The Role of Search Engine Optimization in Search Marketing *

Ron Berman and Zsolt Katona †

January 22, 2011

*We thank Pedro Gardete, Ganesh Iyer, Shachar Kariv, John Morgan, Miklos Sarvary, Dana Sisak, Kenneth Wilbur, Yi Zhu and seminar participants at HKUST, University of Florida, University of Houston, UT Austin, and Yale for their useful comments.

†Ron Berman is a Ph.D. student and Zsolt Katona is an Assistant Professor at the Haas School of Business, UC Berkeley, 94720-1900 CA. E-mail: ron_berman@haas.berkeley.edu, zskatona@haas.berkeley.edu.

The Role of Search Engine Optimization in Search Marketing

Abstract

Web sites invest significant resources in trying to influence their visibility among online search results. In addition to paying for sponsored links, they invest in methods known as search engine optimization (SEO) that improve the ranking of a site among the search results without improving its quality. We study the economic incentives of Web sites to invest in SEO and its implications on search engine and advertiser payoffs. We find that the process is equivalent to an all-pay auction with noise and headstarts. Our results show that, under certain conditions, a positive level of search engine optimization improves the search engine's ranking quality and thus the satisfaction of its visitors. In particular, if the quality of sites coincides with their valuation for visitors then search engine optimization serves as a mechanism that improves the ranking by correcting measurement errors. While this benefits consumers and increases traffic to the search engine, sites participating in search engine optimization could be worse off due to wasteful spending unless their valuation for traffic is very high. We also investigate how search engine optimization affects the revenues from sponsored links. Surprisingly, we find that in many cases search engine revenues are increased by SEO.

1 Introduction

Online search engines are among the most popular tools that consumers use to discover information on the Web. As a result, search engine marketing is becoming a dominant form of online advertising. By utilizing search marketing, Web sites that wish to expose their content and merchandize to consumers can reach them when they search for specific keywords providing invaluable targeting opportunities.

In order to accommodate advertisers, most search engines have divided their search results page into an *organic* and a *sponsored* part. The left side of the screen is typically used to display organic results as a ranked list of site links ordered according to their relevance for the search query. The parts above and to the right of the organic results are often used to display sponsored links which are typically auctioned to advertisers using various mechanisms. Selling sponsored links is typically the leading revenue stream for search engines, and in many cases it is the only revenue stream. During the sales process, advertisers submit bids for having their ads placed among the sponsored links, and generally the highest bidders win the most visible links[1], usually on the top of the list.

In addition to buying sponsored links, many websites and advertisers try to find their way to the top of the organic results list by influencing the search engine's ranking algorithm. Since the organic links are viewed by consumers as more trustworthy, websites receive positive benefits from visitors arriving through clicks on them. The collection of different actions that a site can take to improve its position on the organic list is called *search engine optimization* (SEO). Improving one's position can be accomplished either by making the site more relevant for consumers, or by investing in techniques that affect the search engine's quality ranking process solely. These two types of SEO techniques are sometimes referred to as *white hat SEO* and *black hat SEO* respectively. The important difference is that white hat SEO improves the site content, thus increasing visitor satisfaction and making the site more relevant, while black hat SEO only improves the ranking of a site among search results without affecting its quality.

[1]The more sophisticated auction mechanisms also take into account parameters such as the the likelihood of a click on a given link, or the quality of the landing page the ad leads to, estimating them from historical click-through data.

Examples of black hat techniques are eliciting external linking to the site or changing the site's pages to manipulate the ranking process of the search engine. Our focus in this paper is on black hat SEO methods, which we plainly call *SEO*. White hat methods can be seen as purely content investments, and we refer to them briefly in our concluding remarks.

Influencing the relevance measurement of search engines requires an investment of resources, many times in the form of a service company being hired to perform SEO. Search engines typically take a stance against black hat SEO and consider it cheating. In some cases, websites caught conducting SEO activities are removed from the organic list[2]. To set the rules, search engines sometimes publish guidelines describing undesired practices. Google, for example, prohibits buying incoming links to increase one's PageRank[3]. Yahoo, on the other hand, simply does not give weight to a paid link if they think it is not valuable to consumers[4]. In addition to simply stating what they consider allowable, search engines can also invest significant amounts in reducing the effectiveness of certain SEO activities[5].

To justify their position, search engines typically claim that manipulation of search engine results hurts consumer satisfaction and decreases the welfare of "honest" sites. In contrast to that, search engines also convey a puzzling message that the auction mechanism for sponsored links ensures that the best advertisers will obtain the links of highest quality, resulting in higher social and consumer welfare. Is not the case of SEO similar? If the most resourceful sites are the ones providing the best links, why not let them invest in improving their rankings?

We stipulate that a major reason for search engines' reluctance in allowing SEO is the trade-off advertisers face between investing in sponsored links and investing in influencing organic rankings. As a result, search engines may be unhappy if sites spend significant amounts on SEO activities instead of on paid links and content creation. One possible solution is to allow payments for organic links and to pocket the money that sites would have otherwise paid to

[2]BBC News reported that Google has blacklisted BMW.de for breaching its guidelines. See http://news.bbc.co.uk/2/hi/technology/4685750.stm

[3]Google Webmaster Central: http://www.google.com/support/webmasters/bin/answer.py?answer=66356

[4]Interview with Priyank Garg, director of product management for Yahoo! Search Technology: http://www.stonetemple.com/articles/interview-priyank-garg.shtml

[5]In response to Google's regular updates of its search algorithm, different sites shuffle up and down wildly in its search rankings. This phenomenon, which happens two or three times a year is called the "Google Dance" by search professionals who give names to these events as they do for hurricanes (see "Dancing with Google's spiders", *The Economist*, March 9, 2006).

third parties. An example of such an implementation, Baidu, the leading Chinese search engine and the world's third largest, does accept payments for organic links[6].

The above examples show that it is not clear what role search engine optimization plays in the online advertising ecosystem and whether it is necessarily detrimental. Our goal is to explore the economics of the SEO process and its effects on consumers, advertisers and search engines. Specifically, we focus on the interaction between investment in SEO and in sponsored links, and the resulting effects. By doing so, we are able to uncover the conditions under which manipulation of ranking results is harmful to consumers[7] and other stakeholders. We are also able to provide recommendations to search engines on SEO policy, and to advertisers on how to optimally invest in or against SEO.

Our main results in Section 4 show that black hat SEO can be advantageous to the search engine and can increase traffic and consumer welfare in equilibrium. In particular, if sites' valuation for traffic is aligned with their relevance (quality) then consumers are better off with some positive level of SEO than without, resulting in a higher traffic to the search engine. If, on the other hand, there are sites which extract high value from visitors yet provide them with little value then SEO is generally detrimental to the search engine and consumer welfare. An example of such a "bad" site, which is often called a *spam site*, is a site that advertises products for a very low price to lure visitors, but later on uses the visitors' credit card details for fraudulent activities[8].

According to our main results, SEO can be beneficial to consumers under some conditions by moving the higher quality sites higher among the organic results. Although this may result in higher traffic to the search engine, it is not clear what implication it has on the search engine's profit from sponsored links. Normally, more traffic to the search page implies higher revenues for the search engine. Moreover, the size of the audience depends on the quality of the service which in this case is the quality of search results visitors can expect. This logic yields that search engines should offer the highest quality organic results to maximize revenues. However, as the organic and sponsored lists are competing for consumer attention and the same sites

[6]Baidu scandal makes it to CCTV: http://shanghaiist.com/2008/11/23/baidu_scandal_makes_it_to_cctv.php
[7]Wilbur and Zhu (2009) study click fraud driven by a similar motivation.
[8]Researchers estimate (Benczur et al. 2008) that 10-20% of Web sites constitute spam.

may appear on both lists, the search engine would have an incentive to offer suboptimal results on the organic side (White 2009, Taylor 2010).

It is not clear how the search engine should resolve this potential conflict between the two lists. Even though organic links bring in visitors, if the results are too satisfactory they do not click on the sponsored links. Furthermore, if advertisers receive enough visits through organic links, they potentially have lower willingness to pay for sponsored visits. Thus, one may argue that advertisers who spend resources on SEO, will spend less on sponsored links. We investigate this problem in Section 5 where we study the interaction between the SEO contest for organic links and the auction for sponsored links. We find that, surprisingly, SEO not only leads to higher traffic in many cases, but also increases search engine revenues under certain conditions. This follows from a more general finding that sponsored revenues are not necessarily hurt by better quality organic link. Indeed, when the best quality site acquires the top organic link, the second best site might pay a higher price for the sponsored link than in the reverse case. As part of the SEO profitability analysis we identify the exact conditions under which providing high quality organic links is profitable for the search engine.

Despite the apparent importance of the topic, there has been very little research done on search engine optimization. At the same time, search engine optimization has grown to become a multi-billion dollar business[9]. Many papers have focused on the sponsored side of the search page and some on the interaction between the two lists. In all of these cases, however, the ranking of a website in the organic list is given as exogenous, and the possibility of investing in SEO is ignored, although marketers often face the problem whether and how much to invest in SEO. Our results provide useful insights to firms involved in search marketing and to search engines. Furthermore, given that search engines are a major gateway for information discovery, there is an emerging debate on the fairness of search results and ranking algorithms, and the possibility of regulating search engines. We hope our results will contribute to this discussion.

The rest of the paper is organized as follows. Section 2 gives an overview of a small selection of this diverse literature and other research areas that are relevant to our work. In Section 3 we describe the main model and in Section 4 we present the equilibrium outcomes of a simplified

[9]See the survey conducted by seomoz.com at http://www.seomoz.org/dp/seo-industry-survey-results.

case with only one organic link. Next, we examine the interaction between the SEO game and the sponsored auction in Section 5. Finally, Section 6 generalizes our model in several ways to show that our main results are robust, and introduces new results on analysis of contests for multiple items with asymmetric players. All proofs and technical details appear in the Appendix.

2 Relevant Literature

The rapid growth of the online advertising industry led to a surge in research dedicated to this phenomenon. We review just a select few of the large volume of works related to our research. Works such as those by Rutz and Bucklin (2007) and Ghose and Yang (2009) focus on consumer response to search advertising and the different characteristics that impact advertising efficiency. Another major stream of research, including works by Edelman et al. (2007), Varian (2007), Zhu and Wilbur (2010) and Athey and Nekipelov (2010) focus mostly on the auction mechanism used by the different search engines to allocate their advertising slots. Other recent examples, such as those by Chen and He (2006), Athey and Ellison (2009) and Aggarwal et al. (2008) analyze models that include both consumers and advertisers as active players.

A number of recent papers study the interplay between the organic list and the sponsored list. Katona and Sarvary (2010) show that the top organic sites may not have an incentive to bid for sponsored links. In an empirical piece, Yang and Ghose (2010) show that organic links have a positive effect on click-through-rates on paid links, potentially increasing profits. Taylor (2010), White (2009) and Xu et al. (2009) study how the incentives of the search engine to provide high quality organic results are affected by potential losses on sponsored links. The general notion is that search engines have an incentive to provide lower quality results in order to maximize revenues.

Our model makes extensive use of methodologies related to all-pay auctions and contests. For a survey of the literature on contests[10], see Konrad (2007) and Sisak (2009). Specifically, our analysis takes into account asymmetries among websites as well as ranking error of the search engine. Kirkegaard (2009) describes the equilibria in contests with asymmetric players,

[10]Contests are all-pay competitions among bidders where the bids do not reach the auctioneer

while Siegel (2009) analyzed such games under more general conditions. Our application is unique in that it considers the cases where the initial asymmetry is biased by noise inherent in the quality measurement process. Krishna (2007) and Athey and Nekipelov (2010) are two of the few examples taking noise into consideration in an auction setting. This noise is the main reason for the initially inefficient allocation of organic link slots, which can be corrected by allowing for SEO.

Little attention was given to search engine optimization, although the use of SEO techniques is common practice. Works such as those by Pasquale (2006) Bracha and Pasquale (2007) and Mercadante (2008) consider the implications of search results manipulation using the traditional view that "cheating" has strictly negative results. Different options for regulation and the need (and legality) for it are examined. The economic implications of using bribes in contests is analyzed by Clark and Riis (2000). An important result is that allocative efficiency is not necessarily degraded by a bribery procedure, but might increase depending on the contest's parameters.

The work of Xing and Lin (2006) resembles ours the most by defining "algorithm quality" and "algorithm robustness" to describe the search engine's ability to accurately identify relevant websites. Their paper shows that when advertisers' valuation for organic links is high enough, providers of SEO services are profitable, while search engines' profits suffer. The strategic effect of the increase in consumer's satisfaction in not taken into account in their analysis. An earlier work by Sen (2005) develops a theoretical model that examines the optimal strategy of mixing between investing in SEO and buying ad placements. The model surprisingly shows that SEO should not exist as part of an equilibrium strategy.

3 Model

We set up a static game to model the competition of advertisers for consumers who search for a specific key phrase online. We assume there is a monopolistic search engine (SE) that provides search results to consumers by displaying links to one of n websites that can also buy sponsored links from the search engine. Whenever a consumer enters the search phrase, the search engine ranks the different websites according to a scoring mechanism, and presents the n links ranked according to their scores. The search engine, advertisers, and consumers each have different incentives and characteristics affecting their decisions as described below.

3.1 The Search Engine

A search engine is a website that provides searching results as a service to its visitors: they enter queries (search phrases) into a search form and the SE returns a k number of results for this query displaying them in an ordered list. This list - often referred to as the *organic list* - contains a number of links to other websites in the order of the relevance of their content for the given search phrase. In our model, we focus on a single keyword and we assume that the *relevance*, or *quality* of a search result is essentially the probability that a consumer is satisfied with the site once clicking on the link[11]. In addition to the organic results, the search engine displays m *sponsored links* to generate revenue. Sponsored links are typically displayed above and rightward of the organic results for a search query and look similar to the search results, but are clearly marked as advertisement. These links are sold to advertisers through an auction in which they submit bids and are awarded different positions on the page. The outcome of the auction is usually determined by the order of the bids - corrected for the differences in the likelihood that consumers click on a particular link - and each advertiser pays the next highest (corrected) bid. We assume that there is a second price auction to determine the allocation of the sponsored link and the highest bidder receives the link.

In order to rank websites, the search engine uses information gathered from crawling algorithms and data mining methods on the Internet. Let q_i denote the relevance of site i in the context of a given keyword. It is reasonable to assume that the search engine can only measure quality with an error, and cannot observe it directly. The initial quality score that the SE assigns to site i is thus $s_i^S = q_i + \sigma\varepsilon_i$, where ε_i are assumed to be independent and are drawn from the same distribution and σ is a scaling parameter measuring the standard deviation of the error. If the Web sites do not take any action the results will be ordered according to the s_i^S's as assigned by the search engine. If, however, Web sites can invest in SEO, they have

[11] The results remain unchanged if we assume that quality is the expected utility a random consumer gets when clicking on a link

the option to influence their position after observing the initial scores[12]. The effectiveness of SEO is measured by the parameter α in the following way. If site i invests a b_i amount in SEO, its final score becomes $s_i^F = s_i^S + \alpha b_i$. That is, depending on the effectiveness of SEO, sites can influence their score to a varying extent which in turn determines their final location in the organic list of search results. The parameter α essentially measures how easy it is to change one's ranking using SEO methods. That is, $1/\alpha$ can be interpreted as the cost of SEO which, among other factors, can be influenced by the search engine. Indeed, if the search engine ignores the possibility of SEO activities, α presumably increases. We do not explicitly model the search engine's decision to invest in changing α, we rather compare the SE's payoff under different α's providing a recommendation on the optimal values.

We assume that the SE receives T amount of traffic that is a function of how satisfied consumers are with the search results. Let U denote the expected satisfaction of a consumer from visiting the search engine[13]. Then, it is natural to assume that $T = f(U)$ where $f()$ is an increasing function, since the higher the expected satisfaction of a visitor the higher the likelihood of a visit. We assume that the expectations are rational in that they match the actual satisfaction level of the consumers. However, one could imagine a situation in which it is hard for consumers to form reasonable expectations - that would be captured by a constant $f()$ function. Essentially, $f()$ measures consumers' sensitivity to the quality they can expect at the search engine and their ability to form correct expectations. Using the above notation, the search engine's profit is

$$\pi_{SE} = \int_0^T \pi_S(t)dt, \tag{1}$$

where $\pi_S(t)$ is the SE's revenue from sponsored links for visitor t.

3.2 Websites/Advertisers

We assume that there are n websites providing informational content or products to consumers and that those sites derive some utility from the visiting consumers. We index the sites in

[12]We do not explicitly model who conducts the SEO activities. It could be the site itself, a third party, or in the extreme case, the search engine itself.

[13]This includes expected utility from visiting sponsored links. Our results are very similar if we assume that consumers only include organic links in their expectations.

decreasing order of their quality: $q_1 \geq q_2 \geq \ldots q_n$. The quality of a site is essentially the probability that a consumer that visits the site is satisfied[14]. The sites' profits primarily depend on their traffic. We assume that site i has a valuation of $R_i(t)$, for t amount of clicks. Let $r_i(t) = R_i'(t)$ be the incremental valuation of site i for an extra click, where we assume that $R_i()$ is differentiable and $r_i()$ is non-negative and weakly decreasing. In essence, we assume that each site has a positive valuation for any click, but the valuation is non-increasing. This is a relatively flexible setup, as it allows us to incorporate traditional decreasing returns arguments, but also to capture a typical scenario of the digital world, in which there is no difference between two clicks. Furthermore, it also allows us to capture advertisers that have a steady valuation for successive clicks but are bound by cash-flow limitations and operate on a fixed budget. An advertiser has two types of possible costs – investment in SEO and buying sponsored links. We let b_i be the investment in SEO and p_i be the cost of buying sponsored links. The resulting profit function for each advertiser is thus:

$$\pi_i = R_i(t_i) - b_i - p_i \tag{2}$$

The traffic t_i a website experiences depends on the behavior of consumers using the search engine, which we now describe.

3.3 Consumers

The behavior of consumers in our model is relatively simple, but captures different behaviors identified by the literature. Most papers assume that consumers click passively without evaluating the differences between links. Recent papers (Chen and He 2006, Athey and Ellison 2009, Yang and Ghose 2010, Jeziorski and Segal 2009, Yao and Mela 2010) point out that visitors exhibit utility maximizing characteristics. To incorporate this in our model, we assume that a consumer may be able to make inferences about the quality of the links without clicking on them, thus, we assume that a consumer is sophisticated with probability ψ and selects the link that offers the highest quality[15]. If the consumer is not able to determine which is the

[14] One can imagine a detailed model of consumer satisfaction, based on consumer heterogeneity in tastes and preferences. We use this basic $0 - 1$ setup to capture the main differences between sites' qualities.

[15] If the same link appears on both the organic and sponsored list, as a tie-braking rule, we assume that a sophisticated consumer selects the organic link, but this assumption is not crucial for our results.

highest quality link (with probability $1-\psi$) then he or she randomly chooses a link on either the organic list or the sponsored list. Naturally there are differences in probabilities based on a link's position in the list and whether the link is organic or sponsored. The probability with which a non-sophisticated consumer clicks on link i in the organic list is $\gamma\beta_i$ and $(1-\gamma)\beta_i$ in the sponsored list. We naturally assume that $\beta_1 \geq \beta_2 \geq ... \geq \beta_n$ and $\sum_{i=0}^{\infty} \beta_i \leq 1$. That is γ captures the proportion of non-sophisticated clicks on the organic side, whereas β_i capture the order effects. Once the consumer clicks on a link that belongs to site j s/he is satisfied with probability q_j, receiving a utility normalized to 1 if satisfied. To determine the number of clicks received by a site, let Φ_i^j denote an indicator that takes the value of 1 if site i is located in the organic position j and 0 otherwise, let χ_i^j be the same for the jth sponsored link, and finally let Ψ_i denote the indicator that takes a value of 1 if site i is the highest quality site on the entire page. With these, the traffic received by site i is

$$t_i = T\left[\psi\Psi_i + (1-\psi)\left(\gamma\sum_{j=1}^{k}\beta_j\Phi_i^j + (1-\gamma)\sum_{j=1}^{l}\beta_j\chi_i^j\right)\right]. \tag{3}$$

3.4 Timing

The sequence of the game is the following. First the search engine measures the relevance of each website and publishes $s_i^S = q_i + \sigma\varepsilon_i$. Next, the websites, after observing s_i^S, simultaneously decide how much to invest in SEO, changing the scores to $s_i^F = s_i^S + \alpha \cdot b_i$. The search engine then recalculates the scores and displays an ordered list of search results sorted in a decreasing order of the final site scores s_i^F. When the organic ranking has been settled, advertisers bid for the sponsored links and participate in a second price auction that determines the sponsored link to be shown to all visitors[16]. Each visitor clicks on the results according to the sequence defined above and payoffs are realized at the end. Our assumption on the timing of the above events is somewhat simplistic, but it is the most plausible way of capturing Web sites' reactions to their ranking results and their subsequent investment in SEO and bids for sponsored links. Later, in Section 6, we relax our assumption on the information structure and employ an incomplete

[16]One could assume that advertisers can adjust their bids or, at the extreme, bid for each click separately. This setup would not change our results substantially, but simplifying assumptions need to be made to find an equilibrium in the sequential auctions.

information setting. There, we assume that the search engine performs a measurement of the quality each time a ranking is performed, and the error is not public. Websites then have expectations about the error structure and spend their SEO efforts and make their bids in advance.

Variable	**Notation**
Number of sites	n
Number of organic links	k
Number of sponsored links	m
Standard deviation of ranking error	σ
Initial quality score assigned by SE to site i	s_i^S
Final quality score assigned by SE to site i	s_i^F
Effectiveness of SEO (inverse of cost of SEO)	α
Quality of site i	q_i
Valuation of site i for a single link	v_i
Net revenue of site i from click	$R_i(t_i)$
Ranking error for site i	ε_i
SEO investment by site i	b_i
Total amount paid by site i for sponsored link	p_i
Indicator denoting whether site i has organic link j	Φ_i^j
Indicator denoting whether site i has sponsored link j	χ_i^j
Indicator denoting whether site i has the highest quality	Ψ_i
Proportion of sophisticated consumers	ψ
Proportion of organic clicks	γ
Click-through rate of position j in any list	β_j
Expected consumer satisfaction	U
Number of clicks to site i	t_i
Total traffic at SE	$T = f(U)$

Table 1: Summary of Notation

We start our analysis by examining a simple case that illustrates the main forces governing SEO. Here, we do not consider sponsored links (assume $m = 0$ or $\gamma = 1$), but rather focus on the core effect of SEO on the organic ranking and its effects on search engine traffic. For the sake of simplicity, we assume that there is only one organic link displayed on the SE ($k = 1$, also assuming $\beta_1 = 1$) and that there are two bidders ($n = 2$). We examine the effects of the presence of sponsored links in Section 5 and generalize to the case of $n > 2$ sites, and multiple $k > 1$ links in Section 6.

4 SEO Equilibrium - One Organic Link

Assume that two Web sites compete for a single organic link and let the distribution of ε_i take the values of 1 or -1 with equal probabilities. We assume $\sigma > |q_1 - q_2|/2$ to ensure that the error can affect the ordering of sites, otherwise the error never changes the order of results and the setup is equivalent to one with no error. Although very simplistic, this setup allows us to derive our main results on the forces governing the effects of SEO on the ranking of sites and their payoffs. Let v_1 and v_2 denote the sites' valuations for winning the auction, as derived from their valuation functions and the traffic functions:

$$v_1 = R_1(f(q_1)), \quad v_2 = R_2(f(q_2))$$

As a benchmark, let us examine the case in which search engine optimization is *not* possible, i.e., when $\alpha = 0$. In this case sites cannot influence their position among the search results. Since $q_1 \geq q_2$, the expected traffic arriving to the SE is $\frac{3}{4}f(q_1) + \frac{1}{4}f(q_2)$. The error in measurement thus causes a suboptimal allocation of the search link when compared to the first best case of $f(q_1)$, leading to a drop in traffic.

When search engine optimization becomes effective, i.e., when $\alpha > 0$, websites have a tool to influence the order of results. The ability to influence, however, is typically asymmetric, since sites have different starting scores s_i^S. A site that is in the first position in the SE's initial ranking has a headstart and hence can remain the first even if it invests less in SEO than its competitor. Another characteristic of the game sites play is that their SEO investment is sunk no matter what the outcome of the game is. Our first result states that sites essentially participate in an all-pay auction with headstarts.

Lemma 1 *The game that sites play after observing their starting scores is equivalent to an all-pay auction with headstarts.*

Appendix A discusses all-pay auctions with headstarts in detail, together with their equilibrium characteristics and their application to our game. All-pay auctions with complete information typically do not have pure-strategy Nash-equilibria. In a simple auction with two

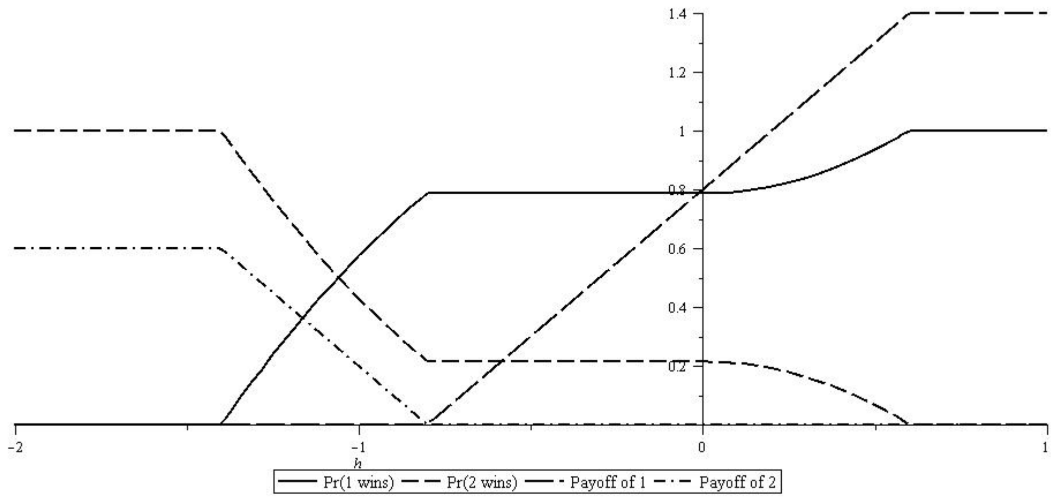

Figure 1: The figure characterizes the mixed strategy equilibrium of an all-pay auction as a function of the headstart of player 1. Sites' valuations are $v_1 = 1.4$ and $v_2 = 0.6$. The probability that player 1 (player 2) wins is weakly increasing(decreasing) in the headstart, similarly to the payoffs.

players with valuations $v_1 > v_2$, both players mix between bidding 0 and v_2 with different distributions[17]. Player 1, having the higher valuation, wins with the higher probability of $v_1/2v_2$ and player 2's surplus is 0. Thus, only the player with the highest valuation makes a positive profit in expectation, but the chance of winning gives an incentive to the other player to submit positive bids. In the case of an all-pay auction with headstarts the equilibrium is very similar and the player with the highest potential score (valuation plus headstart) wins with higher probability and the other player's expected surplus is 0. The winner's expected surplus is equal to the sum of differences in valuations and headstarts. Figure 1 illustrates the probabilities that the two sites win and their payoffs as a function of the headstart.

4.1 The effect of SEO on traffic and consumer welfare

To examine the outcomes of the SEO game, we use $ET(\alpha) = ET(\alpha; \sigma, v_1, v_2, q_1, q_2)$ to denote the expected traffic the search engine receives based on different levels of SEO effectiveness and valuations. In this simple case the probability that the player with the more relevant link wins

[17]See the Appendix for detailed bidding distributions.

the auction (that is, player 1) directly determines the expected traffic as

$$ET(\alpha; \sigma, v_1, v_2, q_1, q_2) = f(q_2) + (f(q_1) - f(q_2))\,\mathrm{Pr}(\text{Site 1 receives the organic link}).$$

Note that the above probability is very closely related not only to the amount of traffic the search engine receives, but also to the a priori expected consumer satisfaction which is $EU = q_2 + (q_1 - q_2)\,\mathrm{Pr}(\text{Site 1 wins})$. For further reference, let $P(\alpha) = P(\alpha; \sigma, v_1, v_2, q_1, q_2)$ denote the probability of the desired outcome, that is, that site 1 receives the organic link.

If there is no SEO, that is, when $\alpha = 0$ (and $\sigma > |q_1 - q_2|/2$), we have $P(0) = 3/4$. Our goal is therefore to determine whether the probability exceeds this value for any positive α SEO effectiveness levels. It is useful, however, to begin with analyzing how the probability depends on valuations and qualities for given α and σ values. The following Lemma summarizes our initial results.

Lemma 2 *For any fixed α and σ, $P(\alpha; \sigma, v_1, v_2, q_1, q_2)$ is increasing in v_1 and q_1 and is decreasing in v_2 and q_2.*

Thus, the probability of the desired outcome increases when the most relevant site becomes even more relevant and also when its valuation for clicks increases. When there is no SEO, i.e., when $\alpha = 0$, the Lemma holds because the efficiency simply does not change with v_1, v_2, q_1, q_2, but when $\alpha > 0$ the efficiency strictly increases and decreases in the respective variables. In essence, the Lemma tells us that no matter how effective SEO is, the less sites valuations are aligned with their relevance levels, the less efficient the rankings are.

The following proposition summarizes the main result of this section, showing how SEO affects the traffic of the search engine.

Proposition 1

1. *For any $\sigma > |q_1 - q_2|/2$, there exists a positive $\hat{\alpha} = \hat{\alpha}(\sigma, v_1, v_2, q_1, q_2)$ SEO effectiveness level such that $ET(\hat{\alpha}) \geq ET(0)$.*

2. *If $v_1/v_2 > 3/2$ then for any $\sigma > |q_1 - q_2|/2$, there exists a positive $\hat{\alpha} = \hat{\alpha}(\sigma, v_1, v_2, q_1, q_2)$ such that $ET(\hat{\alpha}) > ET(0)$.*

3. If $v_1 < v_2$ *and* $\sigma \geq \frac{v_2+v_1}{v_2-v_1}\frac{q_1-q_2}{2}$ *then for any* $\alpha > 0$ *we have* $ET(\alpha) \leq ET(0)$.

The first part of the proposition tells us that for any level of error there is a positive level of SEO that does not reduce the traffic of the search engine. Practically, if the level of SEO effectiveness is very low (that is, its cost is high) then firms will not invest and the ranking will not be altered.

The second point yields more interesting results. Essentially, it demonstrates that positive levels of search engine optimization do improve the the ranking in some cases. When high quality sites value visitors relatively high compared to lower quality sites, SEO is beneficial to both the search engine and consumers regardless of the level of error. The third part states the opposite: when the player with the lower quality has a higher valuation then all positive levels of SEO are detrimental to the efficiency (unless the error is very small). Although our model assumes given deterministic valuations and qualities, one could imagine a setting where these parameters are randomly drawn from a given distribution. In that case, our results yield that the higher the correlation between a site's quality and valuation, the more likely that SEO improves efficiency[18].

The intuition for the results is as follows. The SEO mechanism favors bidders with high valuations. Since the SE cannot perfectly measure site qualities, this mechanism corrects some of the error when valuations increase monotonically in quality. Furthermore, since valuations for the organic link are based on the total expected traffic, they include a component that is based on the site's quality. Thus, a higher quality site has a valuation advantage even if its valuation for a single click is the same as its low quality competitor's. For example, if valuations for additional clicks are not declining and are the same for all sites ($r_i(x) \equiv r$), then $v_i = r \cdot f(q_i)$. Substituting in the results yields that SEO can be beneficial if $f(q_1)/f(q_2) > 3/2$. Thus, if quality differences are high enough, the contest creates a force that favors high quality sites even if the valuations for a single click are similar across sites.

When lower quality sites have very high valuations for traffic, however, SEO creates incentives that are not compatible with the utilities of consumers or the search engine. In this latter

[18] If the joint distribution of v_i and q_i is lognormal then one can show that the condition for part 2 is more likely to hold when the correlation between v_i and q_i is higher.

case, the high valuation sites which are not relevant can get ahead by investing in SEO unless the error is relatively small[19]. Examples are cases of "spammer" sites that mislead consumers. In these cases consumers do not gain any utility from visiting such sites, but the sites may profit from consumer visits.

Our result describe the outcomes in different cases and show that in many cases the SE is better off allowing some positive level of SEO. It is not clear, however, what the optimal level of SEO is in these cases. In particular, how does it depend on the variance of the measurement error? To answer this, let $\hat{A}(\sigma)$ denote the set of α SEO effectiveness levels that maximize the search engine's traffic. For two sets $A_1 \subseteq \mathbb{R}$ and $A_2 \subseteq \mathbb{R}$, we say that $A_1 \succeq A_2$ if and only if for any $\alpha_1 \in A_1$ there is an $\alpha_2 \in A_2$ such that $\alpha_2 \leq \alpha_1$ and for any $\alpha_2' \in A_2$ there is an $\alpha_1' \in A_1$ such that $\alpha_1' \geq \alpha_2'$.

Corollary 1 *If $v_1/v_2 > 3/2$, then the optimal SEO effectiveness is increasing as the variance of the measurement error increases. In particular, for any $\sigma_1 > \sigma_2 > 0$, we have $\hat{A}(\sigma_1) \succeq \hat{A}(\sigma_2)$.*

Figure 2 illustrates the findings by depicting the probability of the higher valuation site getting the organic link as a function of α for different values of σ. As the corollary states, the optimal SEO effectiveness levels are increasing in σ. We have already shown that SEO can be beneficial because it can serve as a mechanism that corrects the search engine's error when measuring how relevant sites are. The above corollary tells us that if the error is larger then higher effectiveness of SEO is required to correct the error. This suggests, somewhat counter intuitively, that investments against SEO on the SE's part complement investment in better search algorithms, and do not substitute them. That is, only search engines that are already very good at estimating true qualities should fight hard against black hat SEO. Nevertheless, measurement error is not always under the control of the search engine, but can depend on external factors and vary from keyword to keyword. Therefore, it may make sense to allow higher levels of SEO in areas where the quality measurement is very noisy. Another factor to consider is that although higher levels of SEO may improve the ranking, the investments sites

[19] When σ is close to $\frac{q_1 - q_2}{2}$ the probability of the desired outcome can slightly exceed $3/4$ for some small $\alpha > 0$ values. This happens because ineffective SEO gives player 1 the chance to win in cases when the error favors player 2 while not affecting the other cases.

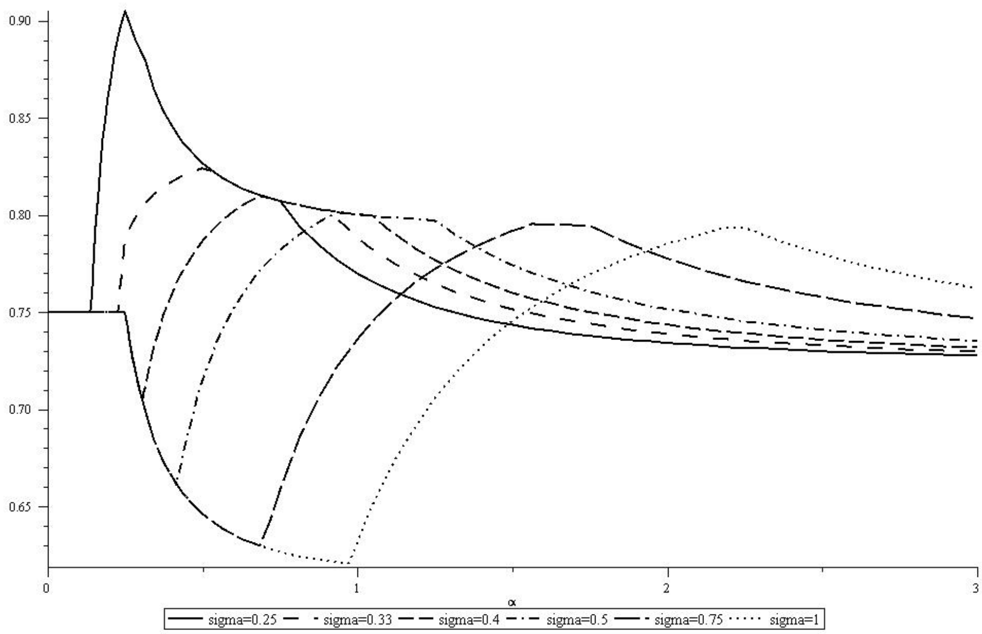

Figure 2: The figure shows the probability that the higher quality site wins the organic link as a function of SEO effectiveness for different error levels ($v_1 = 1.8, v_2 = 1, q_1 - q_2 = 0.25$).

make in it mostly end up at firms offering SEO services and not the SE. One solution to this problem is to allow sites to pay for placing links among organic results, similar to what sites such as *Baidu.com* or *Yelp.com* do. There are arguably many potential pitfalls to including paid links as part of the organic results and we do not suggest that this is always optimal, but our model does offer a plausible explanation as to why some search engines may do so.

4.2 The effect of SEO on advertiser profits

When SEO is effective, sites have a natural incentive to invest in SEO, which they do not have when $\alpha = 0$. In the extreme case of $\alpha \to \infty$ the difference in initial scores dissipates and the game becomes a regular all-pay auction. If, for example, $v_1 > v_2$ then player 1's expected payoff is $v_1 - v_2$, whereas player 2 makes nothing in expectation. Comparing this to the case in which there is no SEO - player 2 making $v_2/4$ and player 1 making $3v_1/4$ - reveals that player 2 is worse off with SEO whereas player 1 is better off iff $v_1 > 4v_2$. This implies that high levels of SEO only increase profits for sites with outstanding valuations. The following corollary provides detailed results on the sites' payoffs.

Corollary 2

1. *If $v_1 > v_2$ then Player 2's payoff is decreasing in α.*

2. *If $v_1 > v_2$ there always exists an $\alpha^* > 0$ such that Player 1 is better off with an SEO effectiveness level of $\alpha = \alpha^*$ than with $\alpha = 0$. If $v_1 > 4v_2$ or $\sigma < \frac{v_1}{v_2}\frac{q_1 - q_2}{2}$ then Player 1 is strictly better off.*

The player with the lower valuation is therefore worse off with higher SEO. Player 1, on the other hand, is better off with a certain positive level of SEO, especially if its valuation is much higher than its competitor's and if the measurement error is small. The intuition from the former follows from the fact that higher levels of SEO emphasize the differences in valuations, and the higher the difference the more likely that the higher valuation wins. For the latter condition, smaller measurement errors make it easier for the player with the higher starting score to win and to take advantage of SEO. The corollary shows that the player with the higher valuation is generally happy with some positive level of SEO.

5 SEO and Sponsored Links

In the previous section, we have focused on the effect of SEO on the traffic that the search engine can maintain as an important determinant of the success and the profit of the SE. Here, we examine the interplay between search engine optimization and the sale of sponsored (paid) links to advertisers to uncover the direct revenue effects of SEO and combine them with the results on traffic.

We first examine the case in which consumers are not sophisticated and randomly choose between the sponsored and the organic link, that is when $\psi = 0$. The following proposition summarizes our results in this case:

Proposition 2

1. *If $R_1(f(q_1)) > (3/2)R_2(f(q_2))$, then for any $\sigma > |q_1 - q_2|/2$, there exists a $\underline{\gamma} < 1$ and a positive $\hat{\alpha} = \hat{\alpha}(\sigma, v_1, v_2, q_1, q_2)$ SEO effectiveness level such that if $\gamma > \underline{\gamma}$ then $ET(\hat{\alpha}) \geq ET(0)$, and*

2. *$E\pi_{SE}(\hat{\alpha}) > E\pi_{SE}(0)$ if and only if $R_1(f(q_1)) - R_1(\gamma f(q_1)) \geq R_2(f(q_2)) - R_2(\gamma f(q_2))$.*

The results show that SEO optimization can be beneficial not only in increasing traffic, but surprisingly, it can also increase the expected revenues of the search engine from sponsored links. In particular, if a high enough proportion of visitors click on the organic links, there is a positive level of SEO for any error level that increases the traffic of the SE by improving the expected quality of the organic link. The intuition for the results on the traffic is similar to our basic model: if the high quality site has a valuation that is high enough relative to its competitor, it has a good incentive to invest in SEO, correcting the error that the SE had possibly made when ranking them. Increased traffic is certainly important, but it may result in lower profits if the revenues from sponsored links go down. The proposition specifies the condition under which SE revenues increase: if the valuation of the higher quality site is high enough for the remaining clicks after acquiring the organic link relative to its competitor. To better understand this scenario the following corollary provides sufficient conditions for it to hold.

Corollary 3 *Expected search engine profits increase with a higher traffic if any of the following hold:*

1. *$|r_1'()| < \delta$ with a sufficiently small $\delta > 0$.*

2. *$q_1/q_2 < 1 + \beta$ with a sufficiently small $\beta > 0$*

3. *$f'()$ is not too high.*

First, when advertisers' revenue functions are decreasing slowly (in particular, the higher quality site's revenue) then sponsored revenues go up as a result of the traffic increase. This is easy to see, if one considers constant marginal revenues. If sites' valuation for a click is constant no matter the amount of clicks, the competition for organic links and the bidding for sponsored links becomes independent. Thus, if SEO result in a higher overall traffic, revenues can only increase as the identity of the sponsored winner does not change. A similar intuition holds when marginal returns are decreasing at a low enough rate. Second, when the quality differences are not too high or when consumers are not very sensitive to expected link quality, revenues also increase. The rationale behind this result is that when the higher quality site wins the organic link, it has a lowered incentive to bid for the sponsored link yielding that the second highest quality site will obtain them. However, the price that site 2 pays is based on site 1's residual valuations (lower than its valuation for the first clicks), which is higher than what site 1 would pay based on site 2's residual valuations. However, if the fact that the higher quality site (1) gets the organic link results in a large increase in traffic, the residual valuation of site 1 for the sponsored link may be lower than the residual valuation of site 2. That is, although the search engine benefits from the fact that site 1's high quality organic link attracts more traffic, the sponsored prices may go down if site 1 becomes "saturated" with organic clicks and has a decreased valuation for them reducing the pressure in the sponsored auction. Condition 2 in the proposition quantifies this exactly, whereas condition 2 in the corollary makes sure this does not happen.

The above logic holds in more general cases as well. Indeed, if there are more players in the sponsored auction, the chances that revenues go down with high traffic are low. If we pool the valuations of many other potential players for the sponsored link, the only case when SE revenues go down is when site 1 and 2's valuations are high enough to surpass the rest of the players even after acquiring the organic link. Otherwise, an increase in traffic will lead to an increase in revenues.

We now relax the important assumption that consumers choose between the organic and sponsored links randomly while ingoring their expected qualities ($\psi = 0$). We consider the case when visitors can possibly identify (with a ψ probability) the best link among the results. To make the analysis tractable we assume a simple linear form for the marginal valuations, where $r_1(t) = r_1 \cdot (1-t)$ and $r_2(t) = r_2 \cdot (1-t)$ and r denotes r_2/r_1.

Proposition 3

1. *For any $0 \leq \psi < 1$ there exist $\underline{\gamma}(r,\psi) \leq \overline{\gamma}(r,\psi)$ such that if and only if $\underline{\gamma}(r,\psi) < \gamma < \overline{\gamma}(r,\psi)$ there is a positive level of SEO that improves expected traffic for any error level. Furthermore, $\overline{\gamma}(r,\psi) = 1$ when $r < \frac{2}{3}\left(1 + \frac{q_1-q_2}{(1-\psi)q_2+\psi q_1}\right)\frac{2-\psi^2 q_1 - q_1 + \psi^2 q_2 - \psi q_2}{2+\psi^2 q_1 - \psi q_1 - \psi^2 q_2 + 2\psi q_2 - q_2}$.*

2. *For any* $0 \leq \psi < 1$*, there is a* $\hat{\gamma}(r, \psi)$ *such that* $ET(\alpha) > ET(0)$ *yields* $E\pi(\alpha) > E\pi(0)$ *iff* $\gamma < \hat{\gamma}(r, \psi)$*.*

3. $\hat{\gamma}(r, \psi)$ *decreases in* ψ *for any* $0 \leq r \leq 1$*.*

Figure 3 summarizes the results. The shaded areas depict the regions in which SEO increase the efficiency of the organic result and thus expected traffic (bounded by $\underline{\gamma}(r, \psi)$ and $\overline{\gamma}(r, \psi)$). There are generally two regions where this happens: when γ is high and r is low and when γ is low, r is high with $\psi > 0$. The former region is a straightforward generalization of our previous results. When many consumers click on the organic links and the second site's valuation is relatively low, the first site has a good incentive to invest heavily in SEO, helping achieve the desired organic result and increasing expected traffic. The former region is more interesting. When there is a positive proportion of sophisticated consumers who click on the higher quality link no matter where it is located the incentives change. The better quality site now has an incentive to invest in SEO not because of a fear of not getting the clicks, but because of the high expected costs of sponsored links. These costs are higher when the second player has a relatively high valuation (r is high) and when many of the non-sophisticated consumers click on the sponsored link (γ is low). This latter force also sheds light on when SE revenues are affected by a high ψ. The solid curve on the figure depicts the boundary of the regions where a higher traffic results in higher or lower profits. The area left of the curve is certainly beneficial for the search engine, since both traffic and profits go up with a certain level of SEO, whereas profits decrease with a higher traffic right of the curve. As it can be gleaned from the figure the area below the curve is shrinking as ψ increases, consistent with the third point of the proposition. The intuition is that when the proportion of sophisticated consumers goes up many consumers will click site 1's link. If site 1 won the organic link, the SE does not generate any profits from clicks on its link, and the SE profits will be hurt. Therefore, when ψ is high the SE has higher profits when the organic link goes to the lower quality site and r is high enough (right of the curve).

In summary, our analysis of the interaction between organic and sponsored links, in addition to showing how our main results generalize, uncovers interesting nuances. SEO can improve

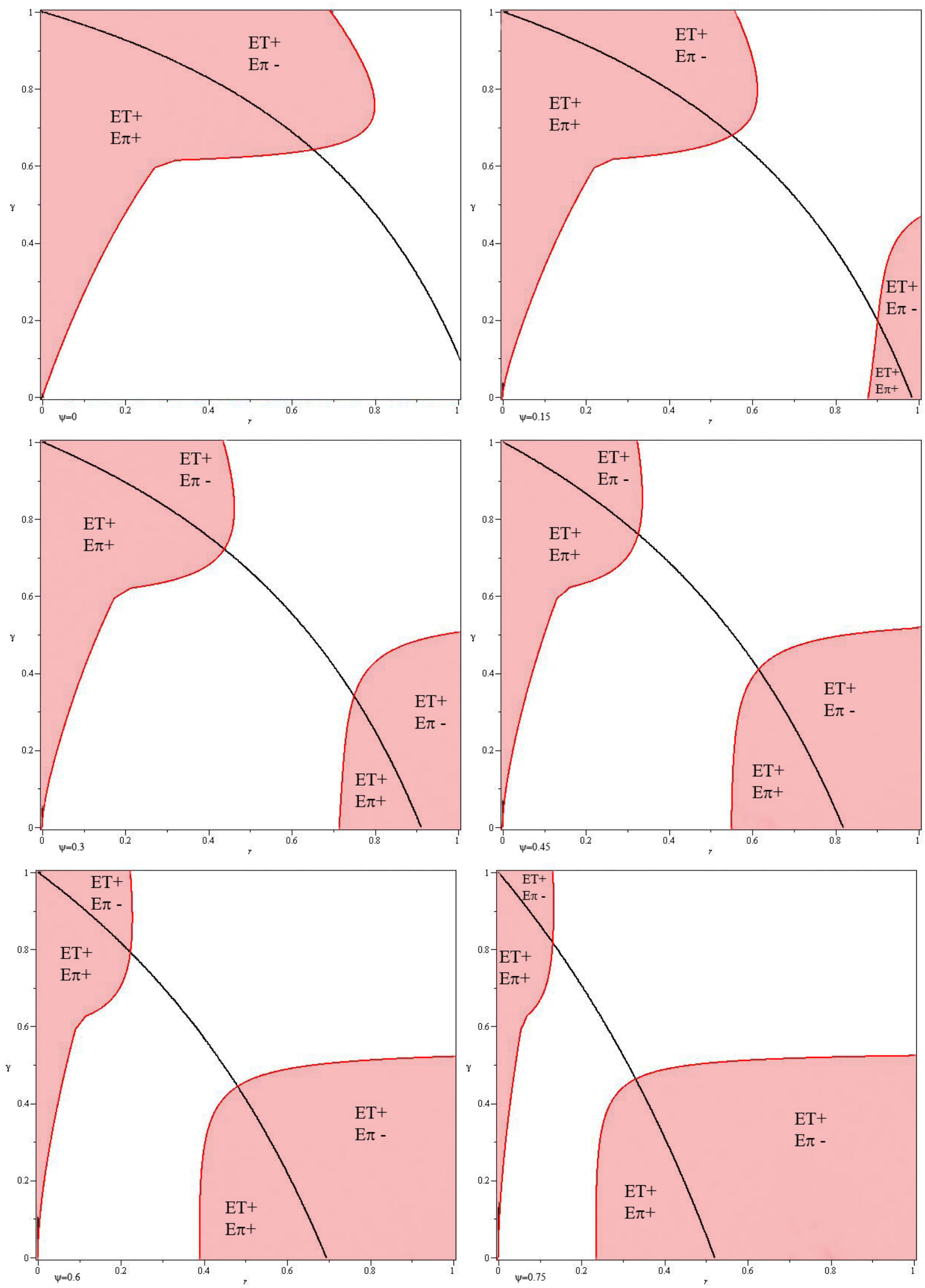

Figure 3: The figure shows the parameter regions in which SEO is beneficial with respect to expected traffic and profits. The shaded areas depict the regions where SEO certainly improves the expected traffic ($ET+$), whereas the solid curve separates the regions where SEO improves (decreases) expected profits, denoted by $E\pi+$ ($E\pi-$). Quality levels are $q_1 = 1$ and $q_2 = 0.8$ and $r = r_2/r_1$.

the traffic of the search engine for the same reasons as in our basic setup and in many cases this translates to increased profits. However, the search engine has to be careful in order to avoid situations in which increased traffic leads to reduced profits. We identify two scenarios when this happens. First, when a good organic list boosts the traffic to the SE and the high valuation sites get satisfied with organic clicks, their valuation for sponsored links decreases, leading to lower revenues. Second, when a high proportion of sophisticated customers click on the higher quality link no matter where it is located, the SE profit suffers if the organic list is better despite an increase in traffic. We would like to emphasize that the latter results on the relationship between traffic and revenue hold in a general setting, regardless of search engine optimization (even if it is non-existent) and shed light on the incentives of the search engine to have high quality organic links. When searchers are able to find the best link among the results with a high likelihood, sponsored and organic links serve as competing substitutes. As a result the search engine is more likely to lose "alternative" revenues from sponsored links by offering high quality organic links.

6 General Model with Multiple Links

In this section we show that our main results are robust under more general assumptions. We focus on showing that SEO is beneficial to improving the ranking of organic links, and defer analysis on the impact of sponsored links and search engine profits to future work. First, we extend our model to allow multiple websites to compete for multiple links in one ranked list. Second, we relax the assumption on the distribution of the search engine's measurement error. Finally, we consider the case of an incomplete information structure, where websites do not know the values of the measurement errors induced by the search engine's algorithm, and analyze the resulting Bayesian Nash equilibrium. We analyze competition among sites on the organic list solely, that is, when $\gamma = 1$ and $\psi = 0$.

The analysis is highly simplified by the use of a multiplicative scoring function instead of an additive one. Thus, the ranking score of site i with quality q_i is $\tilde{s}_i = \tilde{q}_i \cdot \tilde{b}_i^{\alpha} \cdot \tilde{\varepsilon}_i$. This scoring function is equivalent to taking an exponent of our original additive function and maintains its ordinal properties. However, we still assume that a website's effort of $\tilde{b}_i$ costs $\tilde{b}_i$, which results in a convex cost function. That is, the current multiplicative formulation is equivalent to our additive model with an exponential cost of investment in SEO. Such a convex cost function is common for investments and it is realistic in the case SEO, since it is harder to improve one's ranking after reaching a certain level.

The game still consists of n websites that are considered by the search engine for inclusion in the organic list consisting of k links with qualities $\tilde{q}_1 > \tilde{q}_2 > \ldots > \tilde{q}_n$. Let $\tilde{q} = (\tilde{q}_1, \ldots, \tilde{q}_n)$, and let $\tilde{b} = (\tilde{b}_1, \ldots, \tilde{b}_n)$ be the SEO expenditures of the n sites. Regarding the error $\tilde{\varepsilon}_i$, we allow its distribution to be arbitrary with c.d.f $F_{\tilde{\varepsilon}}$ having finite support, a mean of zero and a finite variance normalized to 1. Let $\underline{\varepsilon}$ and $\bar{\varepsilon}$ be the lower and upper boundaries of the support respectively, and assign $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)$. Similarly to Section 4, we assume that the error is large enough that it makes a difference, that is, we assume that $\tilde{q}_i\underline{\varepsilon} < \tilde{q}_{i+1}\bar{\varepsilon}$ for each $1 \leq i \leq n$. Furthermore, let Φ_i^j be an indicator for site i appearing in location j among the top k sites.

Since, we assumed $\psi = 0$, we treat consumer search as an exogenous process and assume that when site i is displayed in location j of the organic list, it receives β_j clicks from a mass one of consumers. We call this quantity the click-through rate. Given sites' click-through rates, we define t_i as the total amount of visitor traffic a site receives in a list of k sites:

$$t_i(\tilde{b}, \tilde{q}) = \mathbb{E}_{\tilde{\varepsilon}}\left[\sum_{j=1}^{k} \beta_j Pr(\Phi_i^j = 1)\right]. \tag{4}$$

The profit of site i is thus $\pi_i(\tilde{b}, \tilde{q}) = R_i(t_i(\tilde{b}, \tilde{q})) - \tilde{b}_i$. We let $\pi = (\pi_1, \ldots, \pi_n)$. The first order conditions necessary for equilibrium are given by

$$\frac{\partial t_i(\tilde{b}_i, \tilde{b}_{-i}, \tilde{q})}{\partial \tilde{b}_i} = \frac{1}{r_i(t_i(\tilde{b}, \tilde{q}))}. \tag{5}$$

Our construction fulfills the conditions of Theorem 1 in Athey and Nekipelov (2010). To see this, we first prove that a proportional increase in the bids of all other players decreases site i's profit by $\tilde{b}_i$, which is a variation on Lemma 1 in Athey and Nekipelov (2010).

Lemma 3 *Assume that $\frac{\partial}{\partial \tilde{b}'}\pi(\tilde{b}, \tilde{q})$ is continuous in $\tilde{b}$. Suppose that $t_i(\tilde{b}, \tilde{q}) > 0$ for all i. Then $\tilde{b}$ is a vector of equilibrium bids satisfying the first order conditions in (5) iff $\frac{d}{d\tau}\pi_i(\tilde{b}, \tau\tilde{b}_{-i}, \tilde{q})|_{\tau=1} = -\tilde{b}_i$ for all $i \leq k$.*

Using Lemma 3, we can rewrite the first order conditions by defining a mapping $\tilde{b} = \lambda(\tau)$ that exists in some neighborhood of $\tau = 1$:

$$\tau \frac{d}{d\tau} \pi_i(\lambda_i(\tau), \tau\lambda_{-i}(\tau), q) = -\tilde{b}_i \tag{6}$$

We let $V = [0, v_1] \times \ldots \times [0, v_k]$ be the support of potential bids of players 1 to k, and define $D_0(\tilde{b}, \tilde{q}) = \frac{\partial}{\partial \tilde{b}'} \pi(\tilde{b}, \tilde{q})$ with the diagonal elements replaced with zeros. The following theorem from Athey and Nekipelov (2010) establishes the conditions under which the mapping $\lambda(\tau)$ exists locally around $\tau = 1$ and globally for $\tau \in [0, 1]$, which yields the equilibrium bids of the players.

Theorem 1 (Athey and Nekipelov (2010)) *Assume that D_0 is continuous in $\tilde{b}$. Suppose that for each $i = 1, \ldots, k$, $t_i(\tilde{b}, \tilde{q}) > 0$, and that each π_i is quasi-concave in $\tilde{b}_i$ on V and for each $\tilde{b}$ its gradient contains at least one non-zero element. Then*

1. *An equilibrium exists if and only if for some $\delta > 0$ the system of equations (6) has a solution on $\tau \in [1 - \delta, 1]$.*

2. *The conditions from part 1 are satisfied for all $\delta \in [0, 1]$ and so an equilibrium exists, if $D_0(\tilde{b}, \tilde{q})$ is locally Lipschitz and non-singular for $\tilde{b} \in V$ except for a finite number of points.*

3. *There is a unique equilibrium if and only if for some $\delta > 0$ the system of equations (6) has a unique solution on $\tau \in [1 - \delta, 1]$.*

4. *The conditions from part 3 are satisfied for all $\delta \in [0, 1]$, so that there is a unique equilibrium, if each element of $\frac{\partial}{\partial \tilde{b}'} \pi(\tilde{b}, \tilde{q})$ is Lipschitz in $\tilde{b}$ and non-singular for $\tilde{b} \in V$*[20].

The theorem shows that under very general conditions, websites would spend non-zero efforts on SEO in equilibrium. We now proceed to analyze how positive levels of SEO effectiveness α affect the satisfaction of consumers from the ranking of the organic list. To analyze the incentives of the different websites, it is easier to transform the multiple links contest into a

[20] Athey and Nekipelov (2010) give example conditions for the non-singularity of the matrix D_0 in their Lemma 2.

game where websites choose the amount of traffic they would like to acquire from organic clicks, which implicitly determines their bids. We define the vector of traffic for each site i given the SEO effectiveness α and the vector of bids $\tilde{b}$ as $t^\alpha(\tilde{b}) = (t_1^\alpha(\tilde{b}), \ldots, t_n^\alpha(\tilde{b}))$. For each player i, fixing the bids of other players as b_{-i}, we can rewrite the first order condition of each player as $\frac{\partial \pi_i}{\partial t_i} = 0$. The expected utility of consumers when searching through links with traffic vector t^α is $EU(t^\alpha) = \sum_i q_i t_i^\alpha$.

Analyzing the result of the SEO game with multiple links is hard. In addition, under certain conditions, such as when the errors are small or α is very large, multiple equilibria might exist as shown in Siegel (2009). We therefore proceed to analyze the special cases defined by Theorem 1 where an internal equilibrium exists for all players and the first order conditions hold for players in equilibrium. For every α we define $\mathbb{T}^\alpha = \{t^\alpha | EU(t^\alpha) \geq EU(t^0)\}$ as the group of all traffic distributions over sites where the expected consumer utility is higher than under the benchmark traffic distribution t^0.

The following proposition shows that under certain conditions, a positive level of SEO can improve consumer satisfaction. These conditions are sufficient, but by no means necessary. We conjecture that much weaker conditions can be found under which SEO improves consumer satisfaction.

Proposition 4 *For each α such that there exists a vector of non-negative functions $M(t) = (M_1(t), \ldots M_k(t))$ with*

$$\frac{M_i(t)}{M_{i+1}(t)} > \frac{t_i \ \left|\frac{\partial \pi_{i+1}}{\partial t_{i+1}}(t)\right|}{t_{i+1} \ \left|\frac{\partial \pi_i}{\partial t_i}(t)\right|} \tag{7}$$

for every $t \in \mathbb{T}^\alpha$ and $\frac{\partial \pi_{i+1}}{\partial t_{i+1}}(t) \neq 0$, the equilibrium distribution of traffic $t^{\alpha *}$ satisfies $EU(t^{\alpha *}) > EU(t^0)$

The conditions in (7) imply that the sequence of bounding function limits the changes in profits of the different players from increased organic traffic. As a result, the existence of such a sequence means that extra traffic does not yield "too steep" changes in players profits and thus their incentives to decrease their expected amount of clicks in equilibrium. In such cases, allowing $\alpha > 0$ improves consumer satisfaction from the resulting quality of ranking and increases total traffic to the search engine.

7 Conclusion

Search engine optimization is a widespread phenomenon which affects consumer search abilities and search marketing decisions tremendously. Our model of the phenomenon reveals that sites investing in improving their search ranking without changing their link's relevance essentially participate in a contest for the top organic links. We find that some level of SEO can be useful to the search engine and its customers because it acts as a mechanism that improves the rankings by placing sites with high valuations for the links higher on the results list. In general, if sites' valuations for consumers are aligned with how much utility consumers gain when visiting them then SEO is beneficial to consumers and increases traffic to the search engine. We also find that when the search algorithms of the SE are less accurate, higher levels of SEO are more likely to be beneficial. Participating sites, on the other hand, might be worse off as they carry the extra burden of having to invest in SEO, whereas if search engine optimization does not exist they do not have to make additional effort. In the case when a low quality site has high valuation for traffic, it can benefit from the presence of SEO and get to an undeserved, better position, hindering search engine payoffs and consumer benefits.

We uncover important details about the interaction between the two sides of the market that the search engine is surrounded by. We contribute to the growing literature examining the implications of the quality of the organic results on sponsored revenues. Surprisingly, in many cases SEO can not only increase the search engine's traffic through improving the organic side, but also its revenues from sponsored links. Nonetheless, there are certain condition under which expected profits are lowered by better organic results. In particular, when consumers are very sensitive to the quality of the search results, the boost in organic traffic may fulfill advertisers' needs for clicks, lowering their willingness to pay for clicks. Moreover, when sophisticated consumers place less trust in the ranking displayed by the search engine and they knowingly click on the higher quality link no matter its location, profits may go down even though traffic increases.

Our paper has important practical implications. Contrary to the popular belief, allowing sites to invest in improving their ranking without improving their relevance can be beneficial to the search engine and the consumers, but can hurt the top sites even if they end up higher in the rankings eventually. Our results explain why some search engines seem to work very hard to reduce the possibility and effectiveness of SEO, while others like Baidu or Yelp offer such services themselves. Our results suggest that when the search algorithms are not very accurate or when SEO methods are hard to identify, allowing some level of black hat SEO does not necessarily compromise the results. Despite the potential advantages, search engines should be careful when considering the profit implications of SEO. In many cases SEO can improve profits, but when consumers become very strategic in their clicking behavior revenues from sponsored links may decline. Search engines usually keep track of their visitors' clicking actions and analyze their behavior, which in combination with our results could prove useful to maximize profits.

Our results also provide important recommendations to websites that are competing for top organic and sponsored links. Contrary to popular belief, sites engaging in SEO are not only sites that wish to achieve a better position than their true quality merits, but also top quality sites that need to defend their position. Moreover, an important problem that advertisers face is how to allocate resources between the different areas of search engine marketing. The paper provides useful guidelines in how much to invest in SEO and how much to leave for the sponsored bidding.

We believe that the economics of search engine optimization is a topic of high importance to both academics and practitioners. In this paper we examine the basic forces of this interesting and unusual ecosystem. Given the complexity of the problem, our model has a number of limitations that could be explored by future research. First of all, we model SEO as a static game, whereas in reality sites invest in SEO dynamically, reacting to each other's and the search engine's actions. Second, we take sites' qualities and valuations as given and we do not account for the possibility of investments that substantially improve content quality and/or valuations for visitors. Finally, we do not explicitly examine how search engines can invest in reducing SEO. Our results yield that under certain conditions having some SEO is beneficial, but we do not determine how much it is worth investing against SEO when it is too effective and is detrimental. Depending on the costs of reducing SEO, it might be unprofitable for the search engine to do so.

References

Aggarwal, Gagan, Jon Feldman, S. Muthukrishnan, Martin Pál. 2008. Sponsored search auctions with markovian users. *WINE '08: Proceedings of the 4th International Workshop on Internet and Network Economics*. Springer-Verlag, Berlin, Heidelberg, 621–628.

Athey, S., D. Nekipelov. 2010. A Structural Model of Sponsored Search Advertising Auctions. *mimeo* .

Athey, Susan, Glenn Ellison. 2009. Position auctions with consumer search. NBER Working Papers 15253, National Bureau of Economic Research, Inc. URL `http://ideas.repec.org/p/nbr/nberwo/15253.html`.

Benczur, Andras A., DAvid Siklosi, Jacint Szabo, Istvan Biro, Zsolt Fekete, Miklos Kurucz, Attila Pereszlenyi, Simon Racz, Adrienn Szabo. 2008. Web spam: a survey with vision for the archivist. IWAW '08 International Web Archiving Workshop. Available at: http://iwaw.europarchive.org/08/IWAW2008-Benczur.pdf.

Bracha, O., F. Pasquale. 2007. Federal Search Commission-Access, Fairness, and Accountability in the Law of Search. *Cornell L. Rev.* **93** 1149.

Chen, Yongmin, Chuan He. 2006. Paid Placement: Advertising and Search on the Internet. *SSRN eLibrary* .

Clark, D.J., C. Riis. 2000. Allocation efficiency in a competitive bribery game. *Journal of Economic Behavior and Organization* **42**(1) 109–124.

Edelman, Benjamin, Michael Ostrovsky, Michael Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review* **97**(1) 242–259. URL `http://ideas.repec.org/a/aea/aecrev/v97y2007i1p242-259.html`.

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

Jeziorski, Przemyslaw, Ilya Segal. 2009. What Makes Them Click: Empirical Analysis of Consumer Demand for Search Advertising. *SSRN eLibrary* .

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

Kirkegaard, René. 2009. Favoritism in asymmetric contests: Head starts and handicaps. *working paper* .

Konrad, Kai A. 2007. Strategy in contests – an introduction. *WZB-Markets and Politics Working Paper No. SP II 2007-01* .

Krishna, Vijay John Morgan. 2007. The winner-take-all principle in small tournaments. *Advances in Applied Microeconomics* **28** 849–862.

Mercadante, J.M. 2008. What is Google's Reputation Score-A Method for Modified Self-Regulation of Search. *Seton Hall Legis. J.* **33** 327.

Pasquale, F. 2006. Rankings, Reductionism, and Responsibility. *Clev. St. L. Rev.* **54** 115.

Rutz, Oliver J., Randolph E. Bucklin. 2007. A Model of Individual Keyword Performance in Paid Search Advertising. *SSRN eLibrary* .

Sen, Ravi. 2005. Optimal search engine marketing strategy. *Int. J. Electron. Commerce* **10**(1) 9–25.

Siegel, Ron. 2009. All-pay contests. *Econometrica* **77**(1) 71–92.

Sisak, Dana. 2009. Multiple-prize contests the optimal allocation of prizes. *Journal of Economic Surveys* **23** 82–114(33).

Taylor, Greg. 2010. Search quality and revenue cannibalisation by competing search engines. *working paper, Available at SSRN: http://ssrn.com/abstract=1233282* .

Varian, Hal R. 2007. Position auctions. *International Journal of Industrial Organization* **25**(6) 1163–1178.

White, Alexander. 2009. Search engines: Left side quality versus right side profits. *working paper, Touluse School of Economics* .

Wilbur, Kenneth C., Yi Zhu. 2009. Click fraud. *Marketing Science* **28**(2) 293–308.

Xing, Bo, Zhangxi Lin. 2006. The impact of search engine optimization on online advertising market. *ICEC '06: Proceedings of the 8th international conference on Electronic commerce*. ACM, New York, NY, USA, 519–529. doi:http://doi.acm.org/10.1145/1151454.1151531.

Xu, Lizhen, Jianqing Chen, Andrew B. Whinston. 2009. Too organic for organic listing? interplay between organic and sponsored listing in search advertising. *working paper, Available at SSRN: http://ssrn.com/abstract=1409450* .

Yang, S., A. Ghose. 2010. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marekting Science* **29** 602–623.

Yao, Song, Carl F. Mela. 2010. A dynamic model of sponsored search advertising. *working paper, Kellogg* .

Zhu, Yi, Kenneth C. Wilbur. 2010. Hybrid advertising auctions. *Marketing Science* forthcoming.

Appendix A - All Pay Auctions with Headstarts

All pay-auctions with headstarts are generalizations of basic all-pay auctions. In traditional all-pay auctions players submit bids for an object that they have different valuations for. The player with the highest bid wins the object, but all players have to pay their bid to the auctioneer (hence the term "all-pay auction"). When the auctioneer does not collect the revenues from the bids which are sunk, the game is called a contest. If players have headstarts then the winner is the player with the highest score - the sum of bid and headstart (see the Appendix and Kirkegaard (2009) for details).

The level of headstart in our model depends on the starting scores and hence on the error. For example, if $q_1 > q_2$ and $\varepsilon_1 = \varepsilon_2 = 1$, the error does not affect the order (which is $q_1 \geq q_2$) nor the difference between the starting scores $(q_1 - q_2)$. Since SEO effectiveness is α, an investment of b only changes the scores by αb, therefore the headstart of site 1 is $\frac{q_1-q_2}{\alpha}$. As the size of the headstart decreases with α, the more effective SEO is, the less the initial difference in scores matters. Even if site 1 is more relevant than site 2, it is not always the case that it has a headstart. If $\varepsilon_1 = -1$ and $\varepsilon_2 = 1$ then $s_1^S = q_1 - \sigma < s_2^S = q_2 + \sigma$ given our assumption on the lower bound on σ. Thus, player 2 has a headstart of $\frac{q_2+2\sigma-q_1}{\alpha}$. By analyzing the outcome of the all-pay auction given the starting scores, we can determine the expected utility of the SE and the websites.

Appendix B - Proofs

Proof of Lemma 1: We decompose the final scores of both sites into a headstart h and a bid as follows: $\tilde{s}_1^F = h + b_1$ and $\tilde{s}_2^F = b_2$ where $h = \frac{s_1^S - s_2^S}{\alpha}$. The decomposed scores have the property that $\tilde{s}_1^F \geq \tilde{s}_2^F \iff s_1^F \geq s_2^F$ for every b_1, b_2 and thus preserve the outcome of the SEO game. Since the investments are sunk and only the winner receives the benefits (with the exception of a draw) the SEO game is equivalent to an all-pay auction with a headstart of $h = \frac{s_1^S - s_2^S}{\alpha}$. In the following, we present the solution of such a game to facilitate the presentation of the remaining proofs. $\square$

Solution of all-pay auctions with headstarts: As derived by Kirkegaard (2009), the generic two player all-pay auction with headstarts has a unique mixed strategy equilibrium. When players valuations are $v_1 \geq v_2$ and player 1 has a headstart of h then s/he wins the auction with the following probabilities:

$$W_1(h) = Pr(1 \text{ wins}|h \geq 0) = \begin{cases} 1 & h > v_2 \\ 1 - \frac{v_2}{2v_1} + \frac{h^2}{2v_1 v_2} & h \leq v_2 \end{cases}$$

$$W_1(h) = Pr(1 \text{ wins}|h < 0) = \begin{cases} 1 - \frac{v_2}{2v_1} & h \geq v_2 - v_1 \\ \frac{v_1^2 - h^2}{2v_1 v_2} & -v_1 \leq h < v_2 - v_1 \\ 0 & \text{otherwise} \end{cases}$$

For completeness, we specify the players' cumulative bidding distributions. When h is positive,

$$F_1(b) = \begin{cases} 0 & b \leq 0 \\ \frac{h+b}{v_2} & b \in (0, v_2 - h] \\ 1 & b > v_2 - h \end{cases} \quad F_2(b) = \begin{cases} 0 & b \leq 0 \\ 1 - \frac{v_2 - h}{v_1} & b \in (0, h] \\ 1 - \frac{v_2 - b}{v_1} & b \in (h, v_2] \\ 1 & b > v_2 \end{cases} \tag{8}$$

When h is negative,

$$F_1(b) = \begin{cases} 0 & b \leq h \\ \frac{b-h}{v_2} & b \in (h, v_2 + h] \\ 1 & b > v_2 + h \end{cases} \quad F_2(b) = \begin{cases} 0 & b \leq 0 \\ 1 - \frac{v_2 - b}{v_1} & b \in (0, v_2] \\ 1 & b > v_2 \end{cases} \tag{9}$$

In our model, the value of the headstart is determined by the different realizations of the errors $\varepsilon_1, \varepsilon_2$. There are four possible realizations with equal probability: $h_1 = h_2 = \frac{q_1 - q_2}{\alpha}$, $h_3 = \frac{q_1 - q_2 + 2\sigma}{\alpha}$ and $h_4 = \frac{q_1 - q_2 - 2\sigma}{\alpha}$.

Proof of Lemma 2: Since $P(\alpha) = \frac{1}{2}W_1(h_1) + \frac{1}{4}W_1(h_3) + \frac{1}{4}W_1(h_4)$ and the headstart does not depend on v_1 and v_2, it is enough to show that $W_1(\cdot)$ is increasing in v_1 and decreasing in v_2. These easily follow from the definition of $W_1(\cdot)$. The results on q_1 and q_2 follow from the fact that h_1, h_3, h_4 are all increasing in q_1 and decreasing in q_2, and $W_1(\cdot)$ depend on them only through h in which it is increasing. $\square$

Proof of Proposition 1: Since $ET(\alpha) = f(q_2) + (f(q_1) - f(q_2))P(\alpha)$, it is enough to examine the function $P(\alpha)$ and compare its values with $P(0)$. We use the notation $P_i = Pr(1 \text{ wins}|h_i)$. Given the above described equilibrium of the two-player all-pay auctions we have $P_i = W_1(h_i)$. We further define $\alpha_1 = \frac{q_1 - q_2}{v_2}$, $\alpha_3 = \frac{q_1 - q_2 + 2\sigma}{v_2}$, $\alpha_4 = \frac{q_2 - q_1 + 2\sigma}{v_1}$, $\alpha_4' = \frac{q_2 - q_1 + 2\sigma}{v_1 - v_2}$. Note that $P_1 = P_2$, since the headstarts in the first two case are equal. Thus $P(\alpha) = \frac{1}{2}P_1 + \frac{1}{4}P_3 + \frac{1}{4}P_4$, and $P_1 = 1$ iff $\alpha \leq \alpha_1$, $P_3 = 1$ iff $\alpha \leq \alpha_3$, $P_4 = 1 - \frac{v_2}{2v_1}$ iff $\alpha \geq \alpha_4'$. Furthermore, it is easy to check that $\alpha_1 \leq \alpha_3$, $\alpha_4 \leq \alpha_3$, and $\alpha_4 \leq \alpha_4'$.

We proceed by separating the three parts of the proposition:

- Part 1: By setting $\alpha = \alpha_1$, we have $P_1 = P_3 = 1$, and thus $P(\alpha) \geq 3/4$ for any σ.

- Part 2: In order to prove this part, we determine the α value that yields the highest efficiency level for a given σ if $v_1/v_2 > 3/2$. As noted above, $P(\alpha)$ is a linear combination of $W_1(h_1)$, $W_1(h_3)$, $W_1(h_4)$. Since $W_1(\cdot)$ is continuous and h_1, h_3, h_4 are all continuous in

α, it follows that $P(\alpha)$ is continuous in α. However, $P(\alpha)$ is not differentiable everywhere, but there are only a finite number of points where it is not. Therefore it suffices to examine the sign of $P'(\alpha)$ to determine whether it is increasing or not. This requires tedious analysis, since depending on the value of σ the formula describing $P(\alpha)$ is different in up to five intervals. We identify five different formulas that $P(\alpha)$ can take in different intervals and take their derivatives:

$$
\begin{aligned}
P'(\alpha) &= P'_I(\alpha) = \frac{(q_1 - q_2 - 2\sigma)^2}{4\alpha^3 v_1 v_2} \text{ if } \alpha_4 \leq \alpha \leq \alpha_1 \& \alpha'_4, \\
P'(\alpha) &= P'_{II}(\alpha) = -\frac{(q_1 - q_2)^2}{2\alpha^3 v_1 v_2} \text{ if } \alpha_1 \leq \alpha \leq \alpha_4, \\
P'(\alpha) &= P'_{III}(\alpha) = -\frac{2(q_1 - q_2)^2 + (q_1 - q_2 + 2\sigma)^2}{4\alpha^3 v_1 v_2} \text{ if } \alpha_3 \& \alpha'_4 \leq \alpha, \\
P'(\alpha) &= P'_{IV}(\alpha) = \frac{4\sigma^2 - (q_1 - q_2)(4\sigma + q_1 - q_2)}{4\alpha^3 v_1 v_2} \text{ if } \alpha_1 \& \alpha_4 \leq \alpha \leq \alpha'_4, \\
P'(\alpha) &= P'_V(\alpha) = -\frac{(q_1 - q_2)(4\sigma + q_1 - q_2)}{2\alpha^3 v_1 v_2} \text{ if } \alpha_3 \& \alpha'_4 \leq \alpha.
\end{aligned}
$$

In any other range the derivative of $P(\alpha)$ is 0. It is clear from the above formulas that $P'_I(\alpha)$ is always positive and that $P'_{II}(\alpha)$, $P'_{III}(\alpha)$, and $P'_V(\alpha)$ are always negative. Furthermore, one can show that

$$P'_{IV}(\alpha) > 0 \text{ iff } \sigma > \frac{1 + \sqrt{2}}{2}(q_1 - q_2).$$

This allows us to determine the maximal $P(\alpha)$ for different values of σ in four different cases.

1. If $\frac{q_1 - q_2}{2} \leq \sigma \leq \frac{v_1}{v_2}\frac{q_1 - q_2}{2}$ then $\alpha_4 \leq \alpha'_4 \leq \alpha_1 \leq \alpha_3$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: $0, P'_I(\alpha), 0, P'_{II}(\alpha), P'_{III}(\alpha)$. Therefore $P(\alpha)$ is first constant, then increasing, then constant again and then strictly decreasing. Thus, any value between α'_4 and α_1 maximizes $P(\alpha)$. Using the notation of Corollary 1, $\hat{A}(\sigma) = [\alpha'_4, \alpha_1]$.

2. If $\frac{v_1}{v_2}\frac{q_1 - q_2}{2} \leq \sigma \leq \frac{v_1 + v_2}{v_2}\frac{q_1 - q_2}{2}$ then $\alpha_4 \leq \alpha_1 \leq \alpha'_4 \leq \alpha_3$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: 0, $P'_I(\alpha)$, $P'_{IV}(\alpha)$, $P'_{II}(\alpha)$, $P'_{III}(\alpha)$. Therefore $P(\alpha)$ is first constant, then decreasing, then strictly increasing,

then depending on the sign of $P'_{IV}(\alpha)$ increasing or decreasing, and finally strictly decreasing. Therefore if $\sigma < \frac{1+\sqrt{2}}{2}(q_1 - q_2)$ then α_1 maximizes $P(\alpha)$, that is $\hat{A}(\sigma) = \{\alpha_1\}$. If $\sigma = \frac{1+\sqrt{2}}{2}(q_1 - q_2)$ then $P(\alpha)$ is constant between α_1 and α'_4, that is $\hat{A}(\sigma) = [\alpha_1, \alpha'_4]$. Finally, if $\sigma = \frac{1+\sqrt{2}}{2}(q_1 - q_2)$ then $\hat{A}(\sigma) = \{\alpha'_4\}$.

3. If $\frac{v_1+v_2}{v_2}\frac{q_1-q_2}{2} \leq \sigma \leq \frac{v_1}{2v_2-v_1}\frac{q_1-q_2}{2}$ then $\alpha_1 \leq \alpha_4 \leq \alpha'_4 \leq \alpha_3$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: 0, $P'_{II}(\alpha)$, $P'_{IV}(\alpha)$, $P'_{II}(\alpha)$, $P'_{III}(\alpha)$. In this case $P'_{IV}(\alpha) > 0$ since $\sigma \geq \frac{v_1+v_2}{v_2}\frac{q_1-q_2}{2} \geq (1+\frac{3}{2})\frac{q_1-q_2}{2} > (1+\sqrt{2})\frac{q_1-q_2}{2}$. Therefore $P(\alpha)$ is first constant, then decreasing, then strictly increasing again and finally strictly decreasing. Thus, there are two candidates for the argmax: α_1 and α'_4. One can show that $P_{IV}(\alpha'_4) > P_{II}(\alpha_1)$ iff $v_1 > \sqrt{2}v_2$, therefore α'_4 maximizes $P(\alpha)$ in this case.

4. If $\frac{v_1}{2v_2-v_1}\frac{q_1-q_2}{2} \leq \sigma$ then $\alpha_1 \leq \alpha_4 \leq \alpha_3 \leq \alpha'_4$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: 0, $P'_{II}(\alpha)$, $P'_{IV}(\alpha)$, $P'_{V}(\alpha)$, $P'_{III}(\alpha)$. Similarly to the previous case $P'_{IV}(\alpha) > 0$, therefore $P(\alpha)$ is first constant, then decreasing, then strictly increasing again and finally strictly decreasing. Comparing the two candidates for the argmax yields that $P_{IV}(\alpha_3) > P_{II}(\alpha_1)$ iff $v_1 > (3/2)v_2$, that is α_3 maximizes $P(\alpha)$ in this case.

In each of the cases above, it is clear that the maximum is higher than $P(0) = 3/4$. In cases 1 and 2, $P(\alpha)$ is strictly increasing after a constant value of 3/4 and in cases 3 and 4 we directly compared to $P_{II}(\alpha_1) = 3/4$. This completes the proof of Part 2.

- Part 3: One can derive the efficiency function for different cases as in Part 2. It follows that if $\sigma \geq \frac{v_2+v_1}{v_2-v_1}\frac{q_1-q_2}{2}$ then $P'(\alpha)$ is first 0 then negative and finally positive. Therefore $P(\alpha)$ either has a maximum in $\alpha = 0$ or as it approaches infinity. However,

$$P(\alpha) \underset{\alpha\to\infty}{\longrightarrow} \frac{v_1}{2v_2} \leq \frac{1}{2} < \frac{3}{4} = P(0).$$

□

Proof of Corollary 1: In the proof of Proposition 1, we determined the values of α

that maximize $P(\alpha)$ for different σ's. In summary:

$$\hat{A}(\sigma) = \begin{cases} [\alpha_4', \alpha_1] & \text{if } \frac{q_1-q_2}{2} \leq \sigma \leq \frac{v_1}{v_2}\frac{q_1-q_2}{2} \\ \alpha_1 & \text{if } \frac{v_1}{v_2}\frac{q_1-q_2}{2} < \sigma \leq (1+\sqrt{2})\frac{q_1-q_2}{2} \\ [\alpha_1, \alpha_4'] & \text{if } \sigma = (1+\sqrt{2})\frac{q_1-q_2}{2} \\ \alpha_4' & \text{if } (1+\sqrt{2})\frac{q_1-q_2}{2} < \sigma \leq \frac{v_1+v_2}{v_2}\frac{q_1-q_2}{2} \\ \alpha_4' & \text{if } \frac{v_1+v_2}{v_2}\frac{q_1-q_2}{2} \leq \sigma \leq \frac{v_1}{2v_2-v_1}\frac{q_1-q_2}{2} \\ \alpha_3 & \text{if } \frac{v_1}{2v_2-v_1}\frac{q_1-q_2}{2} \leq \sigma \end{cases}$$

It is straightforward to check that all of α_1, α_3, and α_4' are increasing in σ and that the $\hat{A}(\sigma)$ is increasing over the entire range. □

Proof of Corollary 2: First, we describe the payoffs of the two players in an all-pay auction with headstarts. When players follow the mixed strategies described in (8) and (9), player 1's payoff is:

$$\pi_1(h) = \begin{cases} 0 & h \leq v_2 - v_1 \\ v_1 - v_2 + h & v_2 - v_1 < h < v_2 \\ 1 & h \geq v_2 \end{cases}$$

where h is the headstart of player 1. The payoff of player 2 can be obtained from the same formula by changing the roles. Then, we get player i's total payoff by linearly combining the above quantities:

$$\pi_1 = \frac{1}{2}\pi_1(h_1) + \frac{1}{4}\pi_1(h_3) + \frac{1}{4}\pi_1(h_4)$$

Then following the same steps as in the proof of Proposition 1, we can determine the values of α that maximize Player 1's payoff for different σ's. We get the following results. If $v_1 \leq 3v_2$ then

$$\arg\max_{\alpha} \pi_1 = \begin{cases} \alpha_1 & \text{if } \sigma \leq \frac{v_1}{v_2}\frac{q_1-q_2}{2} \\ [0, \alpha_1] & \text{if } \frac{v_1}{v_2}\frac{q_1-q_2}{2} < \sigma \end{cases}$$

In case of $3v_2 < v_1 \leq 4v_2$

$$\arg\max_{\alpha} \pi_1 = \begin{cases} \alpha_1 & \text{if } \sigma \leq \frac{v_1}{v_2}\frac{q_1-q_2}{2} \\ \alpha_3 & \text{if } \frac{v_1}{v_2}\frac{q_1-q_2}{2} < \sigma \leq \frac{v_1}{4v_2-v_1}\frac{q_1-q_2}{2} \\ [0, \alpha_1] & \text{if } \frac{v_1}{4v_2-v_1}\frac{q_1-q_2}{2} < \sigma \end{cases}$$

Finally, when $v_1 \leq 4v_2$ we have

$$\arg\max_{\alpha} \pi_1 = \begin{cases} \alpha_1 & \text{if } \sigma \leq \frac{v_1}{v_2}\frac{q_1-q_2}{2} \\ s_3 & \text{if } \frac{v_1}{v_2}\frac{q_1-q_2}{2} < \sigma \end{cases}$$

It is easy to see that with exception of the two cases when the optimal α is anywhere between 0 and α_1, Player 1 is strictly better off with a particular positive level of SEO than without it. $\square$

PROOF OF PROPOSITION 2: We use backward induction and start by analyzing the last stage of the game, in which sites bid for sponsored link. Normally, there are multiple equilibria in a second price auction with known valuation, but if we eliminate weakly dominated strategies the only equilibrium is in which bidders bid their valuations and, thus, the highest bidder wins. We can determine site's valuations based on the outcome of the organic allocation. For example, if site 1 gets the organic link its valuation for the total traffic from sponsored link becomes $v_{s,1,1} = R_1(f(q_1)) - R_1(\gamma f(q_1))$, whereas site 2's valuation is $v_{s,2,1} = R_2((1-\gamma)f(q_1))$. If site 2 gets the organic link then sites valuations are $v_{s,1,2} = R_1((1-\gamma)f(q_2))$ and $v_{s,2,2} = R_2(f(q_2)) - R_2(\gamma f(q_2))$, respectively. Consequently, one can determine the outcome of the auction and the payoff of player i given that player k gets the organic link (where $j \neq i$):

$$\pi_{s,i,k} = \max(0, v_{s,i,k} - v_{s,j,k}).$$

One can the determine a players valuation in the SEO game. The players have the same valuation for the organic clicks as before, but depending on whether they get it or not, they may have an "option" value from the sponsored clicks. Therefore, player i's valuation of acquiring the organic link is

$$v_i = R_i(f(q_i)) + \pi_{s,i,i} - \pi_{s,i,j},$$

where $i \neq j$. We can then apply the results of Proposition 1 to prove part 1 using the above v_1 and v_2. We just need to show that there is a $\overline{\gamma} < 1$ such that $v_1 > (3/2)v_2$ as long as $\gamma > \overline{\gamma}$. One can determine the above valuations in two cases. If $R_2((1-\gamma)f(q_1)) > R_1(f(q_1)) - R_1(\gamma f(q_1))$ then

$$\begin{aligned} v_1 &= R_1(\gamma f(q_1)) - R_1((1-\gamma)f(q_2)) + R_2(f(q_2)) - R_2(\gamma f(q_2)), \\ v_2 &= R_2(\gamma f(q_2)) - R_2((1-\gamma)f(q_1)) + R_1(f(q_1)) - R_1(\gamma f(q_1)). \end{aligned}$$

Otherwise,

$$\begin{aligned} v_1 &= R_1(f(q_1)) - R_2((1-\gamma)f(q_1)) - R_1(f(q_2)) + R_2(f(q_2)) - R_2(\gamma f(q_2)), \\ v_2 &= R_2(\gamma f(q_2)). \end{aligned}$$

It is easy to check that when $\gamma = 1$ then $v_1/v_2 = R_1(f(q_1))/R_2(f(q_2)) > 3/2$. Since v_1/v_2 is continuous in γ, there is a $\overline{\gamma}$ for which $v_1/v_2 = 3/2$, proving part 1.

For part 2, one has to determine the revenues of the search engine from the sponsored links. Let $\pi_{s,SE,k}$ denote the revenue of the search engine when site k receives the organic link. It is easy to determine that

$$\begin{aligned} \pi_{s,SE,1} &= \min(R_1(f(q_1)) - R_1(\gamma f(q_1)), R_2((1-\gamma)f(q_1))), \\ \pi_{s,SE,2} &= R_2(f(q_2)) - R_2(\gamma f(q_2)). \end{aligned}$$

Since $R_2((1-\gamma)f(q_1)) > R_2(f(q_2)) - R_2(\gamma f(q_2))$, we have that $\pi_{s,SE,1} > \pi_{s,SE,2}$ iff $R_1(f(q_1)) - R_1(\gamma f(q_1)) > R_2(f(q_2)) - R_2(\gamma f(q_2))$. Since $E\pi_{SE}(\alpha) = P(\alpha)(\pi_{s,SE,1} - \pi_{s,SE,2}) + \pi_{s,SE,2}$ and $P(\hat{\alpha}) > P(0)$, it follows that $E\pi_{SE}(\hat{\alpha}) > E\pi_{SE}(0)$ iff $\pi_{s,SE,1} > \pi_{s,SE,2}$, completing the proof. $\square$

Proof of Corollary 3: For part 1, let $\overline{\delta} = r_1(0)\frac{f(q_1)-f(q_2)}{f(q_1)^2}$. If $|r_1'()| < \overline{\delta}$ then

$$\frac{r_1(f(q_1))}{r_2((1-\gamma)f(q_2))} \geq \frac{r_1(0) - f(q_1)\delta}{r_1(0)} > \frac{f(q_2)}{f(q_1)}$$

yielding that the condition in the second part of the proposition is satisfied.

For parts 2 and 3, one can check that $R_1(f(q_1)) - R_1(\gamma f(q_1)) > R_2(f(q_2)) - R_2(\gamma f(q_2))$ if $f(q_1) = f(q_2)$. Since R_i and f are continuous, if $f(q_1)$ is only slightly higher, the inequality still holds. This implies that if q_1/q_2 is close enough to 1 or f is not too much increasing, the condition of the second part of the proposition is satisfied. $\square$

Proof of Proposition 3: We derive sites' valuation and the search engine's expected profits similarly to how we do in the proof of Proposition 2. We get

$$\begin{aligned} v_1 &= \max\{R_1((\psi + (1-\psi)\gamma)f(q_2 + (\psi + (1-\psi)\gamma)(q_1 - q_2))), \\ &\quad R_1(f(q_1)) - R_2((1 - (\psi + (1-\psi)\gamma))f(q_2 + (\psi + (1-\psi)\gamma)(q_1 - q_2)))\} \\ &\quad - R_1((1 - (1-\psi)\gamma)f(q_2 + (1 - (1-\psi)\gamma)(q_1 - q_2))) + R_2(f(q_2)) - R_2((\psi + (1-\psi)\gamma)f(q_2)), \\ v_2 &= R_2((1-\psi)\gamma f(q_2 + (1 - (1-\psi)\gamma)(q_1 - q_2))) \\ &\quad - \max\{0, R_2((1-\psi)(1-\gamma)f(q_2 + (\psi + (1-\psi)\gamma)(q_1 - q_2))) \\ &\quad - R_1(f(q_1)) + R_1((\psi + (1-\psi)\gamma)f(q_1))\}. \\ \pi_{s,SE,1} &= \min\{R_1(f(q_1)) - R_1((\psi + (1-\psi)\gamma)f(q_1)), \\ &\quad R_2((1 - (\psi + (1-\psi)\gamma))f(q_2 + (\psi + (1-\psi)\gamma)(q_1 - q_2)))\}, \\ \pi_{s,SE,2} &= R_2(f(q_2)) - R_2((\psi + (1-\psi)\gamma)f(q_2)). \end{aligned}$$

One can use the above expressions to determine the parameter regions in which which $v_1 > (3/2)v_2$ yielding the functions $\underline{\gamma}(r,\psi)$ and $\overline{\gamma}(r,\psi)$. One can also check that $v_1 > (3/2)v_2$ holds when $\gamma = 1$ and

$$r < \frac{2}{3}\left(1 + \frac{q_1 - q_2}{(1-\psi)q_2 + \psi q_1}\right)\frac{2 - \psi^2 q_1 - q_1 + \psi^2 q_2 - \psi q_2}{2 + \psi^2 q_1 - \psi q_1 - \psi^2 q_2 + 2\psi q_2 - q_2}.$$

Furthermore, solving $\pi_{s,SE,1} = \pi_{s,SE,2}$ yields

$$\hat{\gamma} = 1 - \frac{2(f(q_1)(f(q_1)-1) - rf(q_2)(f(q_2)-1)}{(1-\psi)(f(q_1)^2 - rf(q_2)^2)},$$

which is clearly decreasing in ψ, completing the proof. □

PROOF OF LEMMA 3: We denote by $P_i^j(\tilde{b}, \tilde{q})$ the probability that site i appears in location j among the top k sites. This probability equals $P_i^j(\tilde{b}, \tilde{q}) = \int \Phi_i^j(\tilde{b}, \tilde{q}, \tilde{\varepsilon}) dF_{\tilde{\varepsilon}}(\tilde{\varepsilon})$. The total number of clicks site i gets, t_i, is therefore $t_i(\tilde{b}, \tilde{q}) = \sum_{j=1}^{J} \beta_j P_i^j(\tilde{b}, \tilde{q})$.

A proportional increase of all bids in $\tilde{b}$ does not change the expected rankings of the sites, and keeps the expected number of clicks constant for all sites: $t_i(\tilde{b}, \tilde{q}) = t_i(\eta\tilde{b}, \tilde{q})$ for any $\eta \neq 0$. Since t_i is homogeneous of degree zero, by Euler's homogeneous function theorem, $\sum_{l=1}^{k} \tilde{b}_l \frac{\partial}{\partial \tilde{b}_l} R_i(t_i(\tilde{b}, \tilde{q})) = r_i \sum_{l=1}^{k} \tilde{b}_l \frac{\partial}{\partial \tilde{b}_l} t_i(\tilde{b}, \tilde{q}) = 0$.

As a result, the following holds:

$$\sum_{l=1}^{k} \frac{\partial}{\partial \tilde{b}_l} \pi_i(\tilde{b}, \tilde{q}) \cdot \tilde{b}_l = \sum_{l=1}^{k} (\tilde{b}_l \frac{\partial}{\partial \tilde{b}_l} R_i(t_i(\tilde{b}, \tilde{q}))) - \tilde{b}_i = -\tilde{b}_i \tag{10}$$

This identity can be rewritten as:

$$\frac{\partial}{\partial \tau} \pi_i(\tilde{b}_i, \tau\tilde{b}_{-i}, \tilde{q})|_{\tau=1} + \tilde{b}_i \frac{\partial}{\partial \tilde{b}_i} \pi_i(\tilde{b}_i, \tilde{b}_{-i}, \tilde{q}) = -\tilde{b}_i \tag{11}$$

Thus, the FOC $\frac{\partial}{\partial \tilde{b}_i} \pi_i(\tilde{b}_i, \tilde{b}_{-i}, \tilde{q}) = 0$ holds for $\tilde{b}_i > 0$ iff $\frac{\partial}{\partial \tau} \pi_i(\tilde{b}_i, \tau\tilde{b}_{-i}, \tilde{q})|_{\tau=1} = -\tilde{b}_i$ □

PROOF OF PROPOSITION 4: Recall that $\mathbb{T}^\alpha$ contains all traffic distributions $t = (t_1, \ldots, t_n)$ for which the expected utility of consumers is weakly greater with an SEO effectiveness level of α than with $\alpha = 0$, implying $EU(t) = \sum_i q_i \cdot t_i \geq \sum_i q_i \cdot t_i^0$.

Let $\beta = \sum_j \beta_j$ be the sum of the exogenous click-through rates. If we normalize the sum of clicks $\sum_i t_i$ to 1 we have $\beta = \sum_i t_i$. We then define, for each α, the mapping

$$F_\alpha : \frac{(t_1, \ldots, t_n)}{\beta} \to \frac{(M_1(t) \cdot |\frac{\partial \pi_1}{\partial t_1}(t)| + t_1, \ldots, M_n(t) \cdot |\frac{\partial \pi_n}{\partial t_n}(t)| + t_n)}{\beta + \sum_i M_i(t) |\frac{\partial \pi_i}{\partial t_i}|}.$$

Above, for convenience of notation, α was dropped and the first orders $\frac{\partial \pi_i}{\partial t_i}$ as well as the traffic distributions t_i are given under the specific α for each F_α. To simplify exposition we assign $\tilde{t} = \frac{t}{\beta}$ as the normalized traffic vector. This mapping has several special properties:

- The mapping maps a given traffic distribution to another, implicitly setting the required bids to reach this traffic distribution. The input and output distributions are normalized to one, so the mapping is closed on traffic distributions. In addition, the mapping is continuous.

- The fixed points of each mapping F_α are the equilibrium distributions of the SEO game. To see this, note that when the first order conditions hold and are equal zero, the mapping has a fixed point, and vice-versa.

- The set of traffic distributions superior to $U(t^0)$ (which is $\mathbb{T}^\alpha$) is convex.

As a result, showing that the fixed points of F_α are superior to t^0 would prove that SEO increases consumer utility in equilibrium. To see this, let $t \in \mathbb{T}^\alpha$. Then

$$U(F(\tilde{t})) - U(\tilde{t}) = \sum_i q_i \left(\frac{t_i + M_i |\frac{\partial \pi_i}{\partial t_i}|}{\beta + \sum_j M_j |\frac{\partial \pi_j}{\partial t_j}|} - \frac{t_i}{\beta} \right) = \sum_i q_i \frac{\beta M_i |\frac{\partial \pi_i}{\partial t_i}| - t_i \sum_j M_j |\frac{\partial \pi_j}{\partial t_j}|}{\beta(\beta + \sum_j M_j |\frac{\partial \pi_j}{\partial t_j}|)} \tag{12}$$

As $M_i(t)$ are non-negative and $\beta = \sum_i t_i$, the difference in utilities is positive when:

$$\sum_i q_i \left(\beta M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| - t_i \sum_j M_j \left| \frac{\partial \pi_j}{\partial t_j} \right| \right) = \sum_j t_j \left(\sum_i M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| (q_i - q_j) \right) > 0 \tag{13}$$

Fix i, j and assume $i < j$, then $q_i \geq q_j$. Looking at the couples of additions in the sum for i, j we get

$$t_j M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| (q_i - q_j) + t_i M_j \left| \frac{\partial \pi_j}{\partial t_j} \right| (q_j - q_i) = \left(t_j M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| - t_i M_j \left| \frac{\partial \pi_j}{\partial t_j} \right| \right) (q_i - q_j) \tag{14}$$

which is larger than zero when condition (7) holds.

This shows that the set $\mathbb{T}^\alpha$ is convex and closed under the continuous mapping F_α. As a result, Brouwer's fixed point theorem tells us that a fixed point of F_α exists in $\mathbb{T}^\alpha$, which concludes the proof. $\square$