How Does Ranking Affect User Choice in Online Search?

Mark Glick · Greg Richards · Margarita Sapozhnikov · Paul Seabright

Abstract This paper investigates whether a search engine's ordering of algorithmic results has an important effect on website traffic. A website's ranking on a search engine results page is positively correlated with the clicks that it receives. This could result from the search engine's accurately predicting the websites relevance to users. Or it could result from users merely clicking on the highest ranked links, regardless of the website's relevance. Using a unique dataset, we find that a website's rank, not just

M. Glick
Department of Economics, University of Utah, Salt Lake City, UT 84112, USA
e-mail: glick@economics.utah.edu

G. Richards
Keystone Strategy, 2000 Sierra Point Parkway, Suite 701, Brisbane, CA 94005, USA
e-mail: grichards@keystonestrategy.com

M. Sapozhnikov (✉)
Keystone Strategy, 150 CambridgePark Drive, Cambridge, MA 02140, USA
e-mail: msapozhnikov@keystonestrategy.com

P. Seabright
Toulouse School of Economics (IAST), 21 allée de Brienne, 31015 Toulouse, France
e-mail: Paul.Seabright@tse-fr.eu

Springer

its relevance, strongly and significantly affects the likelihood of a click. We also find evidence that rank influences CTRs partly by controlling access to the scarce attention of users, but primarily by substituting the reputational capital of the search engine for the reputation of individual websites.

1 Introduction

Recent anti-trust investigations of the internet search market in the US and Europe have considered to what extent search engines have the ability to influence traffic to websites. It is well known that the ranking (i.e. the hierarchical physical location on a search-results page) of websites[1] is positively correlated with click-through rates (CTRs).[2] If this correlation reflected a causal impact of ranking on CTRs, then search engines with a large share of total search activity would influence a large amount of traffic to websites.

How could a correlation between rank and CTRs arise if there were no causal impact of the former on the latter? This might occur through reverse causation: The search engine might accurately predict the relevance of websites to users (and therefore their likely future CTRs) and then place websites on the page as a function of this prediction.

Using a unique dataset of individual search behavior we show that there is indeed a strong positive correlation between the rank of a website on a given search engine results page (SERP) and the probability that an individual will click on that website. Although part of this correlation can be explained by the predicted relevance of the website, there is a substantial direct causal impact even when this is taken into account.

We find that being at the top of the ranking in the algorithmic search results has a large and statistically significant causal impact on the odds of receiving a user click, and that moving the website from rank 1 to rank 2 on the same page decreases the odds of a click by between one third and two thirds depending on the specific search that is undertaken. We concede that no single statistical method completely eliminates endogeneity concerns; however, our results are robust and all evidence points to very high economic significance of the algorithmic rank.

There are no studies to our knowledge in the economic literature that estimate systematically the effect of rank in the algorithmic search results using individual user data. Athey and Meidan (2011) and Athey and Nekipelov (2011) are important papers on the analysis of user behavior in the paid results. Previous research has indicated that CTRs can increase markedly for results placed at the top of a results page compared

---

[1] We use the term website to describe the search result hyperlink and associated domain name that when clicked takes a user to the webpage that is associated with that hyperlink.

[2] Search engines display two types of results: These are often called paid results and algorithmic results. Our analysis focuses on algorithmic results which are results that are ranked based on how relevant the search engine infers them to be to users (rather than based on payments from the website).

to other ranks and pages (Smith and Brynjolfsson 2001; Xu and Kim 2008; Ghose and Yang 2009. Of these studies, Xu and Kim (2008) is based on a small-sample laboratory experiment, and Xu and Kim (2008) uses data on paid search from an advertiser. The closest study in spirit to ours is Smith and Brynjolfsson (2001), which uses individual data from an Internet book-purchase site, but this relates to purchases of homogeneous books rather than searches for heterogeneous websites.

Armstrong et al. (2009) and Armstrong and Zhou (2011) study the welfare implications of "prominence" in search markets. Among studies investigating the impact of placement on sponsored search results, Jerath et al. (2011) demonstrate the existence of a "position paradox" where advertisements at higher positions obtain more clicks, but this effect can be offset by a superior firm reputation. The paradox is that the superior firm may make higher profits from bidding for lower ranked positions. In a similar vein, Baye et al. (2012), using search data that were aggregated by retailer, consider the impact of product position and product reputation on the organic search results page on CTRs and find that both are important factors.

We proceed as follows. In Sect. 2 we describe the data selection process and provide summary statistics for our dataset. We also describe the nature of the search engine algorithm and provide descriptive evidence about the determinants of ranking. In Sect. 3 we demonstrate econometrically the effect of ranking on the probability of clicking on a website. Section 4 investigates the contribution of reputation and conspicuousness to enabling page rank to influence click probabilities. Section 5 concludes.

2 Data

2.1 Query Term Selection

Microsoft Corporation provided us with access to the database of the Bing log files for individual user searches during November and December 2010 and January 2011. All search engines store data from user sessions in detailed logs. The Bing logs contain recorded observations for each of the millions of Bing user queries, including for each query: a record of the date and time; all websites that were displayed on the SERP generated from the search; each website's position on the SERPs; and which websites were clicked. For each website that appeared in a set of search results, we know at what rank it appeared in each view and whether it was clicked on during that view.

In order to isolate the impact of website relevance from that of page rank, we need query terms where the website relevance to the user query remains reasonably constant during the time period of study, while the ranking of websites varies (even if only slightly). We also need to eliminate as far as possible other confusing influences. To find suitable data, we first categorized a list of available query terms and then eliminated the non-suitable categories until we arrived at a final list of queries.

A first type of unsuitable query is one that generates what are known are "highly monetized" results. For example, the query term "airline tickets" signals the intent to shop for airline tickets on-line and, because it is defined in generic terms, occurs with relatively high frequency. The intent to make a purchase and the high frequency make

this query attractive to the advertisers and the results page is highly monetized: There is a large volume of ads. The ads distract from the algorithmic results and introduce more "noise" into the algorithmic click behavior data. In order to predict click behavior on the algorithmic results we would need to know all of the paid results as well (whose presence might well be endogenous). As a consequence, these queries are not suitable for our analysis.

A second type of unsuitable query is what is known as "superfresh". Consider the query term "Obama approval rating". The intent is to look for current news, and every day (sometimes every hour) a different set of websites will be most relevant and appear in the top ranks. This variability in website relevance, which we cannot directly observe and for which we cannot control, makes such query terms unsuitable for our analysis.

A third type of unsuitable query is "navigational": Where the user has a prior intent to navigate to a specific website. An example of this is one of the most frequent queries—"facebook"—and the search results display the different subpages of this website. Although a large proportion of query terms have some corresponding domain name and thus could in theory be navigational, queries become unsuitable for our purposes only when such query terms regularly appear in among the top results on the page.

Finally, query terms that arise from non-uniform intent across users are also unsuitable. One example is the query "eclipse". Based on the websites that are displayed on the results page, this search has at least three possible intents: to learn about a solar or lunar eclipse, to find information about a software product that is known as Eclipse, and to search for one of the Twilight Saga books with this title (which is a teenage vampire romance novel).

Thus, we manually sorted through an extensive list of queries, and found four query terms that were suitable for our purposes. In alphabetical order these are: "Free Movies", "Fun Games", "Phone Numbers" and "Sports".[3] Although some of these query terms are now monetized, none were so at the time of our study. None related to newsworthy events that might have had an impact on relevance. None were primarily navigational, and none showed significant evidence of non-uniform intent.[4]

2.2 Algorithm

Algorithms are sometimes patented (the Google PageRank algorithm is covered by U.S. Patent No. 6,285,999) and exact formulas are held as trade secrets. However, the general characteristics of search algorithms are known. The paper that introduced

---

[3] In our data we identify blended search results (those compiled by the search engine, usually with multiple links and an image in one installment), and omit SERPs in which blended search results occupy any of the top three ranks. We omit the SERPs that have two or more clicks to different websites on the same page, and count two or more clicks to the same website on the same SERP as one.

[4] Only "Sports" had the corresponding domain "sports.com" appearing among the top five domains, and it gathered many fewer views and clicks in the top three positions than did the top three domains. The only evidence we could find of non-uniform intent in our four queries was the appearance of the domain "Wikipedia.com" in the results for "Phone Numbers", and this domain also appeared very rarely in the top three positions.

Google Brin and Page (1998) states that "Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems." The fundamental ranking techniques of a search engine algorithm depend on natural language processing of the content of websites, topological analysis of the connections between websites, and analysis of the interactions of consumers with search results, among other things.

A Search Engine algorithm proceeds in two steps: choosing the websites that match the query term and then putting them in ranking order. The first step uses keyword focused measures, which examine the placement and count of the query term words in a website name and anchor text.[5] Once the set of websites to be displayed in the SERP is determined, they are ranked using natural language techniques, static rank[6] and user behavior data, such as prior website traffic and prior CTR.

This obviously raises a concern about reverse causality: It may be previous CTRs that determine ranking rather than ranking that determines future CTRs. Based on discussions with the engineers who provided us with the Bing data, we believe that at the time of our study (11/1/2010–1/31/2011), and for our selected query terms, the Bing algorithm relied on website CTRs that were calculated over long prior periods of time, and was refreshed only occasionally. As we illustrate further below, fluctuations in the CTR over short periods of time do not seem to be a determinant in Bing ranking for the query terms that we selected.

During the study period, some instability remained in the relatively new Bing algorithm, which can cause variation in ranks and is most probably the cause of the variation in page rank in our data.[7] In addition, during this study period, the results of the Bing algorithm were not personalized to user characteristics, which further alleviates many potential data concerns.

2.3 Sample Statistics

Our sample consists of those websites that appear on Bing on the first SERP (in positions 1–10) for each of the four query terms considered. "Free Movies" resulted in views for 262 such distinct websites, "Fun Games" for 158, "Phone Numbers" for 322, and "Sports" for 996.

However, not all websites had views in all ten positions. As an illustration, Table 1 displays the top five websites (as determined by the total number of views for the time period of our analysis) for the query term "Phone Numbers"; they are displayed in the order of frequency of appearance in Rank 1.

For each of the five websites, Table 1 shows how many views each website had in each rank during our sample period, and what the website CTRs were in each rank.

---

[5] Anchor text is the text in a hyperlink that leads to the website and website content.

[6] Static Rank is computed based on the ontological map of all web pages, consisting of nodes and links between them. Given these interconnections, Static Rank assigns a score to each website. This score represents the probability that a person starting at a random page and randomly clicking on links will arrive at the website in question.

[7] Variation in ranking can be caused by maintenance operations on some of the servers, for example.

Table 1 Top five websites for "Phone Numbers"

Website/rank	1	2	3	4	5	6	7	8	9	10	Total
phonenumber.com											
Views	17,075	13,315	1,417	9	6	5	1	0	3	1	31,832
CTR	0.295	0.168	0.1	0	0	0	0	–	0.333	0	0.233
whitepages.com											
Views	14,652	16,558	580	1	10	5	0	0	7	13	31,826
CTR	0.274	0.154	0.097	0	0.1	0	–	–	0	0	0.208
switchboard.com											
Views	80	1,893	29,734	36	19	22	6	4	1	3	31,798
CTR	0.625	0.098	0.054	0.111	0.158	0.091	0	0	0	0	0.058
anywho.com											
Views	5	0	15	8,645	6,185	1,933	9,653	3,650	1,428	201	31,715
CTR	0.8	–	0.067	0.028	0.021	0.014	0.015	0.012	0.009	0	0.019
en.wikipedia.org											
Views	1	12	8	229	4,055	21,648	4,142	1,288	349	74	31,806
CTR	0	0	0	0	0.001	0.001	0.001	0.003	0	0	0.001

For example, website phonenumbers.com had 17,075 views in rank 1, and 29.5 % of the views resulted in a click-through (CTR is 0.295). The statistics for each query term show that being in the top rank is associated with higher CTRs for each domain.

In addition, the frequency with which the top three websites appear in the top rank is also often, though not always, reflected in the ordering of their CTRs when they appear in the second rank, suggesting that some of the ranking frequency may reflect perceived website relevance. In particular, two websites—phonenumber.com and whitepages.com—are competing for the top spot on the page. Phonenumber.com has 17,075 views in rank 1 (with top rank CTR of 0.295) and whitepage.com has 14,652 (CTR is 0.274): When one website is in rank 1, the other website is usually displayed in rank 2. Phonenumber.com is slightly more relevant to the user query, since it is being clicked on more often in nearly every rank compared to whitepages.com.[8] This is consistent with the observation that phonenumber.com is observed in rank 1 more often.

Tables 2, 3 and 4 present the same statistics for the other three query terms, and display broadly similar characteristics.

These data naturally raise the question of what triggers changes in ranking. In particular, we are interested in whether the data are consistent with our claim that changes in ranking are more likely to reflect random events than to have been triggered by prior changes in CTRs. To examine this further, Fig. 1 has the time series of the daily CTR (dotted line) and daily percent of views in Rank 1 (solid line) for the two leading websites for the "Phone Numbers" query.

Our main concern is whether the changes in CTR trigger the switch between the ranks for these websites. This does not appear to be the case. It is easy to observe the level change in CTR once a website is displayed in Rank 1 more often, and the changes in CTR appear to occur after—rather than to precede—the switch between the ranks.

However, visual inspection of Fig. 1 is not the way to settle the question. Our conjecture is confirmed by Granger causality tests that were run for both websites. The summary statistics of the sample used for the Granger causality tests can be found in Table 5, and the results of the Granger causality tests are reported in Table 6. Note that the proportion of views in which the domain appears in rank 1 may sum to greater than unity across websites: For instance, it would be possible for two domains each to appear in rank 1 whenever they are viewed (thus 100 % of the time), provided the domains are never viewed on the same page

To determine the direction of causality between daily percentage of views in which the website appears in Rank 1 and its daily CTR, we perform a Wald test for the null hypothesis that lagged values of the former can be excluded from a regression of the latter, and vice versa. For the "Phone Numbers" query, we can clearly reject the null hypothesis that prior page rank has no effect on current CTRs: The F-statistic for the exclusion of the percentage of time spent in Rank 1 from the equation for CTR is

---

[8] The exception is rank 5 where phonenumber.com has six views and a CTR of zero (zero clicks out of six views) and whitepages.com has 10 views and a CTR of 0.1 (one click out of ten views). This difference seems to be attributable to chance variation, given the miniscule number of views and clicks involved.

Table 2 Top five websites for "Free Movies"

Website/rank	1	2	3	4	5	6	7	8	9	10	Total
fancast.com											
Views	20,613	2,866	158	45	18	32	56	134	87	94	24,103
CTR	0.213	0.116	0.076	0.022	0	0.125	0.018	0.022	0.023	0	0.197
freemoviescinema.com											
Views	3,231	7,879	321	40	325	4,866	4,385	1,215	214	216	22,692
CTR	0.217	0.111	0.103	0.175	0.046	0.038	0.033	0.021	0.033	0.023	0.088
hulu.com											
Views	440	13,031	10,364	107	98	117	78	37	15	1	24,288
CTR	0.13	0.102	0.078	0.075	0.02	0.068	0.038	0.027	0.067	0	0.09
ovguide.com											
Views	5	73	53	3	229	3,594	9,259	3,329	858	66	17,469
CTR	0.6	0.151	0.019	0	0.039	0.038	0.03	0.026	0.031	0	0.032
free-new-movies.com											
Views	0	374	41	68	57	166	2,319	8,738	5,638	5,470	22,871
CTR	–	0.126	0.073	0.044	0.053	0.03	0.028	0.024	0.021	0.024	0.026

Table 3 Top five websites for "Fun Games"

Website/rank	1	2	3	4	5	6	7	8	9	10	Total
mostfungames.com											
Views	99,894	1,198	669	58	308	190	3	19	1	0	102,340
CTR	0.433	0.198	0.157	0.138	0.114	0.089	0	0	0	–	0.426
addictinggames.com											
Views	2,418	86,142	14,400	936	110	43	21	9	5	1	104,085
CTR	0.361	0.095	0.072	0.036	0.055	0.023	0	0	0	0	0.097
bumarcade.com											
Views	1,401	16,438	80,228	963	2,410	1,884	913	25	9	5	104,276
CTR	0.251	0.073	0.038	0.028	0.017	0.012	0.008	0	0	0	0.045
funny-games.biz											
Views	201	85	7,909	27,979	33,570	25,842	8,011	258	51	59	103,965
CTR	0.483	0.118	0.037	0.025	0.017	0.013	0.011	0.008	0.02	0	0.020
bored.com											
Views	83	1	25	83	10,572	8,424	19,471	27,549	22,809	11,272	100,289
CTR	0.036	0	0.04	0	0.013	0.012	0.006	0.006	0.006	0.005	0.007

Table 4 Top five websites for "Sports"

Website/rank	1	2	3	4	5	6	7	8	9	10	Total
sports.yahoo.com											
Views	66,744	48,290	448	4	40	8	6	8	13	7	115,608
CTR	0.273	0.194	0.076	0	0.05	0	0	0.125	0.231	0.143	0.239
espn.go.com											
Views	48,027	65,521	1,873	15	304	95	10	7	19	18	115,889
CTR	0.207	0.132	0.1	0.067	0.066	0.084	0.1	0	0.053	0	0.162
msn.foxsports.com											
Views	875	664	104,481	284	6,816	421	133	7	50	77	113,808
CTR	0.717	0.13	0.101	0.049	0.049	0.076	0.03	0	0.02	0.052	0.102
sportsillustrated.cnn.com											
Views	166	1	131	261	38,243	14,065	7,527	14,548	19,488	9,470	103,900
CTR	0.663	1	0.031	0.031	0.023	0.019	0.014	0.014	0.016	0.016	0.020
sports.com											
Views	21	343	2,557	1,717	41,364	38,514	27,411	3,549	450	40	115,966
CTR	0.286	0	0.007	0.008	0.006	0.004	0.005	0.002	0.009	0	0.005
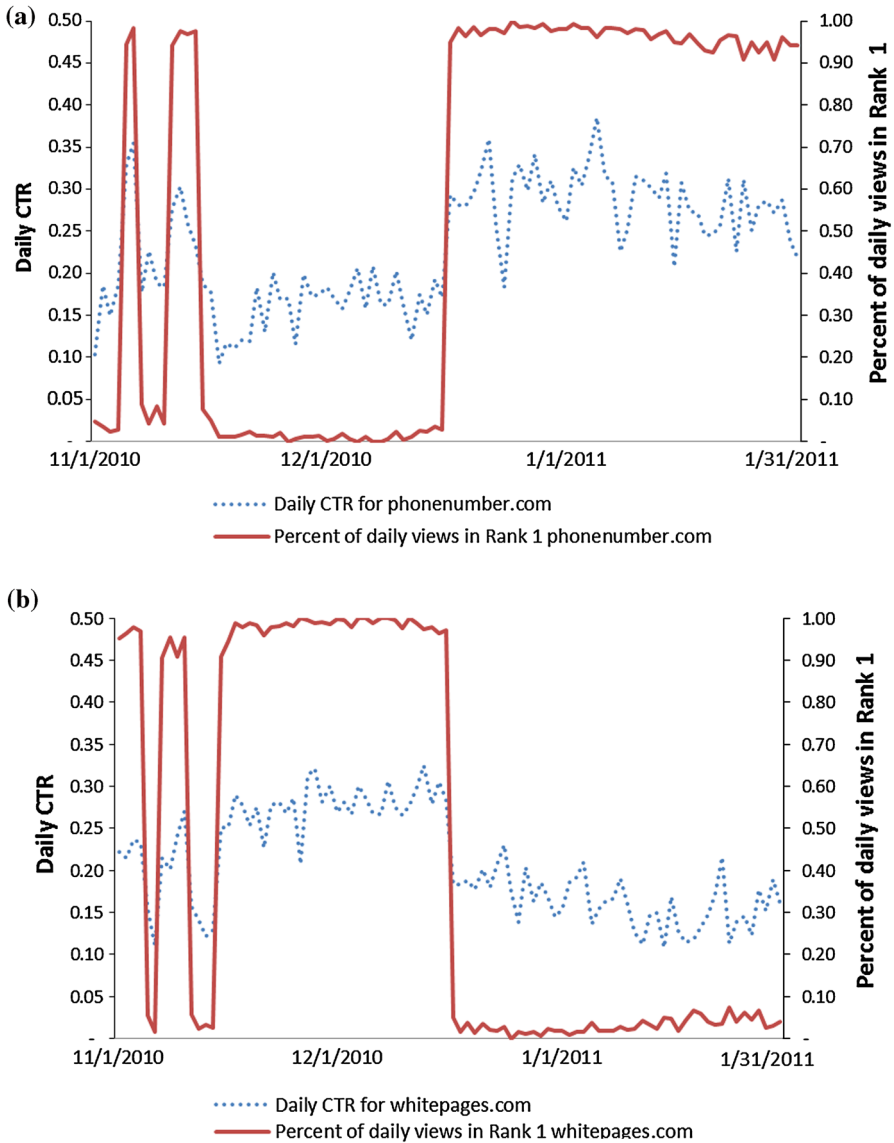
Fig. 1 CTR and % of views in Rank 1 daily. Query "Phone Numbers", **a** phonenumber.com, **b** whitepages.com

significant at 1 % for one domain and 0.1 % for the other. On the other hand, we fail to reject the null hypothesis that prior CTR has no effect on current page rank.

For the other queries the evidence is more mixed. For "Sports" the results are similar to "Phone Numbers" but at slightly lower levels of significance (5 %). For "Fun Games" there is no evidence of Granger-causality in either direction, while for "Free Movies" there is evidence of two-way causality for one domain and none for the others.

Table 5 Summary statistics for domains used in granger causality tests

Query term	Domain	CTR				% Rank 1				N
		Mean	Min	Max	SD	Mean	Min	Max	SD	
Phone Numbers	phonenumber.com	0.234	0.093	0.386	0.071	0.555	0.000	1.000	0.471	92
Phone Numbers	whitepages.com	0.207	0.108	0.325	0.063	0.442	0.000	1.000	0.471	92
Free Movies	fancast.com	0.154	0.000	1.000	0.141	0.642	0.000	1.000	0.444	67
Free Movies	freemoviescinema.com	0.146	0.000	1.000	0.208	0.237	0.000	1.000	0.354	82
Free Movies	indiemoviesonline.com	0.181	0.000	1.000	0.198	0.718	0.000	1.000	0.298	82
Fun Games	didigames.com	0.284	0.000	1.000	0.400	0.573	0.000	1.000	0.471	25
Fun Games	mostfungames.com	0.425	0.319	0.538	0.045	0.976	0.921	1.000	0.017	92
Sports	espn.go.com	0.162	0.106	0.227	0.026	0.419	0.007	0.997	0.452	92
Sports	sports.yahoo.com	0.239	0.178	0.320	0.036	0.574	0.002	0.993	0.448	92

Table 6 Granger causality: daily CTR and % in rank 1

Query term	Domain	F-Statistics	
		Predict CTR exclude % rank 1	Predict % rank 1 exclude CTR
Phone Numbers	phonenumber.com	6.5457***	0.08701
Phone Numbers	whitepages.com	8.4886***	0.33142
Free Movies	fancast.com	2.0542	0.21121
Free Movies	freemoviescinema.com	3.5557**	5.1608***
Free Movies	indiemoviesonline.com	2.9848	0.42144
Fun Games	didigames.com	8.0374	9.1387
Fun Games	mostfungames.com	0.09275	0.29655
Sports	espn.go.com	4.3336**	2.7442
Sports	sports.yahoo.com	4.5793**	1.5155

** $p < 0.05$; *** $p < 0.01$

Overall, for two query terms we can clearly accept the hypothesis, suggested to us by Bing engineers, that prior CTR is not used to determine the rank of the website. For the other query terms there is evidence of possible influence of CTR on page rank for only one of the domains used. On balance the hypothesis of lack of reverse causality seems broadly plausible given the evidence available to us.

3 Econometric Estimation

In order to estimate the effect of page rank on click probabilities we use the multinomial logit model that was developed by McFadden and used for a large variety of situations in which users make a single choice from a range of discrete options. This means that instead of estimating determinants of CTRs over a given time period we estimate the odds that a website in a given page rank is clicked on, relative to a website in the baseline Rank 10, averaged across all SERPs that gave rise to a user click.[9] This therefore allows us to abstract from the many factors that can affect CTRs, such as time of day, since these factors do not vary between alternatives that are presented to the user in a given page view.

The results are presented in Tables 7, 8, 9 and 10 for the four query terms. For ease of interpretation the coefficients are presented as odds ratios, so that the effect of a given rank should be understood as the odds that the user clicks on a website in that rank divided by the odds of clicking on a website in rank 10. An odds ratio of 1 would therefore imply no effect: the rank in question was no more likely to be clicked on than is rank 10. Odds ratios less than one imply a negative effect, odds ratios greater than one imply a positive effect.

---

[9] We use Rank 10 as the baseline since ten is the maximum number of results that appear in a single page view.

Table 7 Page rank and domain reputation as determinants of click odds using rank only

	(1) Sports	(2) Phone Numbers	(3) Fun Games	(4) Free Movies
Rank 1	105.8***	220.5***	145.0***	11.30***
	(72.79)	(26.93)	(84.02)	(50.07)
Rank 2	66.45***	120.6***	31.30***	5.657***
	(65.36)	(23.89)	(57.48)	(34.43)
Rank 3	39.73***	43.58***	14.67***	3.653***
	(57.10)	(18.73)	(44.11)	(24.75)
Rank 4	7.049***	8.212***	4.059***	1.187*
	(28.84)	(10.15)	(21.39)	(2.57)
Rank 5	5.860***	10.71***	4.836***	1.649***
	(25.77)	(11.21)	(23.50)	(5.72)
Rank 6	2.324***	2.733***	3.282***	1.926***
	(11.22)	(4.51)	(17.19)	(10.34)
Rank 7	1.629***	6.579***	2.263***	1.376***
	(6.10)	(8.84)	(11.38)	(5.23)
Rank 8	1.697***	3.628***	1.691***	1.085
	(6.67)	(5.79)	(7.07)	(1.26)
Rank 9	1.708***	2.424***	1.280**	0.951
	(6.71)	(3.69)	(3.11)	(−0.76)
Observations	619528	134907	577590	111,161

Exponentiated coefficients; t statistics in parentheses * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

There is a large variation among the query terms in the magnitude of the rank effects, but the broad qualitative findings are remarkably similar. Table 7 gives the effect of rank without controlling for website relevance for each of the four query terms. We can see that being in rank 1 increases the odds of being clicked on, relative to rank 10, by between 11 times (for "Free Movies") and 220 times (for "Phone Numbers"). This is roughly twice as large as the effect of being in rank 2, though the exact proportion varies somewhat between query terms.

There are two ways in which we control for website relevance. The first, as reported in Table 8 for each of the four query terms, is to control for the mean rank of a website over the whole sample period. This is based on the idea that the mean rank of the website does reflect the search engine's estimate of its likely relevance to users, while deviations within the sample period from this mean rank do not reflect variations in likely relevance.

Our "Mean Rank" variable is the inverse of the arithmetic mean of the rank number, so that higher values of the variable reflect higher ranks (ie those closer to rank 1). Controlling for Mean Rank lowers the odds ratio for rank 1 by over half for all queries except "Free movies", where it has a small lowering effect.

Our second way of controlling for website relevance, as reported in Table 9 for each of the four query terms, is to use a dummy variable that we call "Brand" for any website that appears in rank 1 during the sample period more than 0.5 per cent

Table 8 Page rank and domain reputation as determinants of click odds using rank and mean rank

	(1) Sports	(2) Phone Numbers	(3) Fun Games	(4) Free Movies
Rank 1	39.02***	101.2***	50.23***	10.95***
	(48.35)	(21.27)	(55.34)	(39.41)
Rank 2	25.02***	56.68***	20.09***	5.574***
	(42.64)	(18.65)	(48.29)	(32.23)
Rank 3	26.99***	30.28***	10.79***	3.607***
	(49.63)	(16.60)	(38.40)	(23.56)
Rank 4	5.512***	6.982***	3.470***	1.180*
	(24.94)	(9.34)	(18.92)	(2.46)
Rank 5	5.211***	10.00***	4.419***	1.639***
	(24.00)	(10.88)	(22.12)	(5.64)
Rank 6	2.137***	2.566***	3.065***	1.919***
	(10.10)	(4.23)	(16.19)	(10.25)
Rank 7	1.545***	6.231***	2.198***	1.373***
	(5.44)	(8.58)	(10.98)	(5.19)
Rank 8	1.648***	3.529***	1.666***	1.083
	(6.30)	(5.66)	(6.87)	(1.24)
Rank 9	1.674***	2.384***	1.268**	0.951
	(6.45)	(3.62)	(2.99)	(−0.76)
Mean rank	6.165***	4.295***	3.703***	1.055
	(24.71)	(9.33)	(27.40)	(0.85)
Observations	619,528	134,907	577,590	111,161

Exponentiated coefficients; t statistics in parentheses
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

of the total number of SERP observations.[10] This definition captures the idea that such websites are likely to be perceived as more relevant. Adding this variable to the specification that includes Mean Rank reduces further to a small extent the odds ratio for rank 1, except for "Phone Numbers" where it increases the ratio slightly, probably due to collinearity with Mean Rank.

As a robustness check we use separate fixed effects for each of the "Brand" websites instead of a single dummy variable, as reported in Table 10 for each of the four query terms. This lowers substantially the coefficient on Mean Rank, turning it negative in three cases out of four, without substantially altering the coefficients on Rank. This appears to indicate that the fixed effects and the Mean Rank variable are substantially collinear.

Overall, it is striking that even after these controls for relevance there is a large, statistically and economically very significant effect of being in rank 1 as compared to rank 10. Even in the most conservative specification (number 3), the odds ratios vary

---

[10] Our formal definition is that it appears more than 0.05 % of the total domain-rank observations; since not every SERP has the full ten ranks this is almost but not quite equal to 0.5 % of the total SERPs. We have experimented with different percentage definitions and nothing of importance depends on the precise proportion.

Table 9 Page rank and domain reputation as determinants of click odds using rank, mean rank and brand

	(1) Sports	(2) Phone Numbers	(3) Fun Games	(4) Free Movies
Rank 1	15.68***	124.4***	46.41***	8.925***
	(33.76)	(19.64)	(49.47)	(32.67)
Rank 2	10.01***	69.62***	18.21***	4.473***
	(28.31)	(17.33)	(39.44)	(24.33)
Rank 3	8.271***	38.20***	9.825***	3.282***
	(27.97)	(15.00)	(31.44)	(21.10)
Rank 4	6.271***	7.851***	3.477***	1.190**
	(26.77)	(9.45)	(18.95)	(2.59)
Rank 5	4.458***	10.49***	4.395***	1.625***
	(21.50)	(11.01)	(22.02)	(5.53)
Rank 6	2.207***	2.687***	3.057***	1.759***
	(10.52)	(4.40)	(16.15)	(8.70)
Rank 7	1.589***	6.482***	2.196***	1.311***
	(5.79)	(8.72)	(10.97)	(4.40)
Rank 8	1.683***	3.602***	1.667***	1.072
	(6.56)	(5.75)	(6.88)	(1.08)
Rank 9	1.694***	2.411***	1.268**	0.951
	(6.60)	(3.67)	(2.99)	(−0.75)
Mean rank	2.306***	1.511	3.558***	0.976
	(10.31)	(0.69)	(25.17)	(−0.39)
Brand	4.312***	1.425	1.119*	1.291***
	(34.53)	(1.79)	(2.49)	(7.08)
Observations	619,528	134,907	577,590	111,161

Exponentiated coefficients; t statistics in parentheses
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

from around 9 (for "Free Movies") to over 120 (for "Phone Numbers"), and this effect is at least 50 % higher and sometimes more than twice as high as the effect of being in rank 2. The effects also decline as rank declines, roughly but not strictly monotonically.

4 Forces Behind the Impact of Rank

If page rank exerts a strong causal influence on the likelihood that users click on a website, what is the reason for that effect? In particular, to what extent is it due to the fact that higher ranked websites are more conspicuous on the page, and to what extent is it due to the reputation of the search engine for delivering relevant results in the higher ranks?

To explore this question we make use of a simple insight: The reputation of the search engine for relevance will be a substitute for any reputation for relevance that the website may have in its own right. Websites with strong positive reputations will require less assistance from the reputation of the search engine.

Table 10 Page rank and domain reputation as determinants of click odds using rank, mean rank and domain fixed effects

	(1) Sports	(2) Phone Numbers	(3) Fun Games	(4) Free Movies
Rank 1	44.33***	326.7***	49.25***	9.979***
	(41.08)	(20.54)	(46.37)	(30.08)
Rank 2	29.36***	184.0***	16.48***	5.538***
	(36.52)	(18.51)	(34.00)	(23.48)
Rank 3	9.544***	101.8***	10.66***	4.478***
	(26.41)	(16.47)	(29.22)	(19.19)
Rank 4	9.792***	12.66***	3.879***	1.469***
	(24.66)	(11.18)	(15.39)	(4.92)
Rank 5	5.226***	12.59***	4.636***	1.791***
	(22.14)	(11.79)	(21.19)	(6.53)
Rank 6	2.560***	3.262***	3.185***	1.888***
	(12.05)	(5.23)	(16.07)	(9.42)
Rank 7	1.746***	7.661***	2.241***	1.386***
	(6.87)	(9.45)	(11.14)	(5.19)
Rank 8	1.776***	3.908***	1.685***	1.105
	(7.21)	(6.10)	(7.00)	(1.55)
Rank 9	1.759***	2.517***	1.276**	0.957
	(7.06)	(3.84)	(3.07)	(−0.65)
Mean rank	0.0812***	0.0192***	1.345	0.191***
	(−5.35)	(−4.55)	(0.58)	(−5.18)
Domain 1	6.990***	6.130***	1.812***	3.495***
	(9.61)	(6.28)	(3.91)	(6.77)
Domain 2	7.665***	5.154***	1.201*	1.405***
	(28.51)	(6.08)	(2.03)	(7.88)
Domain 3	11.79***		2.367*	1.543***
	(10.33)		(2.16)	(7.13)
Domain 4				2.123***
				(6.32)
Observations	619,528	134,907	577,590	111,161

Exponentiated coefficients; t statistics in parentheses
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

We can therefore compare two alternative models of the process by which users search: In the first, reputation-based model, users compare all the domains that appear on a page and decide which is most likely to meet their needs, based on combining information based on the page rank (given the reputation of the search engine for reliability) with information based on the domain's own reputation. In this case we expect to see that the higher the reputation of the domain in its own right, the less additional benefit it will gain from being in a high rank.

In the second, conspicuousness-based model, users begin at the most conspicuous point on the results page (typically though not necessarily the first rank), and decide

Table 11 Interaction of page rank and reputation using mean rank in top ranks

	(1) Sports	(2) Phone Numbers	(3) Fun Games	(4) Free Movies
Rank 1	42.33***	89.94***	75.62***	12.35***
	(40.80)	(17.42)	(53.23)	(34.68)
Rank 2	27.12***	50.49***	26.27***	6.182***
	(36.16)	(15.31)	(48.53)	(28.82)
Rank 3	25.85***	31.73***	8.278***	3.055***
	(45.24)	(16.25)	(31.71)	(14.66)
Rank 4	5.364***	7.130***	3.034***	1.076
	(23.79)	(9.37)	(16.59)	(1.00)
Rank 5	5.141***	10.08***	4.054***	1.592***
	(23.61)	(10.90)	(20.66)	(5.28)
Rank 6	2.118***	2.587***	2.886***	1.840***
	(9.94)	(4.26)	(15.26)	(9.39)
Rank 7	1.536***	6.276***	2.146***	1.334***
	(5.36)	(8.61)	(10.64)	(4.67)
Rank 8	1.642***	3.541***	1.645***	1.067
	(6.25)	(5.68)	(6.70)	(1.00)
Rank 9	1.670***	2.388***	1.259**	0.948
	(6.42)	(3.63)	(2.90)	(−0.80)
Mean Rank	7.551***	3.573***	11.00***	2.117**
	(13.63)	(4.73)	(21.88)	(3.24)
Mean tank in top two ranks	0.748	1.396	0.248***	0.472**
	(−1.57)	(0.85)	(−10.58)	(−3.09)
Observations	619,528	134,907	577,590	111,161

Exponentiated coefficients; t statistics in parentheses
* $p < 0.05$; ** $p < 0.01$;
*** $p < 0.001$

whether to click or to continue to the next result. In this setting the reasons why a user will click immediately may be situational (such as that the user is in a hurry) or based on recognition of the domain as one with a good reputation for relevance. In this model of sequential choice, the websites with high own reputations will benefit more rather than less from being in a high rank. They have more to gain from being brought to the user's attention since they are more likely to hold such attention and convert it into a decision to click.

This suggests looking for interaction relationships between our rank variable and our separate measures of website relevance: Mean Rank and Brand. If the positive impact of being in a high rank is due principally to reputation, we should observe a smaller additional effect of reputation (as measured by our relevance indicators) for websites that appear in the higher ranks. Conversely, if it is due principally to conspicuousness, we should observe a larger additional effect of reputation (as measured by our relevance indicators) for websites that appear in higher ranks.

Table 12 Interaction of page rank and reputation using brand in top ranks

	(1) Sports	(2) Phone Numbers	(3) Fun Games	(4) Free Movies
Rank 1	39.43***	152.0***	73.61***	9.216***
	(33.61)	(23.41)	(36.25)	(18.66)
Rank 2	24.91***	84.56***	15.87***	4.648***
	(29.45)	(20.77)	(23.20)	(13.17)
Rank 3	7.579***	41.42***	9.949***	3.259***
	(26.13)	(18.46)	(29.84)	(21.41)
Rank 4	7.003***	8.219***	4.018***	1.187*
	(28.73)	(10.16)	(21.22)	(2.57)
Rank 5	4.498***	10.70***	4.716***	1.621***
	(21.57)	(11.20)	(23.10)	(5.53)
Rank 6	2.286***	2.735***	3.233***	1.755***
	(11.00)	(4.51)	(16.97)	(8.65)
Rank 7	1.628***	6.582***	2.251***	1.309***
	(6.09)	(8.85)	(11.31)	(4.37)
Rank 8	1.706***	3.632***	1.691***	1.072
	(6.74)	(5.79)	(7.07)	(1.07)
Rank 9	1.710***	2.424***	1.280**	0.951
	(6.72)	(3.69)	(3.11)	(−0.75)
Brand	5.659***	1.848***	1.521***	1.294***
	(39.47)	(7.20)	(8.35)	(6.91)
Brand in top two ranks	0.479***	0.787	1.301*	0.952
	(−7.28)	(−1.86)	(2.27)	(−0.43)
Observations	619,528	134,907	577,590	111,161

Exponentiated coefficients;
t statistics in parentheses
* $p < 0.05$; ** $p < 0.01$;
*** $p < 0.001$

Tables 11, 12 and 13 explore this question by interacting our relevance measures with page rank for each of the four query terms. For both Mean Rank and Brand, we include an interaction term for the variable for the first two ranks only.[11] If the coefficient on this interaction variable is greater than one, relevance is more important for websites in higher ranks; if it is less than one, relevance is less important in higher ranks.

The results tend to indicate a lower effect of Mean Rank in the top two ranks than in the remaining ranks, but the evidence is not unequivocal. For two out of four query terms the interaction term is strongly and significantly <1, while for one other query it is insignificantly less than one and for the other it is insignificantly greater than one. For Brand three of the interaction terms are less than one but only one is significantly so, while the other is significantly >1. When both sets of regressors are included

---

[11] Interacting the variable for more than two ranks creates collinearity problems since a large proportion of the observations are dominated by the presence of domains that appear in the top ranks.

Table 13 Interaction of page rank and reputation using both mean rank in top ranks and brand in top ranks

	(1) Sports	(2) Phone Numbers	(3) Fun Games	(4) Free Movies
Rank 1	20.37***	20.16***	91.15***	8.671***
	(25.60)	(8.72)	(38.15)	(16.14)
Rank 2	13.07***	11.28***	32.30***	4.362***
	(21.91)	(7.07)	(29.22)	(11.73)
Rank 3	8.360***	182.7***	8.670***	3.676***
	(26.64)	(17.47)	(28.75)	(15.07)
Rank 4	8.230***	16.69***	3.008***	1.288**
	(28.89)	(12.21)	(16.35)	(3.14)
Rank 5	4.821***	14.10***	4.051***	1.667***
	(22.26)	(12.27)	(20.64)	(5.75)
Rank 6	2.414***	3.664***	2.882***	1.783***
	(11.64)	(5.72)	(15.24)	(8.80)
Rank 7	1.685***	8.448***	2.144***	1.327***
	(6.50)	(9.88)	(10.63)	(4.56)
Rank 8	1.740***	4.104***	1.643***	1.083
	(6.98)	(6.32)	(6.68)	(1.24)
Rank 9	1.733***	2.583***	1.258**	0.954
	(6.89)	(3.95)	(2.89)	(−0.71)
Mean rank	0.298***	0.00138***	12.02***	0.534
	(−5.97)	(−6.76)	(18.69)	(−1.82)
Mean rank in top two ranks	19.96***	32629.0***	0.237***	1.858
	(12.87)	(8.62)	(−9.03)	(1.77)
Brand	6.670***	14.82***	0.929	1.382***
	(36.67)	(8.37)	(−1.27)	(6.27)
Brand in top two ranks	0.210***	0.0293***	0.867	0.886
	(−14.07)	(−8.76)	(−1.12)	(−0.99)
Observations	619,528	134,907	577,590	111,161

Exponentiated coefficients; *t* statistics in parentheses * *p* < 0.05; ** *p* < 0.01; *** *p* < 0.001

together, no clear pattern emerges, though the coefficients indicate the likelihood of significant collinearity.

On balance, the evidence is suggestive rather than conclusive. Nevertheless, it suggests that reputation is a stronger force than conspicuousness in explaining the causal impact of page rank on click probabilities, but that conspicuousness has a role to play as well.

5 Conclusion

We have shown in this paper that when a website appears in a high rank on a Search Engine Results Page it has a substantial and highly significant positive causal effect on

the probability that a user will click on the website. We have done so using a unique data set that allows us to abstract from the fact that search engines determine rank partly by predicting the likely relevance of websites to user needs.

We have shown that this estimation is robust to possible concerns about the endogeneity of page ranking. We have further provided evidence that suggests that rank influences CTRs somewhat more by substituting the reputational capital of the search engine for the reputation of individual websites. However, there is also some evidence that conspicuousness plays a role as well, which implies that one of the assets that search engines deploy is access to the scarce attention of users.

References

Armstrong, M., Vickers, J., & Zhou, J. (2009). Prominence and consumer search. *RAND Journal of Economics*, *40*, 209–233.

Armstrong, M., & Zhou, J. (2011). Paying for prominence. *The Economic Journal*, *121*, F368–F395.

Athey, S., & Meidan, M. (2011). Exchange rate fluctuations, consumer demand, and advertising: The case of internet search. Working paper.

Athey, S., & Nekipelov, D. (2011) A structural model of sponsored search advertising auctions. Working paper.

Baye, M., De Los Santos B., & Wildenbeest, M. (2012) Whats in a name? Measuring prominence and its impact on organic traffic from search engines, Kelley School of Business BEPP 9.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*, 107–117.

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

Jerath, K., Ma, L., Park, Y. H., & Srinivason, K. (2011). A position paradox in sponsored search actions. *Marketing Science*, *30*, 612–627.

Smith, M., & Brynjolfsson, E. (2001). Consumer decision-making at an internet shopbot: Brand still matters. *The Journal of Industrial Economics*, *49*, 541–558.

Xu, Y., & Kim, H. W. (2008). Order effect and vendor inspection in online comparison shopping. *Journal of Retailing*, *84*, 477–486.