \frac{d}{db_k}E\sum_{s=1}^{S_k}\left[c_k^{(s)}|b_k\right]$. In this case the expected increase in profits will be

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

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:

$$\forall k : \frac{d}{db_k} \mathbb{E}\left[v_k | b_k\right] = \lambda \frac{d}{db_k} \mathbb{E}\left[c_k | b_k\right].$$ (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 needs to compute $\mathbb{E}\left[v_k | b_k\right]$ and $\mathbb{E}\left[c_k | b_k\right]$ 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 during the day. 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 $\mathbb{E}\left[v | b\right]$, 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. Given that the competing advertisers' bids are drawn from $F(.)$, the probability of being at position i conditional on a bid b is

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

The position is determined by a Bernoulli process, where the probability that a competitor bids more than b and is placed higher is equal to $1 - F(b)$. Recollect that the positions start from 0, i.e., the topmost ad has position $pos = 0$, and 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

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

where α and γ are keyword specific constants. α represents the overall attractiveness of the ad and γ captures the impact of position on clicks. This functional form does not explicitly consider a number of other factors, e.g. number of words in the keyword, whether the advertiser appears in the organic results or not, presence of dominant competitors etc., that might affect CTR (Yao and Mela, 2011, Agarwal et al., 2011, Katona and Sarvary, 2010, Ghose and Yang, 2009). It focuses only on the impact of position on CTR because ad position is the primary mechanism through which bid

10

impacts CTR. However, the parameter α captures the effect that ad/keyword-level attributes like the number of words in the keyword etc. have on the overall attractiveness of the ad. γ on the other hand captures the change in CTR with respect to position, all other factors held constant, which is consistent with prior research (Katona and Sarvary, 2010, Ghose and Yang, 2009).[7] This function also assumes that consumer behavior is i.i.d and ignores heterogeneity across consumers. We use this assumption not only for model tractability but also because search engines do not provide user-level data on impressions and clicks. Several papers that focus on keyword-level models, also assume i.i.d. consumer behavior (Agarwal et al., 2011, Yang and Ghose, 2010, Ghose and Yang, 2009). Given that the consumers click in the aforementioned manner, the probability of a click conditional on the bid b is given by

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

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

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

It follows from Proposition 1 that

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

We now derive an expression for $\mathbb{E}\left[c|b\right]$. 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

---

[7]However, if the presence of a dominant competitor introduces discontinuities in how position affects CTR (e.g., CTR depends on whether the advertiser is above or below the dominant competitor), the functional form fails to capture the same.

the bid and the position is given by

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

Applying,

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

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

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

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

$$
\begin{aligned}
\mathbb{E}\left[c|b\right] &= \mathbb{E}\left[\delta\underline{b}|b\right] \tag{12}\\
&= \mathbb{E}\left[\underline{b}|b, \delta = 1\right]\Pr\left\{\delta = 1|b\right\}\\
&= \alpha\gamma^{-N}\left(b[1 + (\gamma - 1)F(b)]^{N} - \int_{0}^{b}[1 - F(b) + \gamma F(\underline{b})]^{N}d\underline{b}\right).
\end{aligned}
$$

Using Proposition 2 we can derive that

$$\frac{d\mathbb{E}[c|b]}{db} = \alpha N\gamma^{-N}f(b)\left((\gamma - 1)b[1 + (\gamma - 1)F(b)]^{N-1} + \int_{0}^{b}[1 - F(b) + \gamma F(\underline{b})]^{N-1}d\underline{b}\right). \tag{13}$$

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

$$
\begin{aligned}
\frac{d\mathbb{E}[v|b]}{db} &= \lambda\frac{d\mathbb{E}[c|b]}{db},\\
\frac{1}{\lambda} &= \frac{1}{Ew}\left(b + \frac{\int_{0}^{b}[1 - F(b) + \gamma F(\underline{b})]^{N-1}d\underline{b}}{(\gamma - 1)[1 + (\gamma - 1)F(b)]^{N-1}}\right).
\end{aligned}
$$

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

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

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 distributions 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 per-click (RPC)), 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 5. The optimal bids should satisfy equation (14) and the budget constraint,

$$\sum_k \mu_k \mathbb{E}\left[c_k\right] = D.$$

These conditions are sufficient to compute bids. The budget constraint can be rewritten as

$$\sum_k \mu_k \alpha_k \gamma_k^{-N_k}\left(b_k[1 + (\gamma_k - 1)F_k(b_k)]^{N_k} - \int_0^{b_k}[1 - F_k(b_k) + \gamma_k F_k(\underline{b})]^{N_k}d\underline{b}\right) = D. \qquad (15)$$

For a given *const* in Equation (15), we compute the bid that satisfies the equation for every keyword. Then we use Equation (15) 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 online and through mail-order catalogs. 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 that the firm uses to advertise on Google. These keywords are a part of the North-East campaign of the advertiser that targets consumers in the northeastern parts of the Unites States. The daily record for each keyword has the following fields,

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

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 the type of data available to advertisers in sponsored search. Advertisers only get summary reports from search engines and do not usually have information on clicks and position for each individual ad impression. We present the summary statistics at the keyword-level for a three-month period (March 01-May 31, 2011) prior to the field implementation in Table 2.

The mean average bid across all keywords during this period is $1.18 and the minimum and maximum average bids for any keyword during this period is 35¢ and $10, respectively. We also observe that the mean average revenue *per-click* (RPC) is $4.33 where as the mean average CPC is 75¢, however there is a huge variation in the profitability across the keywords as indicated by the large standard deviation in the average RPC. These 247 keywords belong to 29 unique product categories which span frozen meats, sea foods and desserts. A comprehensive list of these product categories appears in Table 3. We *randomly* divided these 29 product categories into three distinct treatment groups. The random assignment ensures that the product categories in the three groups

14

Table 2: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Avg 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
Avg RPC	4.33	14.30	0.00	158.96
Days Shown	84.72	16.51	54.00	92.00

are well match as shown in the treatment effect literature (Angelucci and Giorgi, 2009). The bids for the first group continued to be controlled by the firm. This group forms the control group for our experiment. The other two groups represent the two treated groups and their bids are determined by the myopic bidding policy (Group I) outlined in Section 3 and the forward looking policy (Group II) that we outline in Section 6. The control group is used to account for any time trends that might enter the analysis due to seasonality in retail, search engine design changes and other such factors. The three groups are fairly well matched in terms of impressions, clicks, cost and revenues of their keywords. Summary statistics for the three groups are presented in Table 4.

Table 3: 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
Fillet Mignon	Orange Roughy	

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 (summary statistics for this period is in Table 2). During this period, there were 1.36 million impressions of the ads for the 247 keywords and they received 11,651 clicks in total. The total weekly cost of these ads was $964 and the weekly gross

Table 4: Summary for the three groups of keywords.

	Control	Group I	Group II
Products Categories	8	10	11
Keywords	55	89	103
Impressions	7474.7	5336.1	4335.8
Clicks	66.6	52.2	30.6
Avg Bid	1.32	1.27	1.14
Avg CPC	0.84	0.91	0.74
Avg Pos	3.22	3.60	2.55
CTR	0.03	0.03	0.04
Avg RPC	4.80	5.36	4.35
Avg RPC/Avg CPC	5.72	5.89	5.87
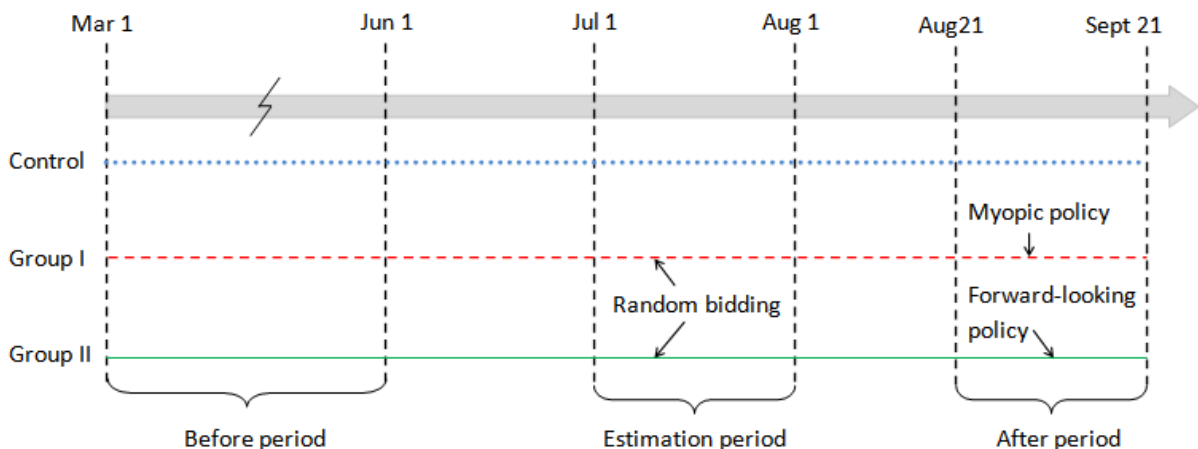
Figure 1: Illustration of the Timeline for the various data collection periods.

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 (μ) for each keyword.

The second period spans 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 significant increase in online activity during this month due to Father's Day. The summary statistics for this period is presented in Table 5. During the estimation period, 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. The variation in position causes changes in the CTR and CPC and helps the identification of the parameters of our model. A

Table 5: Summary Statistics for the Random Bidding Period

	Mean	Standard Deviation	Minimum	Maximum
Bid	1.01	0.93	0.05	3.00
Avg CPC	0.77	0.58	0.00	2.89
Avg Pos	2.96	1.87	1.00	11.08
Impressions	2100.45	5115.73	1.00	43498.00
Clicks	19.00	35.78	0.00	277.00
CTR	0.03	0.06	0.00	0.50
Cost	18.29	38.51	0.00	278.97
Revenue	23.46	75.65	0.00	586.92
Gross Profit	5.18	67.51	-150.43	541.22
RPC	1.28	4.31	0.00	29.32

comparison of the variation in the position, CTR and CPC in the "before" and "estimation" periods (at a keyword level) is shown in Table 6. The exact identification strategy is discussed in Section 5. We also observe that there is decrease in the profitability of the campaign during this period as the bids for keywords in Group I and Group II are chosen randomly.

Table 6: Variation in keyword attributes in the "before" and "estimation" periods.

	"before"	"estimation"
Num. Bids	1.12	4.00
S.D. Pos	0.87	3.19
S.D. CTR	0.01	0.03
S.D. CPC	0.14	0.32

Finally, optimal bids are computed based on estimated parameters and deployed by the firm between August 21 and September 21, 2011. Data from the after period is used to assess the effectiveness of the bidding policies proposed in this paper. 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

We now 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 ($ctr_t = cl_t/i_t$) given the bid for each keyword. These moments are as follows:

$$\mathbb{E}\left[pos_t | b_t\right] = N_t \left(1 - F\left(b_t\right)\right), \tag{16}$$

$$\mathbb{E}\left[\underline{b}_t | b_t, \delta_t = 1\right] = \int_{x < b_t} x \, d\left(\frac{1 - F\left(b_t\right) + \gamma F\left(x\right)}{1 - \left(1 - \gamma\right) F\left(b_t\right)}\right)^{N_t}, \tag{17}$$

$$\mathbb{E}\left[\delta_t | b_t\right] = \alpha \gamma^{-N_t} \left(1 - \left(1 - \gamma\right) F\left(b_t\right)\right)^{N_t}. \tag{18}$$

The observed moments can be expressed in terms of the analytical moments as follows:

$$avgpos_t = \mathbb{E}\left[pos_t | b_t\right] + \xi_{1t},$$

$$avgcpc_t = \mathbb{E}\left[\underline{b}_t | b_t, \delta_t = 1\right] + \xi_{2t},$$

$$ctr_t = \mathbb{E}\left[\delta_t | b_t\right] + \xi_{3t},$$

where $\boldsymbol{\xi}_t = \left(\xi_{1t}, \xi_{2t}, \xi_{3t}\right)'$ are the random shocks. As the dataset contains only daily aggregates, we cannot directly estimate the distribution function $F(.)$ using nonparametric approaches since we have very few bids for each keyword. We therefore use a parametric form for $F(.)$, 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 $F(.)$ 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 et al., 2011) shows that the

Weibull distribution is reasonably good for modeling the bids.[8] Note that we are not assuming that the distribution of bids for keywords is the same across the two datasets, rather the bids are from the same family (Weibull) and the parameters can vary across keywords. The Weibull distribution has the following cumulative distribution function

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

It is defined by two parameters θ and λ. Therefore, we have four unknown parameters for any keyword $(\lambda, \theta, \alpha, \gamma)$ and 3 moment conditions for every unique bid.

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 a 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. Hansen et al. (1996) and Wooldridge (2001) suggest that the optimal weighting matrix is given by $\mathbb{E}[\boldsymbol{\xi}(\beta)'\boldsymbol{\xi}(\beta)]^{-1}$. As we do not know $\mathbb{E}[\boldsymbol{\xi}(\beta)'\boldsymbol{\xi}(\beta)]$, an iterative-GMM estimator is used (see Hansen et al., 1996) wherein 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 computed by taking the total revenues from the keyword in the "before" period and dividing it by the total number of clicks for that keyword in the same period. It should be kept in mind that we only consider sales through the online channel in this analysis. The advertiser attributes revenues from a purchase to the keyword that generated the session in which the purchase was made. One drawback with the approach is that it does not account for indirect benefits such as awareness. As stated above, we address that later in the paper.

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 data from the "before" period cannot be used to estimate the parameters of this model - (i) insufficient variation in bids, and (ii) potential

---

[8]The authors test several distributions such as Normal, Log-Normal, Gamma, Exponential and Logit but the Weibull distribution fits their data the best. Note however that our framework is flexible enough and other distribution can be easily accommodated.

endogeneity in advertiser's bids.

Limited Variation in Bids

In typical SSA campaigns advertisers change their bids infrequently, sometimes 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 are 1.12. Because our model is under-identified 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.

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 $\mathbb{E}[b\boldsymbol{\xi}] = 0$ or the bids and the random shock are independent of each other. However, the firm might increase the bid for a particular keyword if there is a random increase in demand, e.g. on a sunny weekend. These random shocks are observed by the advertisers but we as researchers are not aware of them. Since the firm is bidding strategically, it is very likely that the bids for a particular keyword are correlated with these random shock in the before period. We address this endogeneity issue by using random bids in the estimation period. This randomization of bids ensures that they are independent of the random shocks.

The random bids in the "estimation" period address the limited variations in bids and the endogeneity of bids. However, we still need to make several parametric assumptions in the model, e.g. exponential decay of the CTR curve and Weibull distribution for F(.) because there is insufficient keyword-level data to perform non-parametric estimation. We also require that the distribution $F(.)$ does not change during the estimation period as competitive response to the random bids being set during this period. This seems like a reasonable assumption given the muted short-term competitive response in sponsored search as pointed out by Rutz and Bucklin (2011). We revisit this assumption in Section 7.2.

5.3 Estimation Details and Results

In order to estimate the parameters, a nonlinear solver is used in our implementation.[9] The parameter estimates for a few representative keywords are shown in Table 7. $\overline{N}$ represents the mean number of daily competitive 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 2. A complete table is available from the authors upon request.

Table 7: Parameter estimates for a sample subset of keywords.

keyword	λ	θ	α	γ	$\overline{N}$	$Ew(\$)$
beef sirloin steak	1.7651	0.5351	0.0266	2.1237	9.5	0.00
	(0.4927)	(0.2821)	(0.0115)	(0.2742)		
Steak Burger	0.6697	2.1944	0.0069	1.2915	5.1	1.53
	(0.4035)	(0.3731)	(0.0008)	(0.0902)		
cheesecakes	0.9736	1.3265	0.0004	1.6091	7.0	1.08
	(0.2064)	(0.4270)	(0.0000)	(0.2405)		
Porterhouse Steak	1.1413	0.8639	0.0085	1.1661	4.6	0.43
	(0.5118)	(0.1821)	(0.0015)	(0.3711)		
smoke salmon	1.3414	1.1752	0.0073	1.0255	10.1	6.62
	(0.5429)	(0.4520)	(0.0012)	(0.3989)		
corned beef	1.5368	0.7492	0.0018	1.0175	10.7	3.80
	(0.8126)	(0.5781)	(0.0004)	(0.7045)		
hot dog order	1.0769	1.0869	0.0101	1.6486	7.3	3.00
	(0.4410)	(0.7503)	(0.0036)	(0.2446)		
birthday gifts	1.1756	0.8420	0.0009	1.0659	40.2	5.74
	(0.6781)	(0.4176)	(0.0000)	(0.7850)		
birthday present	0.7524	1.3841	0.0122	1.0434	7.1	0.45
	(0.6721)	(0.4816)	(0.0057)	(0.9381)		
lobster bisque	1.311	1.0074	0.0145	1.9293	11.3	0.00
	(0.3928)	(0.5025)	(0.0037)	(0.4117)		

Although there is significant heterogeneity across keywords, the estimated parameter values are fairly typical in sponsored search. 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., 2006, Craswell et al., 2008). There is also considerable variation in the *expected revenue per-click* (Ew) and the bid distributions (λ, θ) across keywords.

---

[9]We use the Fletcher-Xu hybrid method provided as a part of the ClsSolve routine in TOMLAB.
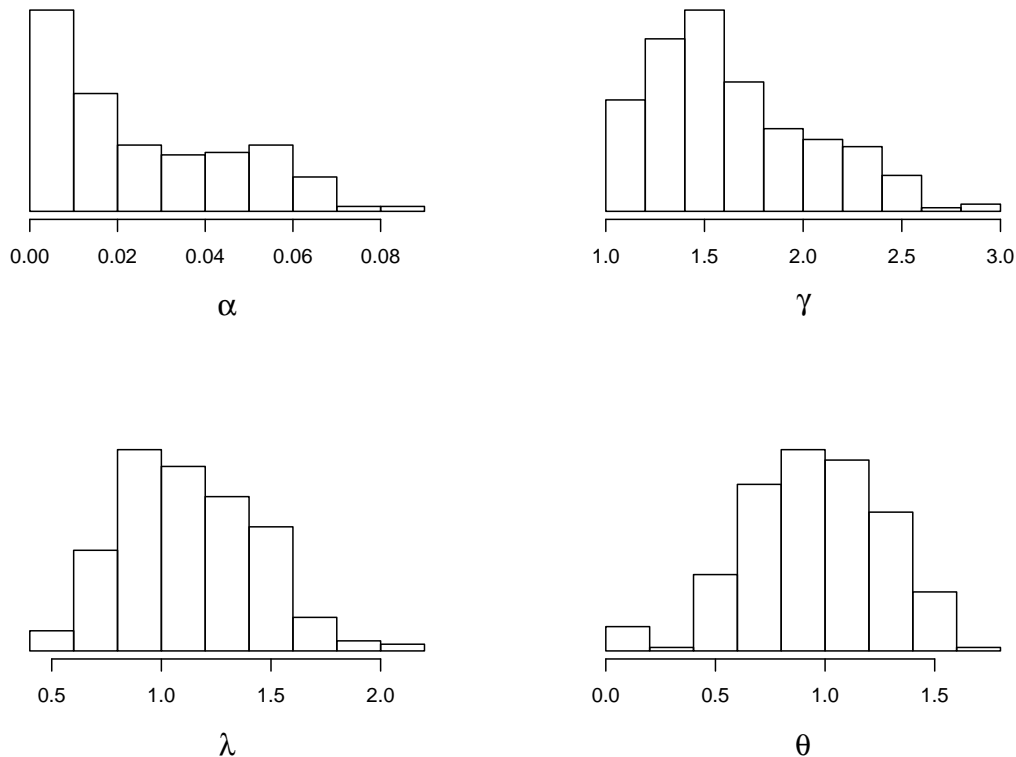
Figure 2: Distribution of estimated parameters across keywords.

5.4 Field Implementation

Once we estimate parameters α, γ, λ and θ for all keywords, we estimate the optimal bids for these keywords. In this section we focus on the myopic policy outlined in Section 3 and discuss the results of the field implementation for keywords in Group I.

For the keywords in Group I, we use a daily budget $D = \$72.00$ based on the mean weekly spend of around \$500 during the 3 month "before" period. The bids are recomputed after two weeks when we re-estimate the parameters $(\alpha, \gamma, \lambda, \theta)$ using newly available data. The bids are recomputed to account for changes in competitor bids and consumer click behavior. However, the bids do not change much during this re-computation. Bids for a sample of keywords are below in Table 8.

The rationale for these bids can be inferred from the parameters listed in Table 7. Consider, for example, bids for keywords "smoke salmon", "hot dogs order" and "birthday gifts". Our algorithm suggests increasing their bids. From Table 7, 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. Interestingly,

Table 8: 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 keyword "birthday gifts" has a very high 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. This policy also decreases the bids for keywords like "beef sirloin steak", "lobster bisque" and "birthday present". The bids for "beef sirloin steak" and "lobster bisque" are decreased because they are not profitable. The bid for "birthday present" is decreased because (i) it is not very profitable and (ii) it is possible to get a similar number of clicks at a lower position (low γ) for much cheaper.

The suggested bids were deployed in the field by the advertiser for a period of 4 weeks. During the 12 weeks in the "before" period, the firm spent a total of $5937.58 on the keywords in Group I 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 period, the total cost (revenues) associated with the control keywords was $4701.52 ($9776.2) and $1667.54 ($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

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

The performance of the advertising campaign increases by 75.38% on a DiD basis indicating that the myopic policy outperforms the advertisers bidding policy. In the next section we discuss some of the drivers of this performance gain. Surprisingly, we notice that there is an absolute decrease

in the ROI in the campaign compared to the "before" time period and this decrease is particularly notable for the Control group. This is partly because of seasonality in meat sales. In addition, there were changes in the manner in which the search engine displayed search results. From July onwards, the search engine started highlighting the top ads by using a light pink background color, which resulted in an increase in the CTR of the top ads. [10] For the Control group we see an increase in the CTR from 0.89% to 1.4% even though there was no apparent change in the advertiser's bidding policy and for the keywords in Group II we see a change from 1.04% to 1.15%. We observed that this not only resulted in an increased CTR for the keywords, but also a decrease in their performance during this time. This change negatively affects the performance of our policy as some the underlying parameters that were used to compute the optimal bids have changed. However, the control group allows us to control for such changes in the search engine policy. Since the keywords in the Control Group and Group I are matched as a result of the randomization procedure, the change in the display scheme has the same effect (on average) on the keywords in both groups. Hence, the DiD approach eliminates the influence of the policy change and any other changes in the environment and measures the difference in performance between the myopic policy and the policy adopted by our partner advertiser. It should also be noted that, although the parameters of the model are recomputed during the experiment, the estimates did not reflect the changes in the CTR model. Since the data during the field experiment (1.5 week) were pooled with data from the estimation period (4 weeks), the changes during the experiment did not have a significant impact on the parameter estimates.

5.5   Analysis of the Field Experiment

In the previous section we presented the improvement that the myopic policy offers over the advertiser's policy. In this section, we discuss in further detail the factors that lead to the improvement in the campaign's performance. There are two main sources of improvement – Firstly, a comprehensive model that captures the effect of bid on position, position on CTR and eventually the bid on the value (v_k) and cost (c_k), helps us in improving the bids for each keywords. Secondly, since bids for the entire portfolio are determined jointly, the advertising budget can be distributed from less profitable keywords to relatively more profitable keywords (based on the aforementioned model). As both these approaches are concurrently applied to the portfolio, it is difficult to disentangle the

---

[10]Several analysts suggest that the pink background for the ads is indistinguishable from the page background and users mistake these ads for organic links.
http://www.plymarketing.com/ppc/6-reasons-googles-new-ad-layout-should-really-piss-you-off/

effect of these drivers on the campaign performance. However, we can demonstrate how both these decisions affect the bids and profitability of keywords in the campaign. The following table contains a list of sample keywords and their performance during the field experiment. For each keyword, the table reports the average revenue per click (RPC), original bid, new bid, average cost per click (CPC) and the percentage change in gross profits under the new bid.

Table 9: Changes in keyword performance

keyword	Avg RPC($)	Old bids($)	New bids($)	Avg CPC	ΔGross Profit
buy orange roughy	2.95	0.35	0.85	0.29	343.33
filet mignon	5.00	1.50	2.55	1.83	210.37
buy lobster online	3.92	1.91	2.35	1.86	148.19
smoked salmon lox	5.85	1.00	3.00	2.03	38.24
bbq Beef	3.83	1.52	1.00	0.96	42.56
lobster delivery	1.62	2.25	0.95	0.87	-1.310
precooked bacon	1.57	1.50	1.05	0.68	-38.73
birthday gifts	5.74	0.96	1.75	1.73	-204.92

From the sample presented in Table 9 we observe that the advertiser was initially placing low bids for keywords like "buy orange roughy" and "filet mignon" . These keywords are in fact very profitable (have a high RPC and a low CPC) as they do not face intense competition. Using the analytical model presented in Section 3, we can ascertain not only the consumer response parameter (CTR v/s position) but also the competitive landscape associated with a keyword, which are subsequently incorporated in the bidding process. E.g. since there is very little competition for "buy orange roughy", the myopic policy recommends increasing the bid for this keyword. Even though the bid has been increased significantly, the CPC in the "after" period is considerably low (29¢), which further validates the low level of competition for this keyword. Interestingly, the policy reduces the bid for "bbq Beef". This is because the decay in the CTR with position is very low for this keyword ($\gamma = 1.0338$). Hence a lower bid can decrease the costs without affecting the revenues considerably. This complex interaction between the bid, the revenue and the cost associated with a keyword cannot be predicted in a modeless manner. Our approach explicitly captures the relationship between these variable and hence outperforms the original heuristics adopted by our advertiser.

We also observe that the bids for several moderately profitable keywords like "lobster delivery" and "precooked bacon" have been reduced in the after period. These keywords were profitable but the optimization reduces the bids for these keywords because of the budget constraint. A portion of the budget invested in these keywords can be diverted to other keywords that deliver higher profits. Consequently, their bids are lowered and they generate lower revenue and, interestingly, lower gross

profits during the field experiment as compared to the "before" period. It should be noted that this decrease in profits is more than offset by the investment in relatively more profitable keywords. In summary, the myopic policy focuses on relatively cheap and profitable keywords and reduces the bids of other keywords to maximize the profits from these keywords. Surprisingly, we see that the gross profits for "birthday gifts", a high performing keyword in the "before" period, decrease considerably during the field experiment despite the increase in bid. The myopic policy increases the bid for "birthday gifts" given its high RPC. However, as the CTR increased significantly during the field experiment (due to the changes by the search engine), the value per impression v_k dropped considerably and this keyword incurred a loss.

6   Incorporating Interdependence between Keywords

The preceding discussion assumes that keywords are independent of each other and the consumer click behavior is i.i.d.. In reality, consumers may search across several keywords before making a purchase decision and this might lead to interaction between keywords. For example, a consumer might begin his search with a generic keyword like "fillet mignon" but may eventually purchase using another keyword such as "Walmart fillet mignon". While searching for fillet mignon, he could have been exposed to ads from Walmart, causing Walmart to be part of his consideration set. Not accounting for such spillovers may cause the advertiser to undervalue "fillet mignon" and overvalue "Walmart fillet mignon". This example illustrates that there is value in accounting for these interactions while making bidding decisions. One way to capture this interaction is a full factorial design, where we consider spillovers for every possible subset of the portfolio of keywords and decide the optimal bids for keywords in this subset. However, the problem is NP hard and requires significant resources to assess the performance of each subset. In this paper we will focus on a specific kind of interaction proposed by Rutz and Bucklin (2011). We categorize the keywords into two groups – generic and branded – and explore how these two groups of keywords interact.

A generic keyword does not contain the brand name of the firm (e.g. "fillet mignon") whereas a branded keyword does (e.g. "Walmart fillet mignon"). Advertising on generic keywords can help create awareness about the brand/product which can then increase the likelihood that the brand is a part of the consumer's consideration set and, in turn, result in greater number of branded searches. Rutz and Bucklin (2011) show that there are considerable spillovers from generic to branded search activity in sponsored search. Methods which do not account for awareness might undervalue some

keywords. E.g., in our dataset, clicks on generic keywords are usually more expensive than on branded keywords (e.g., $0.88 v/s $0.45) and less profitable (e.g., $2.89 v/s $7.80). If we just look at the RPC and CPC of the keywords, it is more profitable to invest in branded keywords as compared to generic keywords. However, as pointed out earlier, bidding on expensive generic keywords might lead to future branded search and more clicks on the profitable branded keywords. Hence, the advertiser should incorporate this spillover effect while making his bidding decisions. In the following discussion we present a model that accounts for this dynamic interaction between the generic and branded keywords while computing optimal bids.

6.1   Measuring Interactions

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 effect on future search activity. We use the Nerlove-Arrow model (Rutz and Bucklin, 2011, Naik and Sawyer, 1998, Nerlove and Arrow, 1962) to capture the evolution of awareness

$$\frac{dA_t}{dt} = -(1 - \eta^A)A_t + \boldsymbol{\beta}\boldsymbol{X}_t, \tag{19}$$

where A_t refers to the awareness level at time t, $(1-\eta^A)$ measures the decay of awareness with time, $\boldsymbol{X}_t$ is a vector of covariates that capture the search activity at time t and $\boldsymbol{\beta}$ captures the extent to which different kinds of search activity affect the level of awareness. According to the Nerlove-Arrow model, brand awareness decays over time since consumers forget about a brand as time goes by. Search activity, on the other hand, reinforces brand awareness. This increased awareness, in turn, can lead to further branded search activity. We divide the keywords into two groups – G (generic) and B (branded) – and explore how search activity related to these keywords affects the level of awareness. The two search activities that we observe in our dataset are impressions and clicks for each keyword in the campaign. Prior results suggest that ad impressions do not have a significant impact on brand awareness but clicks on ads increase brand awareness (Rutz and Bucklin, 2011). This is because an ad impression does not guarantee that the ad is seen by the consumer and, further, mere exposure to an ad may not have an impact on the consumer unless the consumer pays sufficient attention to the ad (e.g., by clicking it). We incorporate this finding in our model and assume that generic and branded clicks may increase brand awareness (which is latent in our model and cannot be directly observed). An increase in this latent awareness can lead

to more search and hence more generic or branded impressions.[11] This interaction is demonstrated in Figure 3.
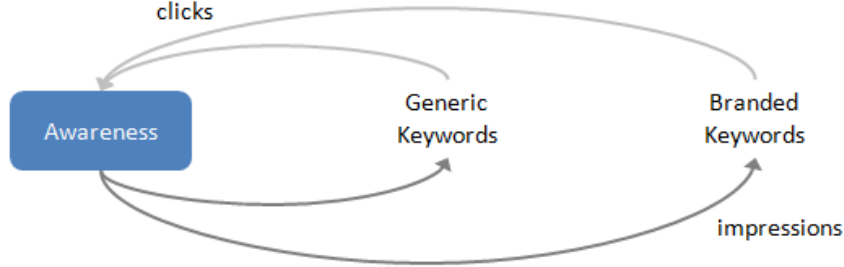

Figure 3: Interaction between search activity and latent awareness.

We first describe how generic and branded clicks affect awareness. 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 (19),

$$A_{t+1} = \eta_A A_t + \beta_G CLK_t^G + \beta_B CLK_t^B + \varepsilon_{t+1}^A, \tag{20}$$

where η_A captures the carry-over rate of awareness and ε_{t+1}^A is the idiosyncratic error term. Like Rutz and Bucklin (2010), we assume that the awareness at time $t+1$ is affected by the generic search activity at time t but in addition we allow for branded search activity to also impact awareness. As highlighted earlier, awareness is not observed in the data and is latent in this state-space model. 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 (2011) who show that awareness does not have a statistically significant impact on *click − through* and *conversion* rates. The expected number of generic impressions at time t is defined as $\mu_t^G = \sum_{k \in G} \mu_{k,t}$ and the expected number of branded impressions at time t is defined as $\mu_t^B = \sum_{k \in B} \mu_{k,t}$, where $\mu_{k,t}$ are the expected number of impressions for keyword k at time t. The expected number of generic and branded impressions

---

[11] We validate this assumption in our dataset by performing a Granger causality test and infer that impressions (both generic and branded) do not lead to more clicks but generic clicks lead to more branded impressions.

evolve with awareness in the following manner,

$$\mu_{t+1}^G = \eta_G \mu_t^G + \gamma^G A_{t+1} + \varepsilon_{t+1}^G, \tag{21}$$

$$\mu_{t+1}^B = \eta_B \mu_t^B + \gamma^B A_{t+1} + \varepsilon_{t+1}^B. \tag{22}$$

It should be noted that the effect of awareness is computed from aggregate data (not individually for each pair of generic/branded keyword).[12] In order to have a parsimonious model we assume that the effect of awareness is homogeneous across all branded keywords. Similarly, the effect across generic keywords is homogeneous.

Combining Equations (20)-(22), we get a state space model whose evolution is as follows

$$
\begin{bmatrix} \mu_{t+1}^G \\ \mu_{t+1}^B \\ A_{t+1} \end{bmatrix}
=
\begin{bmatrix} \eta_G & & \gamma_G \\ & \eta_B & \gamma_B \\ & & \eta_A \end{bmatrix}
\begin{bmatrix} \mu_t^G \\ \mu_t^B \\ A_t \end{bmatrix}
+
\begin{bmatrix} & & \\ & & \\ \beta_G & \beta_B & \end{bmatrix}
\begin{bmatrix} CLK_t^G \\ CLK_t^B \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{t+1}^G \\ \varepsilon_{t+1}^B \\ \varepsilon_{t+1}^A \end{bmatrix}
\tag{23}
$$

where the correlated error terms $\varepsilon_{t+1}^{\cdot}$ account for random shocks and $\varepsilon \sim N(\mathbf{0}, V_\varepsilon)$. The following equation represents how these latent states are linked to the observations,

$$
\begin{bmatrix} IMP_{t+1}^G \\ IMP_{t+1}^B \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}
\begin{bmatrix} \mu_{t+1}^G \\ \mu_{t+1}^B \\ A_{t+1} \end{bmatrix}
+
\begin{bmatrix} \nu_{t+1}^G \\ \nu_{t+1}^B \end{bmatrix}
$$

where IMP_{t+1}^G and IMP_{t+1}^B are generic and branded impressions at time $t+1$, $\nu_{t+1}^{\cdot} \sim N(0, V_\nu)$ is the random shock. We estimate this system of equations using a Dynamic Linear Model (DLM). DLMs have been used in several situations where an important component of the model is unobserved (Rutz and Bucklin, 2011, Bass et al., 2007, Naik and Sawyer, 1998). We estimate this model using a Markov Chain Monte Carlo (MCMC) approach as proposed by West and Harrison (1997). Details of the estimation procedure are outlined in Appendix A3. 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 presented in Table 10.

First, we note that there is a strong positive impact of generic clicks on awareness ($\beta_G > 0$). Second, increased awareness leads to increased branded search activity ($\gamma_B > 0$). Combining these

---

[12]One can conduct this analysis at a generic-branded keyword-pair level or a product level if there is sufficient data and variation in that data. Our dataset is very sparse to get statistical significance at keyword-pair or product level.

Table 10: Estimated Parameters		
Parameter	Mean	95% Conf. Interval
η_G	**0.9515**	[0.9735, 0.9321]
η_B	**0.8664**	[0.8411, 0.8842]
η_A	**0.2418**	[0.2297, 0.2547]
γ_G	0.0232	[-0.0006, 0.0427]
γ_B	**0.1088**	[0.0997, 0.1132]
β_G	**3.4018**	[3.2656, 3.6123]
β_B	0.0208	[-0.0105, 0.0461]

The figures in bold are statistically significant at the 95% level.

results, we conclude that every click on a generic ad increases the number of branded impressions by $\gamma_B \beta_G$ (= 0.38). We also observe that the effects of branded clicks on awareness and of awareness on generic search activity are insignificant ($\beta_B \approx 0, \gamma_G \approx 0$). These findings are consistent with the results reported by Rutz and Bucklin (2011). It appears reasonable that if a consumer is already aware of a brand, then clicking on a branded ad is less likely to change his awareness about that brand. Similarly, awareness about a particular brand does not affect consumer's generic search behavior.

We incorporate these estimates of spillovers into our decision theoretic model in the following manner. Given the statistically insignificant estimates of β_B and γ_G, 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 identical and lead to the same relative increase in the search (or impressions) for these branded keywords. More formally,

$$
\begin{aligned}
\mu_{k,t+1} &= \eta_B \mu_{k,t} + \gamma_{k,B} \beta_G CLK_{G,t} \quad \forall \, k \in B, \\
\mu_{k,t+1} &= \eta_G \mu_{k,t} \quad \forall \, k \in G.
\end{aligned}
\tag{24}
$$

where $\gamma_{k,B} = \gamma_B \frac{\mu_{k,t}}{\mu_{G,t}}$ is the increase in the expected impressions of keyword $k \in B$ at time period $t+1$ for every generic click at time t. The increased impressions for branded keywords, which are usually more profitable, leads to higher revenues in future periods.

6.2 Forward-Looking Policy

As discussed in the previous section, bidding on keywords has two effects - current period revenues and future awareness. As a result, the advertiser faces a trade-off between maximizing current period revenues and 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 for a finite time horizon. Lets denote the planning horizon by T. We assume that the budget in each time period should be less than or equal to D. The multi-period bidding problem is as follows

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

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

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

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

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

$\bar{\boldsymbol{\mu}}$, the vectors of mean impressions, constitute the state-space and the bids, $\bar{\boldsymbol{b}}$, 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 "myopic" policy. In order to find the optimal bids for $t < T$, we use approximate dynamic programming. We assume that the expected number of generic clicks at time t belongs to the set $CLK = \{0, 1, \ldots, M\}$, where M is an arbitrarily large number.[13] For every $CLK \in \{0, 1, \ldots, M\}$, 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

---

[13] $M = 200$ in our analysis.

t. This problem is stated as follows

$$\max_{\{b_t\}} \sum_k \mu_{k,t} \mathbb{E}\left[v_{k,t}|b_{k,t}\right] \quad \text{s.t.} \; C(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \leq D \text{ and } \sum_{k \in G} \mu_{k,t} \mathbb{E}\left[\delta_{k,t}|b_{k,t}\right] = CLK.$$

The optimal policy in this period is to choose a CLK (and the associated bids, $\bar{\boldsymbol{b}}_t$) that maximize the sum of current reward and the optimal future rewards. For the field experiment, we update bids once every two weeks and there are $T = 2$ time periods in total. The optimal bids under this forward-looking policy for some keywords are shown in Table 11 below. We also present the bids that would have been placed if we had used a myopic bidding policy instead. The forward-looking policy increases the bids for some of the generic keywords if they are likely to generate clicks. Accordingly, the bids for some of the less profitable branded keywords are reduced.

Table 11: Bids under the forward-looking policy

Keyword	$Ew(\$)$	Myopic Bids	Forward-Looking Bids
buy barbecue	1.94	0.65	0.85
porterhouse steaks	5.50	2.30	2.40
lobster bisque	0.00	0.00	0.20
Nebraska beef	2.69	1.25	1.25
purchase hot dog	4.47	2.45	2.65
buy top sirloins online	0.00	0.00	0.05
trout fillets	0.00	0.00	0.00
beef sirloin online	0.00	0.00	0.00
BRAND-NAME lobster bisque	6.43	2.45	2.15
BRAND-NAME steak burgers	14.59	3.00	3.00

6.3 Field Implementation

We apply the forward-looking policy to keywords in Group II. A daily budget $D = \$35.00$ is used based on the mean weekly spending of \$250 during the 3 month "before" period. We consider two time periods in our forward looking policy and compute the bids accordingly. The bids computed for the first period are deployed in the field for a period of 2 weeks and the bids computed for the second (last) period are deployed for two weeks thereafter.

During the 12 weeks in the "before" period, the advertiser incurred a cost of \$3052.60 and earned revenues of \$5646.03 for Group II keywords. In the "after" period the cost and revenues were \$1201.67 and \$2075.43, respectively. Using the Difference-in-Difference approach, as in Section

5.4, the improvement in performance is estimated to be

$$
\begin{aligned}
\tau_{FL} &= \Delta ROI_{\text{Group II}} - \Delta ROI_{\text{Control}} \\
&= (72.71\% - 84.96\%) - (-11.20\% - 84.30\%) \\
&= 83.25\%
\end{aligned}
$$

There is a notable increase in the performance of Group II keywords relative to the control group. Further, the forward-looking policy provides performance gains over and above that delivered by the myopic policy ($\tau_{FL} - \tau_M = 7.87\%$). The effectiveness of the forward-looking policy is likely to depend on prior brand awareness among search engine users and also on the duration of the experiment. Thus, the gains may vary in other settings based on prior brand awareness. We expect that the gains from the forward-looking policy will be greater if the field experiment is conducted over a longer duration. We were unable to experiment for an extended period of time due to limitations imposed by our partner advertiser.

7 Discussion

In this section, we contrast our proposed approach with policies commonly used by advertisers in sponsored search. Then we shall discuss some limitations of our models which might limit the applicability of our bidding policies.

7.1 Contrast with Commonly Used Strategies

Our agreement with the advertiser precludes sharing their exact bidding strategy. However, their strategy is fairly typical of strategies used by most advertisers in sponsored search. There are three main reasons why our policies perform better than the policies adopted by these advertisers. Firstly, because of the complexity of bid determination, most advertisers use simple heuristics to determine bids. One common heuristic is to simply raise bids for keywords that generate purchases at a relatively low cost and to reduce bids for keywords that do not generate purchases. While this is a reasonable heuristic, it does not account for details of the bid distribution or how the CTR decays with position. For e.g., for some keywords, reducing bids may reduce clicks significantly but it may not have a significant impact on cost-per-click. Parameters tied to competing bids and click decay have a significant impact on optimal bids. As shown in Table 8 it might not be optimal to

invest heavily in a profitable *but* highly competitive keyword (birthday present). Another challenge for advertisers is that they often manage bids for keywords individually without optimizing the portfolio as a whole. Raising and lowering bids for keywords in equal increments to manage the budget constraint is suboptimal. Optimizing the bids over the entire portfolio helps to move the advertising dollars from poorly performing keywords to profitable ones in the right increments. Thirdly, the forward-looking policy accounts for the two-fold effect that sponsored search ads have - awareness and profits. By ignoring the awareness benefits of generic keywords, advertisers often under-invest in generic keywords and over-invest in branded keywords.

7.2   Competitive Reaction

This paper adopts a decision-theoretic perspective of the bid problem as opposed to an equilibrium perspective. Advertisers have to submit bids based on their current beliefs and may choose to update these bids as their beliefs evolve. Our framework accommodates that by assuming that advertisers can use new data to re-estimate the model parameters and update their bids. If competing advertisers respond instantaneously to changes in bids then this may reduce the effectiveness of our bidding policies or at the very least suggest that bids need to be rapidly and continuously updated. However, current research suggests that competition in sponsored search advertising is fairly subdued (Rutz and Bucklin 2010, Steenkamp et. al. 2005). Our discussions with several managers indicates that bids for these keywords are rarely updated continuously. This is also reflected in our dataset where bids for less than 10% of the keywords were changed in the 12 week "before" period when the advertiser was deciding the bids.

To test whether rapid reaction by competitors render our computed bids ineffective, we compare the difference between the predicted and observed average (*i*) position and (*ii*) *cost per-click* in the after period (presented earlier in Equations (16) and (18)). If competitors react soon to our advertiser's new bids, it would introduce notable errors in our predictions regarding the expected position and cost-per-clicks. For most of the keywords, there is no significant difference between the predicted moments and the daily summaries reported by the search engine, which indicates that there is no significant short-term competitive reaction.[14] There can however be long-term competitive reaction and the model parameters $(\lambda, \theta, \alpha, \gamma)$ can be periodically re-estimated and the bids updated to account for these changes. This estimation would not suffer from endogeneity

---

[14]The Mean Absolute Error (MAE) of these moments averaged across all keywords are shown in Table 12 in Appendix A2.

issues as long as the bids are determined through the proposed algorithm and are uncorrelated with random shocks. Since the issue of endogeneity no longer arises, there might not be a need for a random bidding period.

7.3 Spillover across Groups

While computing the effectiveness of the "forward-looking" bidding policy in Section 6, we implicitly assumed that there are no spillovers across groups. However, spillovers from keywords in one treatment group into keywords in another group might influence the estimate of τ_{FL}. To control for this, we divide keywords into product categories and assign all keywords from a product category into the same treatment group. This experimental design is motivated by the intuition that clicks on keywords related to a particular product will not have any impact on the search behavior for other products, e.g. while clicks for "hot dogs" can spill over to branded keywords within the same product category "BRAND-NAME hot dog", it will have insignificant impact on searches for "salmon" or "BRAND-NAME salmon" . This procedure of random assignment by product categories helps ensure such that most of the spillovers are within treatment groups. This experimental design is based on Angelucci and Giorgi (2009) where they propose a methodology to measure treatment effects with spillovers. Note that the random assignment of the product categories to groups also ensures that even if there are some spillovers across groups, these effects are similar between any given pair of groups. A more sophisticated way to incorporate the spillover effect might be a multi-tier design as proposed by McConnell, Sinclair and Green (2010) but this approach would severely affect the analytical tractability of our approach and has been left as a direction for future research.

7.4 Multi-Channel Retail

In this paper, we assume that the sponsored ads only affects online revenues. The campaign considered for this analysis was restricted to the northeast region where the advertiser has very few brick-and-mortar stores. In addition, the demographics of consumers who visit the advertiser's website are very different from consumers who buy their products from the offline stores. Hence, interactions between channels are not a major concern in the current empirical context. However in a more general situation, there can be interactions between online advertising and offline sales, and Lewis and Reiley (2011) show evidence of online sales positively affecting offline sales. The effectiveness of our techniques would improve with better attribution of revenues to their source. Our approach in this paper ignores multi-channel issues and more sophisticated approaches will

prove useful. Only accounting for online sales might lead to an underestimation of *Ew*. However, given the randomization procedure we employed, we expect no systematic differences in the extent of underestimation between the two groups. This would imply that performance gains we observe in the "after" period for both Groups I and II are underestimated and once the offline revenues are taken into account, the magnitude improvements can be even higher.

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, 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 advertiser's RoI. We extend our basic model to account for secondary effect of these ads - awareness - and show that incorporating awareness into a multi-period bidding problem can help increase revenues further.

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 in advertisers' bidding policies 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

Proofs of Equations and Propositions

Solution of Equation (1)

The constrained optimization problem is as follows

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

The Lagrangian can be written as:

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

KKT Conditions

$$\forall k: \frac{d\mathcal{L}}{db_k} = \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} v_k^{(s)}|b_k\right] - \lambda\frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} c_k^{(s)}|b_k\right] = 0$$

$$\lambda \geq 0,$$

$$D - \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)}\right] \geq 0.$$

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

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

As $rank\left(\frac{d\left(D-\mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)}\right]\right)}{db}\right) > 0$, 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

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

$$\text{or} \quad \frac{d\mathbb{E}[v_k|b_k]}{db_k} = \lambda\frac{d\mathbb{E}[c_k|b_k]}{db_k}.$$

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,

$$F\left(\underline{b} = x | b, pos = i\right)$$

$$= \Pr\left\{\underline{b} < x | b, pos = i\right\},$$

$$= \frac{\Pr\left\{\underline{b} < x, pos = i | b\right\}}{\Pr\left\{pos = i | b\right\}},$$

$$= \frac{\binom{N}{i} \left(1 - F\left(b\right)\right)^i F\left(x\right)^{N-i}}{\binom{N}{i} \left(1 - F\left(b\right)\right)^i F\left(b\right)^{N-i}},$$

$$= \frac{F\left(x\right)^{N-i}}{F\left(b\right)^{N-i}}.$$

Proof of Lemma 2

$$F\left(\underline{b} = x | b, \delta = 1\right)$$

$$= \Pr\left\{\underline{b} < x | b, \delta = 1\right\},$$

$$= \sum_{i=0}^{N} \Pr\left\{\underline{b} < x | b, \delta = 1, pos = i\right\} \times$$

$$\Pr\left\{pos = i | b, \delta = 1\right\},$$

$$= \sum_{i=0}^{N} F\left(x | b, pos = i\right) \times$$

$$\frac{\Pr\left\{\delta = 1 | b, pos = i\right\} \Pr\left\{pos = i | b\right\}}{\Pr\left\{\delta = 1 | b\right\}},$$

$$= \sum_{i=0}^{N} \left(\frac{F\left(x\right)}{F\left(b\right)}\right)^{N-i} \times$$

$$\frac{\frac{\alpha}{\gamma^i} \binom{N}{i} \left(1 - F\left(b\right)\right)^i F\left(b\right)^{N-i}}{\alpha \gamma^{-N} \left(1 + \left(\gamma - 1\right) F\left(b\right)\right)^N},$$

$$= \sum_{i=0}^{N} \binom{N}{i} \left(\gamma F\left(x\right)\right)^{N-i} \frac{\left(1 - F\left(b\right)\right)^i}{\left(1 + \left(\gamma - 1\right) F\left(b\right)\right)^N},$$

$$= \frac{(1 - F(b) + \gamma F(x))^N}{(1 + (\gamma - 1) F(b))^N}.$$

Proof of Proposition 2

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

Proof of Equation 13

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

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 condition (Equation 14).

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

$\Psi'(b) > 0$ 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

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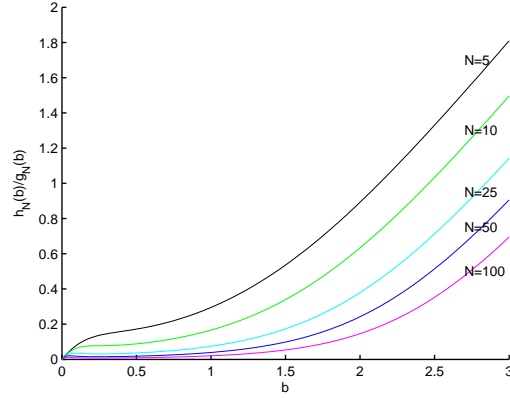\end{aligned}
$$

Figure 4: $h_N(b)/g_N(b)$ 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.

$$= (N-1)f(b)\left[(\gamma-1)\frac{h_{N-1}(b)}{g_{N-1}(b)} + \frac{h_{N-2}(b)}{g_{N-2}(b)}\right].$$

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 (4) 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)

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

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

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 (5) for some sample parameters.
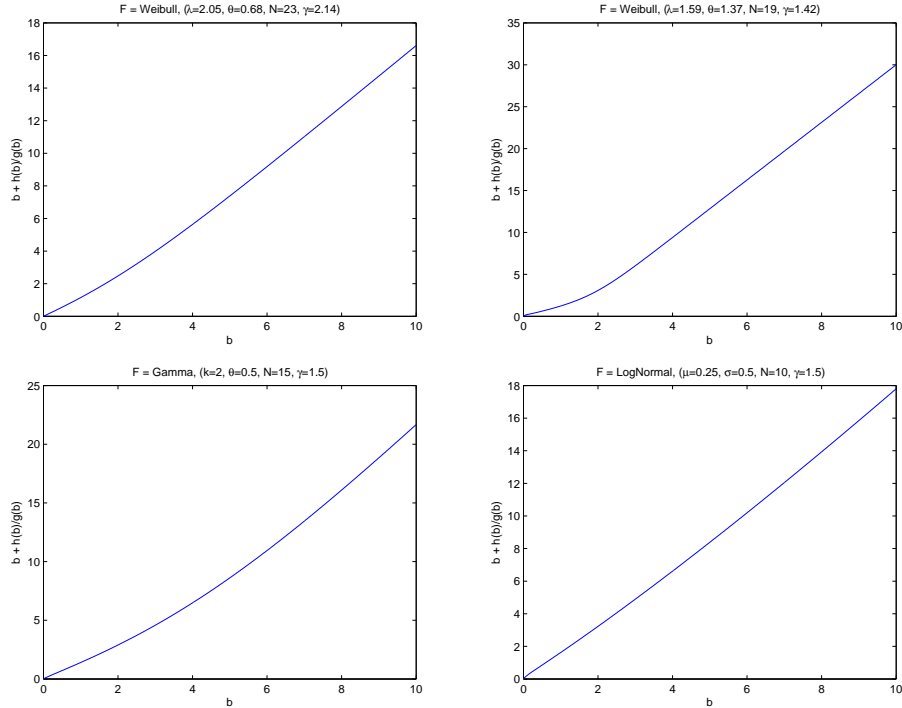
Figure 5: $\Psi(b)$ for various distributions.

Measuring Competitive Reaction

If there is competitive reaction then the predicted average position and CPC would be considerably different from the observed position or CPC as the competitors might change their bids as a response to the changes in bids by the advertiser. If the predicted and observed moments of these quantities are not very different, it suggests that the competitive reaction is subdued. In order to measure competitive reaction, we compute the difference between the predicted daily average position and *cpc* and the mean of these quantities. The MAE is reported in the table below.

Table 12: MAE between the predicted and observed moments

Quantity	MAE
position	0.141
cpc	$0.064

Given that these observed quantities are very close to the predicted values, this provides evidence to suggest that there is very weak competitive reaction during the experimental phase.

Appendix A3: Estimation of DLM parameters

This appendix provides an overview of the sampling procedure used to estimate parameters of the Dynamic Linear Model mentioned in Section 6.1. The sampling procedure mentioned here is an application of the method proposed by West and Harrison (1997). We need to estimate the parameters of the transition matrix (η_A,η_B,η_G, γ_G,γ_B), the effect of generic and branded clicks (β_G,β_B), the covariance matrices (V_ε,V_ν) and the sequence of state vectors $\Phi_T = \{\phi_1, \ldots, \phi_T\}$. We start off with non-informative Gaussian priors for these parameters $\Psi =(\eta_A$,η_B,η_G, γ_G,γ_B,β_G,$\beta_B)$. We also assume that ε_t and ν_t are independent and the priors for V_ε and V_ν are assumed to be inverse Wishart. Given these assumptions, the posteriors distributions of V_ε and V_ν are inverse Wishart and the posteriors for the parameters (η_A,η_B,η_G, γ_G,γ_B,β_G,β_B) are Gaussian.

Let $\boldsymbol{D}_t = \{Y_t, \boldsymbol{D}_{t-1}\}$ denote all the information available to the researcher till time t, e.g. the clicks and impressions till time t. We use a forward-filtering and backward smoothing algorithm (e.g. Rutz and Bucklin, 2011) to sample the state spaces, $\Phi_t | \boldsymbol{D}_t$. Then we sample the parameters (η_A,η_B,η_G, γ_G,γ_B,β_G,β_B) given Φ_t and $\boldsymbol{D}_t$. These estimation steps are described below.

Step 1: Simulation for Φ_T

i) For $t = 1, \ldots, T$, compute m_t and Σ_t, the mean and the variance of the state space at time t. m_t and Σ_t are derived sequentially from the priors m_0 and Σ_0 according to the procedure outlined in West and Harrison (1997, Chapter 4).

ii) Filter-forward step: For $t = T$, sample $p(\phi_T | \boldsymbol{D}_T)$ from the posterior distribution $N(m_T, \Sigma_T)$.

iii) Backward-smoothing step: For $t = T, \ldots, 1$, sample p($\phi_{t-1} | \phi_t, \boldsymbol{D}_T$) conditional on the latest draw ϕ_t.

Step 2: Sampling from $p(\Psi, V_\varepsilon, V_\nu | \Phi_T, D_T)$

We sample the parameters Ψ, V_ε and V_ν sequentially. This is reasonable as the elements of the transition matrix, drift vectors and the error terms are assumed to be independent of each other. Based on these assumptions the Gibbs sampler can be used in a straight-forward manner to draw samples of η_A, η_B, η_G, γ_G, γ_B, β_G, β_B, V_ε and V_ν separately.

Acknowledgment

References

V. Abhishek, K. Hosanagar, and P. Fader. On aggregation bias in sponsored search data: Existence and implications. *SSRN eLibrary*, 2011.

A. Agarwal, K. Hosanagar, and M. D. Smith. Location, location, location: An analysis of profitability and position in online advertising markets. *Journal of Marketing Research*, 48(6):1057–1073, 2011.

G. Aggarwal and J. D. Hartline. Knapsack auctions. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, SODA '06, pages 1083–1092, 2006.

G. Aggarwal, A. Goel, and R. Motwani. Truthful auctions for pricing search keywords. In *Proceedings of the 7th ACM conference on Electronic commerce*, EC '06, pages 1–7, 2006.

K. Ali and M. Scarr. Robust methodologies for modeling web click distributions. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 511–520, 2007.

M. Angelucci and G. D. Giorgi. Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review*, 99(1):486–508, 2009.

F. Bass, N. Bruce, S. Majundar, and B. Murthi. A dynamic bayesian model of the advertising-sales relationship. *Marketing Science*, 26(2):179–195, 2007.

C. Borgs, J. Chayes, N. Immorlica, K. Jain, O. Etesami, and M. Mahdian. Dynamics of bid optimization in online advertisement auctions. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 531–540, 2007.

N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-

bias models. In *Proceedings of the international conference on Web search and web data mining*, WSDM '08, pages 87–94, 2008.

B. Edelman and M. Ostrovsky. Strategic bidder behavior in sponsored search auctions. *Decision Support System*, 43:192–198, February 2007.

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

J. Feldman, S. Muthukrishnan, M. Pal, and C. Stein. Budget optimization in search-based advertising auctions. In *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, pages 40–49, 2007.

J. Feng, H. K. Bhargava, and D. M. Pennock. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS Journal on Computing*, 2005: 137–148, 2006.

A. Ghose and S. Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10):1605–1622, 2009.

L. P. Hansen, J. Heaton, and A. Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economic Statistics*, 14(3):262–280, 1996.

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

R. A. Lewis and D. J. Reiley. Does retail advertising work? measuring the effects of advertising on sales via a controlled experiment on yahoo! *Yahoo! Reasearch Paper Series*, 2011.

M. Mahdian, H. Nazerzadeh, and A. Saberi. Allocating online advertisement space with unreliable estimates. In *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, pages 288–294, 2007.

A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized online matching. *J. ACM*, 54, October 2007.

S. Muthukrishnan, M. Pál, and Z. Svitkina. Stochastic models for budget optimization in search-based advertising. In *Proceedings of the 3rd international conference on Internet and network economics*, WINE'07, 2007.

M. K. M. Naik, Prasad A. and A. G. Sawyer. Planning media schedules in the presence of dynamic advertising quality. *Marketing Science*, 17(3):214–235, 1998.

M. Nerlove and K. J. Arrow. Optimal advertising policy under dynamic conditions. *Economica*, 29 (114):129–142, 1962.

P. Rusmevichientong and D. P. Williamson. An adaptive algorithm for selecting profitable keywords for search-based advertising services. In *Proceedings of the 7th ACM conference on Electronic commerce*, EC '06, pages 260–269, 2006.

O. J. Rutz and R. E. Bucklin. From generic to branded: A model of spillover dynamics in paid search advertising. *Journal of Marketing Research*, 48(1), 2011.

O. J. Rutz, R. E. Bucklin, and G. P. Sonnier. A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *Journal of Marketing Research*, 49(3), June 2012.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, New York, NY, 1997.

J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, October 2001.

S. Yang and A. Ghose. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, page mksc.1100.0552, 2010.

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