An Empirical Analysis of Return on Investment Maximization in Sponsored Search Auctions

Jason Auerbach
Stanford University
jasonaue@cs.stanford.edu

Joel Galenson
Stanford University
galenson@cs.stanford.edu

Mukund Sundararajan
Stanford University
mukunds@cs.stanford.edu

ABSTRACT

We empirically investigate whether advertisers are maximizing their return on investment (ROI) across multiple keywords in sponsored search auctions. Because testing for ROI maximization relies on knowledge of advertisers' private true values per click, we instead use necessary (although not sufficient) conditions for ROI maximizing behavior that rely only on advertisers' bids. We classify advertisers based on the extent to which they satisfy the test conditions. Our results indicate that a large fraction of advertisers in the Yahoo Webscope first price data set may be following ROI-based strategies.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity

General Terms

Sponsored Search Auctions, Empirical Analysis, Campaign Optimization, Bidding Agent, Return on Investment

1. INTRODUCTION

Search engines sell advertising positions alongside organic search results by conducting *sponsored search auctions*. The goal of this paper is to better understand the empirical behavior of bidders (advertisers) in these auctions. There are two reasons why such an understanding is important. Firstly, understanding advertiser needs will allow search engines to develop bidding tools, user interfaces and features that help advertisers achieve their goals. Secondly, as we discuss below, our empirical investigation helps guide theoretical modeling and analysis of these auctions.

Academic research has focused primarily on the single-keyword, single-shot model of a sponsored search auction (see [13] for a good overview). The basic assumption is that bidders strategize independently on each keyword they bid on. This assumption, however, need not hold. Advertisers may not have the resources or sophistication to track each keyword separately, and their spending on each key-

Number of Keywords	Percentage of Bidders Who Bid On This Many Keywords
1	47
2	17
3-5	18
6-25	14
>25	3

Figure 1: Number of keywords bid on by bidders in the Yahoo Webscope data set (described in Section 3.1). Percentages do not sum to 100 due to rounding.

word may be limited by a single advertising budget that applies to all keywords collectively. This may cause advertisers to coordinate their strategies across keywords. If this is indeed happening, the behavior of a large proportion of bidders could be affected. As Figure 1 shows, half of all bidders in our data set (explained in Section 3.1) bid on at least two keywords.

One promising model for multi-keyword bidding strategies is based on the maximization of bidders' return on investment (ROI). Papers exploring ROI in sponsored search auctions include [3, 5, 7, 10]. We draw in particular on the work of [3]. The main result of the paper is that if all advertisers follow a particular ROI-based strategy, then under a certain reasonable model the system reaches a *market equilibrium*. If we can find empirical evidence that advertisers follow a similar strategy, then we can use the vast literature on market equilibria (see chapters 7 and 8 of [8] and the references therein) to better understand sponsored search auctions.

There are two key difficulties in using empirical data to determine whether bidders are maximizing their ROI. Firstly, the definition of ROI relies on a bidder's true value per click, a quantity which is private to the bidder: she does not share her true value with the auctioneer or any other party. Without knowing this value, it is impossible to conclusively say that a bidder is maximizing her ROI. There are, however, some necessary conditions for ROI maximization that rely solely on public information (bids, times, etc) contained in the data sets we used. If a bidder is violating some of these conditions, then we can conclude that she is not maximizing her ROI. For the other bidders, we are limited to saying that they meet certain necessary conditions for ROI maximization.

The second key problem is that few tests will be *exactly* fulfilled by empirical data. To overcome this problem, we follow the approach of [11] and calculate the minimal per-

turbations that must be made to the data in order for the conditions to be fulfilled. If these perturbations are small enough, then we can conclude that the test condition probably holds.

We run several tests on empirical multi-keyword data. We focus in particular on tests that relate to ROI optimization in the model proposed by [3]. ROI-based strategies are supported by some of the most popular bidding agents, such as Apex Pacific's *PPC BidMax* and Omniture's *SearchCenter*. It is quite easy for a bidder to employ an ROI-based strategy: she only needs to know the money spent on and value derived from each keyword. (In fact, knowing just the relative values of his keywords is sufficient.) Therefore, we believe an empirical investigation of ROI-based strategies to be well-motivated.

2. OUR CONTRIBUTION

We formulate empirical tests that check for evidence of ROI maximization and other forms of bidder rationality. These tests use only the public information revealed by bidders. To investigate conditions (or types of rationality) that rely on private information, such as true values per click, we use the public information to build necessary conditions for having the private information for which we seek to test. Our analysis produces the following results:

- 65% of first price bidders display behavior consistent with the ROI-maximization strategy proposed by [3]. Specifically, these bidders need bid perturbations of 5% or less in order to have consistent ratios across time between their bids on different keywords.[1] See Section 6.2 for more details.

- 83% of first price bidders need bid perturbations of 5% or less in order to maintain a consistent lowest-to-highest ordering of their bids on different keywords. This implies that the majority of bidders are at least partially optimizing their ROI. See Section 6.1 for more details.

- For both of the previous two tests, results varied dramatically based on the number of keywords being bid on. For example, 84% of advertisers bidding on only 2 keywords display behavior consistent with the ROI optimization strategy, whereas only 26% of bidders bidding on more than 25 keywords display such behavior.

- Most bidders are sophisticated enough to submit different bids on different keywords. 63% of first price bidders submit meaningfully different bids across their different keywords. (We consider a difference to be meaningful if it cannot be canceled out by perturbations of up to 5%.) See Section 6.3 for more details.

- In the first price data set, only 20% of bidders submit the same bids across time. In the second price data set, 70% of bidders submit the same bids across time. This difference reflects the increased stability of the second price auction. See Section 6.3 for more details.

---

[1]Note that displaying this behavior does not necessarily mean that the ROI-maximization strategy was being followed. For example, submitting identical bids on every keyword or every day is a trivial way of displaying this behavior. This issue is further explored in the body of this paper.

- There are weak but positive correlations between all of the following: number of keywords bid on, frequency of bid updates, bid spread (the mean difference between bids on any two keywords) and mean bid. This indicates, for instance, that the ability to bid on multiple keywords also implies increased bidding activity on any one of them. However, the lack of strong correlation also indicates that bidders in the data set are fairly diverse. See Section 6.4 for more details.

3. SETUP

3.1 Data sets

We ran all tests on Version 1.0 of *Yahoo! Search Marketing advertising bidding data*, provided as part of the Yahoo! Research Alliance Webscope program. The data set contains all bids placed on the top 1,000 keywords during the time period 6/15/2002 to 6/14/2003. There are 10475 bidders in total. Each bidder is assigned a unique anonymous identifier so that he can be tracked across time and keywords. Bid data are sampled in 15 minute time increments (which we then average across an entire day). If a bidder bid on more than 50 keywords in the data set, or more than 90 days, the data were truncated to permit tractable analysis. This truncation does not compromise the analysis: the data are inherently truncated at some level, in the sense that we only have access to bids on the top 1000 keywords, and all bids are sampled from a one year time period. Furthermore, only 10% of bidders were affected by the truncation.

The data are from a first price auction. Although first price auctions are no longer used in practice, we believe that the results from our analyses on the first price data set are also applicable to a second price auction. This is because our analysis focuses on multi-keyword properties. The exact way in which bidders behave in an auction is not relevant to most of our tests; rather, we are interested in how bidders allocate their budgets across keywords, which is a common problem regardless of the type of auction being used. First price auctions were also used in much of the theoretical analysis done by [3].

In some ways, a first price data set is actually preferable to a second price data set for our analysis: a first price auction allows advertisers to control exactly how much they spend per click, and by extension they can control their exact ROI on each keyword. This contrasts with a second price auction, in which advertisers pay the bid of the bidder ranked immediately behind them, thereby not having fine-grained control over their ROI on each keyword. Thus, if advertisers are attempting to maximize their ROI, it should be much easier to see evidence of this behavior in a first price auction than a second price auction.

We also had access to a small second price data set. The data were collected by [2] using the Overture Bid View tool. The data set includes daily bid magnitudes of all bidders on 45 keywords. Each keyword is the name of a car manufacturer. In total, there were 629 bidders in the data set. The data set is limited in size, and is market-specific. Therefore, the second price results should be treated cautiously. They are included mainly for purposes of rough comparison with the first price results, and to justify our belief that results from the first price data set also apply to second price auctions.

3.2 Model and Notation

We use a subset of the model proposed by [3], with a few small changes to fit with our analyses.

A bidder submits bids per click on keywords $1, 2, \cdots, K$ on days $1, 2, \cdots, T$. Let $b_{k,t}$ be the bid submitted on keyword k for day t. If a bidder bids different amounts on keyword k over the course of a single day, we set $b_{k,t}$ to be his average bid for that day. Because this is a first price auction, the bid per click is also the price paid per click. Let $\theta_{k,t}$ be the number of clicks on keyword k over the course of day t. This measurement is slot agnostic, and is well defined even if a bidder resides in different slots over the course of a single day. Each bidder has a budget B_t for each day t. This budget applies to all keywords collectively. In other words,

$$\forall t. \sum_{k=1}^{K} b_{k,t} \cdot \theta_{k,t} \leq B_t. \tag{1}$$

Each bidder is assumed to have a value per click v_k for each keyword k. This value per click is constant across time.

We make the simplifying assumption that the number of clicks a bidder receives is a strictly increasing function of his bid. So if $b_{k,t} > b'_{k,t}$, then $\theta_{k,t} > \theta'_{k,t}$. The purpose of this assumption is to provide an intuitive justification for why an advertiser should attempt to raise his bid on high-ROI keywords and lower his bid on low-ROI keywords (see Section 4). Although this assumption may not be true immediately after a bid change, over the course of an entire day there is a lot of churn, with bidders frequently swapping slots. Thus, when we consider the day as a whole, a bidder will most likely have received more clicks as a result of a higher bid or fewer clicks as a result of a lower bid. Of course, this assumption cannot apply when a bidder is bidding so low that his ad is never appearing or so high that he is in the top slot for the entire day. To compensate for this problem, we ignored bidders in the top slot for some of our empirical analyses.

4. RETURN ON INVESTMENT: A REVIEW OF PRIOR RESEARCH

In this section, we summarize the return on investment (ROI) model proposed by [3], and introduce the ROI maximization condition that we test for in this paper.

Return on investment is the ratio of utility (i.e. value minus cost) over price. For example, if a bidder pays \$1 per click, and gets \$1.50 in value from each click, then his ROI is $\frac{1.50-1}{1} = 0.5$.

ROI optimization is related to value per unit price optimization in the following way:

$$\frac{U}{P} = \frac{V-P}{P} = \frac{V}{P} - 1 \tag{2}$$

where U, P, and V represent the utility, price and value respectively. Thus, maximizing ROI is equivalent to maximizing the value obtained per unit price. If the advertiser is exactly meeting his budget, the total price is fixed: it is always the budget. Thus, if a budget is being exactly met, ROI maximization is equivalent to maximization of the value derived from clicks on the ad.

If a bidder maximizes her overall ROI, then she will, in effect, equalize the marginal ROI she gets from *each* keyword she bids on. If her marginal ROI is higher on one keyword than another, then she could lower her bid slightly on the keyword with lower marginal ROI and raise her bid slightly on the keyword with higher marginal ROI. She could do this in a way that resulted in the same expenditure but higher utility, and thus in higher overall ROI. Because the marginal ROI of a keyword is difficult to estimate (and is sometimes even undefined), [3] approximates it with the ROI. Thus, if a bidder is maximizing her overall ROI, we expect her to have the same ROI on each keyword that she bids on.

Formally, if ROI (and by extension, total value) is being maximized in a budget-constrained setting then on any given day t we expect $\frac{v_k}{b_k} = s_t$ for some constant scale factor s_t. The scale factor s_t may change from day to day in order to always spend exactly the budget. Even if a hard budget does not exist, we may still see behavior similar to this condition. Bidders who are not budget constrained may follow an ROI-like strategy purely for its relative simplicity (see, for example, the strategy described at the very end of this section).

Rearranging this equation gives us the following *ROI maximization condition*:

$$\forall k, t. \left(s_t \cdot b_{k,t} = v_k \right). \tag{3}$$

We test for adherence to this condition. One simple way for bidders to adhere to this condition would be to follow the algorithm proposed in [3]. In this algorithm, bidders always maintain the same relative bids on their keywords. At the end of each day, they see whether their budget was exhausted. If it was exhausted, they scale all bids down by the same factor. It it was not exhausted, they scale all bids up by the same factor. (This scale factor corresponds to the s_t terms in equation (3).) This strategy relies on very little information or sophistication; all a bidder needs to know in order to apply this strategy is the relative value she has for clicks on each keyword and the total amount of money that she is spending.

5. EMPIRICAL ANALYSIS TECHNIQUES

Bidders in a sponsored search auction have both public information (the bids they submit) and private information (their true values per click). The ROI maximization condition is somewhat difficult to test for, because it involves knowledge of private information. Essentially, we overcome this lack of knowledge by using necessary but not sufficient conditions. We follow the approach of [12, 13]. If there exists a set of private values that, along with the public information, is consistent with the ROI maximization condition, then we conclude that bidders may be rationally optimizing their return on investment. This style of analysis is inspired by *revealed preference theory* (see [9] for the more information on this subject).

Sponsored search auctions tend to have a lot of entropy, and it is unlikely that any condition we test for will be exactly fulfilled in a real-world situation. Sources of this seemingly random disorder include inaccurate estimates of values per click, fluctuating search volume and click-through-rates, strategic bidding behavior (e.g. [6]), repeated game bidding strategies, changing participants in the market, changing budgets and inertia towards updating bids. Also, it is possible that bidders are purposely following strategies only in a loose manner. Thus, in all of our tests, we sought to find for which bidders a condition *approximately* held.

As an example, imagine we are trying to test whether the ratio between the magnitudes of the bids on keywords A and B is identical across time. Clearly, if the bids are just a couple of cents away from having an identical ratio, we do not want the bidder to fail the test. The dynamics described above mean that a couple of cents is actually very small in comparison to the general background disorder that occurs in a sponsored search auction.

In order to see whether a condition approximately holds, we assume that all bids have a certain error in them caused by the seemingly random background factors mentioned previously. We then follow the approach of [12, 13]. We find the minimal perturbations to the bids that are required for a test condition to be fulfilled.

The smaller the perturbations, the more likely a condition is to hold. In our tables, we sometimes provide results for a variety of perturbation tolerances. Otherwise, we classify bidders as conforming to a condition if they can do so with bid perturbations of up to 5% of the bid magnitude. We believe that this figure is relatively conservative; in reality, the large amount of entropy in background auction dynamics probably amounts to more than a 5% error.

It is also possible to formalize the effect of uncertainty in the auction setting and argue that a bidder's inability to follow the strategy exactly is due to the noisy information it has. We discuss such an analysis in Section 8 based on ideas from Hal Varian in [12]. We see similar trends using both approaches. We opt to use the informal approach, as it is easier to comprehend.

6. RESULTS

6.1 Weak Test for ROI: Consistent Ordering of Bids Across Keywords

We begin by testing for evidence of attempted ROI maximization (a test for the full ROI maximization condition follows this test). As stated in Section 4, if full ROI maximization is occurring, on any given day, a bidder must have an identical ratio $\frac{v_k}{b_k}$ for every keyword k. Put another way, on every day, the ratio $\frac{b_{k_1}}{b_{k_2}}$ should be the same for any keywords k_1, k_2. Perhaps bidders are not sophisticated enough for this strong condition to hold. If so, a more reasonable condition to test for is whether bidders maintain a consistent *ordering* of their bids on keywords. If a bidder bids more on keyword 1 than keyword 2 on one day, does he do so on all days? A consistent ordering is a necessary condition for full ROI maximization.

We ran an empirical test to see what proportion of bidders meet this condition. We found the minimal perturbations that result in this condition being true. We calculated these perturbations using the following procedure.

1. Re-number the keywords so that keyword 1 has the highest average bid, keyword 2 has the second-highest average bid, and so on.

2. Solve the following quadratic program.

$$\text{minimize} \sum_{k,t} (x_{k,t})^2 \text{ such that}$$
$$\forall k, t (b_{k,t} + x_{k,t}) \geq (b_{k+1,t} + x_{k+1,t})$$
$$\text{where } b_{K+1,t} = x_{K+1,t} = 0 \text{ for all } t$$

(If a bid was not submitted for a specific keyword/time pair, then we ignored that pair in our calculations.)

This quadratic program can in theory be solved using a standard quadratic program solver. (For example, Matlab's `quadprog`.) We found, however, that solving the quadratic program in this way became computationally intractable for some of the more complicated bidders. To work around this problem, we reformulated the program as a gradient descent problem. This reformulation does not affect the results; it merely allows for faster analysis.

Our results show that a large majority of bidders are able to meet this simple condition. If we permit bid perturbations of up to 5%, then 83% of bidders are able to achieve a consistent ordering of their keywords on every day. More detailed results are presented in Figure 2.

6.2 Full ROI Maximization

Recall the ROI maximization condition from equation 3:

$$\forall k, t. \left(s_t \cdot b_{k,t} = v_k \right).$$

For each day t, there is a scale factor s_t between the bids and the values for each keyword k. In other words, for any two keywords k_1, k_2, the ratio $\frac{b_{k_1}}{b_{k_2}}$ is the same on all days.

We sought to discover what proportion of bidders obey this condition. To find the minimal perturbations required to meet the ROI maximization condition, we solved the following program:

$$\text{For each bidder}$$
$$\text{minimize} \sum_{k,t} (x_{k,t})^2 \text{ such that}$$
$$\forall k, t. \ s_t \left(b_{k,t} + x_{k,t} \right) = v_k \text{ and}$$
$$s_1 = 1$$

(We fix $s_1 = 1$ to avoid multiple solutions. If a bid was not submitted for a specific keyword/time pair, then we ignored that pair in our calculations.)

We solved this program using gradient descent. The solution to the program contains the values per click that result in the minimal perturbations required to meet the ROI maximization condition. Because these values per click may not correspond to the true values per click (which are private information), adherence to the ROI maximization condition is a necessary but not sufficient condition for ROI maximization. However, the condition is quite strong, so we suspect that most non-trivial bidders who fulfill the condition are indeed maximizing their ROI.

Because bidders in the top slot cannot expect to receive more clicks when they raise their bid, the ROI model breaks down for them. To compensate for this, we ignored the perturbations of all bidders in the top slot. (This action has only a slight effect. It ignores about 20% of the perturbations and changes the results by at most 4 percentage points.)

Our results show that a majority of bidders are able to fulfill this condition. If we permit bid perturbations of up to 5%, then 65% of bidders obey the ROI maximization condition. More detailed results are presented in Figure 3.

6.3 Eliminating Trivial Bidders

Some of the bidders who passed the consistent ordering or ROI maximization tests passed for a rather trivial reason: they submitted nearly identical bids on every keyword, or

% permissible perturbation on bids	0	5	10	20	30
Number of keywords					
2	75	96	99	100	100
3-5	50	90	97	100	100
6-25	21	68	91	99	100
>25	5	36	75	97	100
Overall	48	83	95	100	100

Figure 2: Results of the *consistent keyword ordering* test. Each cell contains the percentage of bidders who bid on the corresponding number of keywords who are able to maintain a consistent highest-to-lowest bid ordering for their keywords after we apply bid perturbations of up to the corresponding amount.

% permissible perturbations on bids	0	5	10	20	30
Number of keywords					
2	59	84	93	97	97
3-5	36	68	85	97	99
6-25	15	45	66	92	98
>25	6	26	45	88	97
Overall	36	65	80	95	98

Figure 3: Results of the *ROI maximization condition* test. This table follows the same format as Figure 2. Each cell contains the percentage of bidders who are able to meet ROI maximization condition after we apply bid perturbations of up to the corresponding amount.

they submitted nearly identical bids for every day. This may of course mean that they are maximizing their ROI, but it may also mean that they are merely very unsophisticated bidders. We ran tests to see what proportion of bidders submitted nearly identical bids across time and keywords. As always, we allow for a small standard error on bids due to the disordered background dynamics of first price auctions.

For each bidder, we calculated the minimal bid perturbations necessary for the bids on all keywords to be identical. We did this by introducing the variable c_t (the theoretical bid submitted at time t, which is the same for all keywords) and solving the following quadratic program.

$$\text{For each bidder}$$
$$\text{minimize} \sum_{k,t} (x_{k,t})^2 \text{ such that}$$
$$\forall k,t.(b_{k,t} + x_{k,t}) = c_t$$

We found that 37% of bidders submitted nearly identical bids on all keywords. (More formally, 37% of bidders needed perturbations of 5% or less in order to, on each day, have identical bids on all keywords for which bids were submitted on that day.)

Not surprisingly, the percentage of bidders who bid approximately the same amount across all of their keywords varies dramatically based on how many keywords are being bid on. Figure 4 presents these results. As a cautionary note, the reader should be aware that the data set only includes bids on the top 1000 keywords, so it is possible that bidders are also bidding on keywords that were not included in the data set.

To see whether advertisers bid approximately the same amounts on each keyword across time, we solved the following quadratic program:

Number of Keywords Bid on	Percentage of Bidders who Have Approximately Identical Bids on All Keywords
2	60
3-5	35
6-25	19
>25	6
Overall	37

Figure 4: Bidders who can bid identically on all keywords when we apply bid perturbations of 5%.

$$\text{For each bidder}$$
$$\text{minimize} \sum_{k,t} (x_{k,t})^2 \text{ such that}$$
$$\forall k,t.(b_{k,t} + x_{k,t}) = c_k$$

We found that 20% of bidders submitted nearly identical bids across time. (More formally, 20% of bidders needed perturbations of 5% or less in order for each keyword's bid to be identical on all days on which a bid for that keyword was submitted.)

Figure 5 classifies bidders based on the results of this section.

We re-ran the tests from sections 6.1 and 6.2, this time removing the 45% of bidders who bid trivially (i.e. almost identically on all keywords and/or all days). Allowing perturbations of up to 5%, the new data are expressed in Figure 6. This time, only 48% of bidders met the ROI condition. This suggests that many of the bidders who meet the ROI maximization condition do so for the trivial reason that they have identical bids across time and/or keywords.

Number of Keywords Bid on	Percentage of Bidders who Maintain a Consistent Ordering of Their Bids Across Keywords	Percentage of Bidders who Meet the ROI Maximization Condition
2	90	69
3-5	85	53
6-25	62	34
>25	30	19
Overall	74	48

Figure 6: Results of the ROI tests on the 55% of bidders who change their bids by appreciable amounts from day-to-day, and have appreciable differences on their bids on different keywords. We applied perturbations of up to 5% to the bids.
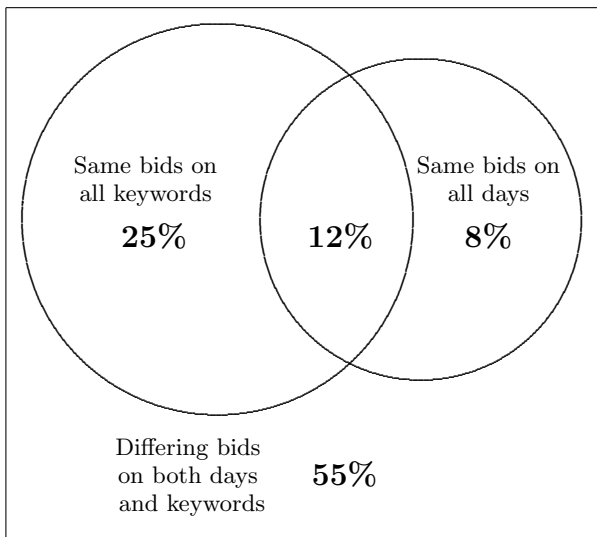

Figure 5: Classification of all bidders based on whether they bid almost identically on all keywords and/or days. Not to scale.

6.4 Correlations

In the course of our analysis of ROI maximization, we found some interesting correlations between other metrics of bidder sophistication. In the first price data set, we found weak positive linear correlations between all of the following: number of keywords bid on, frequency of bid updates, bid spread (the mean difference between bids on any two keywords) and mean bid. In general, all of these metrics are loosely connected to bidder sophistication, so these correlations make sense. The Pearson product-moment correlation coefficients for these correlations ranged between 0.08 and 0.25, so none of the correlations was strong enough to be a significant predictor of bidder behavior.

7. ROI MAXIMIZATION IN SECOND PRICE AUCTIONS

Because the two ROI tests we ran dealt with the allocation of budgets across keywords, and not with the dynamics of a specific keyword's auction, we believe that results from the ROI tests on the first price data set are indicative of results we would expect to see in a second price data set. That being said, there are some differences between the two auctions that may affect ROI performance. It is relatively simple for a bidder to maximize his ROI in a first price auction. Because bidders control exactly how much they pay per click, it is easy to maintain consistent ratios of prices to values across time. In a second price auction bidders instead pay the bid of the bidder ranked immediately behind them, so it is not so easy for bidders to precisely choose their price per click. On the other hand, some (but certainly not all) of the disordered behavior in a first price auction is absent from a second price auction. Figure 7 shows, for example, that bidders update their bids far less often in the second price data set than in the first price data set. It is possible that the increased stability of the market results in increased ROI maximization. To test these hypotheses, we ran our tests on a small second price data set. Because the data set is small and market-specific, we do not recommend drawing broad conclusions from this analysis.

In a second price auction, the consistent ordering test (of Section 6.1) and ROI maximization test (of Section 6.2) can be run using either prices or bids. Running the tests on prices fits in better with the definition of ROI, but because bidders do not have precise control over the price, we choose to present results using bids. This point is largely unimportant, however, because the results using bids are very similar to the results using prices.

Figure 8 compares the first and second price results of the

Proportion of Days On Which a Bid Is Updated	Percentage of Bidders With This Update Frequency	
	First Price	Second Price
[0, 0.2)	25	74
[0.2, 0.4)	1	8
[0.4, 0.6)	1	12
[0.6, 0.8)	2	3
[0.8, 1]	71	3

Figure 7: Bid update frequencies. A bid is considered to have been updated if it changes from one day to the next.

consistent ordering and ROI tests from sections 6.1 and 6.2 respectively. It shows that ROI maximization seems to be slightly more prevalent in the second price data set. However, these results perhaps mask the bigger issue: in the second price data set, there are far more bidders who trivially satisfy the ROI and consistent ordering conditions because they bid almost identically across time or keywords. Figure 9 shows the proportions of second price bidders that fit into these categories. Only 22% of bidders in the second price data set varied their bids by meaningful amounts across both time and keywords. When we rerun the tests using only these bidders, the results are quite different. Figure 10 represents the results of Figure 8, but with the trivial bidders removed from the data set. In this test, ROI maximization is more prevalent in the first price auction. Interestingly, however, the two types of auctions have very similar proportions of non-trivial bidders who fulfill the consistent ordering test. Some of the discrepancies may be related to the difference between the bid and the price in second price auctions.

What conclusions do we draw from this? The fact that so many bidders maintain near-identical bids across time and keywords means that it is harder to determine whether ROI maximization is occurring. This, combined with an already small data set, means that the statistics may reflect particularities of this data set more than general properties of second price auctions. Even so, we see that most of metrics are relatively similar between the first and second price tests. This is in line with our belief that the first price results are indicative of bidder behavior in second price sponsored search auctions.

8. A CHI-SQUARE STATISTICAL ANALYSIS

In this paper, we dealt with perturbations in a relatively straightforward way. We classified bidders based simply on the minimal bid perturbations they required in order to fulfill a certain condition. It would also have been possible to follow the approach of [12] and treat the perturbations in a statistical way. In this section, we explain how such an approach could be used with the perturbations we computed. We briefly summarize the approach of [12] and apply it to the analysis we used in this paper.

When we seek to establish whether a bidder meets a certain condition, we allow perturbations in order to overcome the "errors" in the bids. Here, the term "error" is used to describe the effects of background entropy (see Section 5 for a more thorough explanation) that interfere with a pure adherence to any one strategy. Because the auction entropy
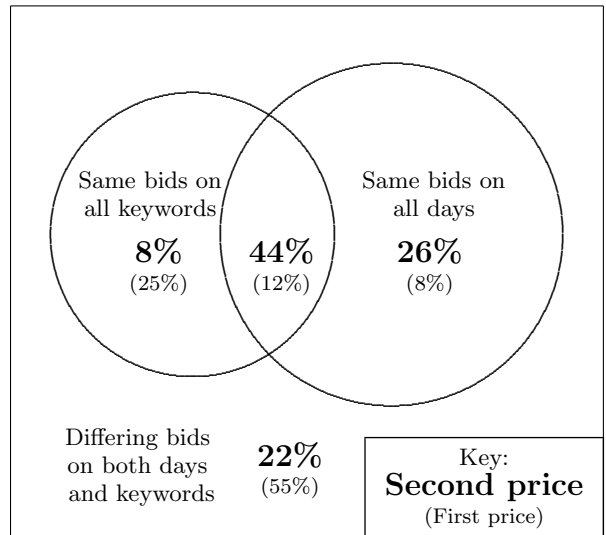

Figure 9: Classification of all bidders in the second price data set based on whether they bid almost identically on all keywords and/or days. Also includes corresponding numbers from the first price data set. (Not to scale.)

is caused by a large number of factors that are essentially acting independently, the error will follow an approximately Gaussian distribution (the central limit theorem could be used for a formal justification). Hence, we could run a standard chi-square goodness-of-fit test to see whether the empirical data are consistent with the theoretical condition we propose. An explanation of such a test follows.

Let $Y = \langle y_1, y_2, \cdots, y_n \rangle$ be a vector of actual bids, and let $Z = \langle z_1, z_2, \cdots, z_n \rangle$ be a vector of bids proposed by a theoretical model. Our null hypothesis is that the data are consistent given a reasonable tolerance on the bids. In other words, the actual bids are "close enough" to the theoretical bids that we accept the model that was used to generate the theoretical bids.

To see whether we should reject this null hypothesis and hence conclude that the theoretical model is insufficient, we compute the χ^2 test statistic

$$T = \sum_{i=1}^{n} \frac{(z_i - y_i)^2}{\sigma^2}.$$

Observe that the equation $z_i - y_i$ is the perturbation required for bid i to equal its theoretical value. If we know σ (the root mean squared error), we can then solve for T. If T is above a threshold C_α that is determined by a certainty level α and the number of degrees of freedom, then we reject the null hypothesis. If we do not know σ, we can fix T based on our desired certainty level, and compute the minimal root mean square error necessary to avoid rejecting the null hypothesis. Because we assume that the bids are unbiased (i.e. they are just as likely to need to be perturbed upwards as they are to need to be perturbed downwards), the root mean squared error is equivalent to the standard error, which is the standard deviation on the error. So if, for example, we expect the background auction entropy to

Number of Keywords Bid On	Percentage of Bidders who Maintain a Consistent Ordering of Their Bids Across Keywords		Percentage of Bidders who Meet the ROI Maximization Condition	
	First Price	Second Price	First Price	Second Price
2	96	100	84	97
3-5	90	100	68	88
6-25	68	88	45	66
>25	36	60	26	44
Overall	83	90	65	79

Figure 8: Comparison of the results from the *consistent ordering test* (Section 6.1) and the *ROI maximization test* (Section 6.2) on first price and second price data sets.

Number of Keywords Bid On	Percentage of Bidders who Maintain a Consistent Ordering of Their Bids Across Keywords		Percentage of Bidders who Meet the ROI Maximization Condition	
	First Price	Second Price	First Price	Second Price
2	90	100	69	57
3-5	85	100	53	59
6-25	62	79	34	42
>25	30	42	19	18
Overall	74	70	48	37

Figure 10: The same as Figure 8, but with trivial bidders (those who submit almost identical bids for all days and/or keywords) removed from the data set.

affect bids by an amount with a standard deviation of x, then we set the standard error to be x.

Of course, the conditions we tested for did not propose exact bids. Rather, they posited certain relationships between bids across time and keywords. To see whether such models are consistent with real-world auction data, we set the theoretical bids to be minimally perturbed real-world bids such that the theoretical test condition holds. For example, if we were testing to see whether the ROI maximization condition holds, then we would set the theoretical bids to be the minimally perturbed empirical bids that resulted in the ROI maximization condition holding. Thus, the minimal possible test statistic is

$$S = \begin{array}{l} \min \sum_{i=1}^{n} \frac{(\jmath_i - y_i)^2}{\sigma^2} \\ \text{such that } \langle \jmath_1, \jmath_2, \cdots, \jmath_n \rangle \text{ is} \\ \text{consistent with the theoretical model} \end{array} \quad (4)$$

Clearly, if S is greater than the critical value C_α, then even with the theoretical data that most resemble the empirical data, we must reject the hypothesized condition. In this case, we conclude that the theoretical model is insufficient.

Although we ran this type of analysis, we chose not to present it in this paper, because we believe that the raw perturbations provide a sufficient and intuitive basis for classifying bidders. (The same decision was made by Varian in [13].) The results of the chi square tests and the results of our informal tests had similar trends.

9. SUMMARY AND CONCLUSIONS

We found evidence that ROI-based strategies may have been employed by a large fraction of bidders in the Yahoo! first price data set. We found that when permitting bid perturbations of up to 5%, 83% of bidders ordered their bids in the same way on all days (a primitive form of ROI maxi-

mization), and 65% of bidders met necessary conditions for full ROI maximization. (Or, alternatively, 35% of bidders are not maximizing their ROI.)

The evidence in a second price auction was less conclusive. We found that over three quarters of bidders meet the ROI maximization condition, but most bidders who meet this condition do so because they have near-identical bids across time or keywords.

Our tests can be applied on a bidder-by-bidder basis to show specific areas in which bidders tend to lack sophistication. Approximately half of all bidders in the first price data set bid almost identically on all days, or keywords, or both. Of the bidders who do bid appreciably different amounts on different days and different keywords, a slight majority appear to not be optimizing their ROI.

Our conclusions are twofold. First, because of this empirical evidence, we suggest that an ROI-based model may be useful when investigating the equilibria of sponsored search auctions, or when modelling the market. On the other hand, in both types of auctions, a large fraction of bidders who cast meaningfully different bids across time and keywords were unable to achieve ROI. This might have been because they were not budget-constrained, because they did not have an accurate assessment of their true values per click on different keywords, or because they lacked sophistication. If it was due to one of the latter two reasons, improved bidding agents or a different market design might result in a more economically efficient market.

10. REFERENCES

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

[2] E. Bice. Sponsored search seminar project. Seminar on Sponsored Search at University of Pennsylvania, Instructor: Michael Kearns, 2006.

[3] C. Borgs, J. T. 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*, pages 531–540, 2007.

[4] M. Cary, A. Das, B. Edelman, I. Giotis, K. Heimerl, A. R. Karlin, C. Mathieu, and M. Schwarz. Greedy bidding strategies for keyword auctions. In *EC '07: Proceedings of the 8th ACM conference on Electronic commerce*, pages 262–271, New York, NY, USA, 2007. ACM.

[5] D. Chakrabarty, Y. Zhou, and R. Lukose. Budget constrained bidding in keyword auctions and online knapsack problems. In *WWW Workshop On Sponsored Search Auctions*, 2007.

[6] B. Edelman and M. Ostrovsky. Strategic bidder behavior in sponsored search auctions. *Decis. Support Syst.*, 43(1):192–198, 2007.

[7] B. Kitts and B. Leblanc. Optimal bidding on keyword auctions. *Electronic Markets*, 14:186–201(16), September 2004.

[8] S. Lahaie, D. M. Pennock, A. Saberi, and R. V. Vohra. Sponsored search auctions. In N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*, chapter 28, pages 699–716. Cambridge University Press, 2007.

[9] P. A. Samuelson. A note on the pure theory of consumer's behaviour. *Economica*, 5(17):61–71, 1938.

[10] B. Szymanski and J. Lee. Impact of roi on bidding and revenue in sponsored search advertisement auctions. In *Second Workshop on Sponsored Search Auctions*, 2006.

[11] H. R. Varian. The nonparametric approach to production analysis. *Econometrica*, 52(3):579–97, 1984.

[12] H. R. Varian. Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics*, 30(1-2):445–458, 1985.

[13] H. R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, December 2007.