How does ranking affect user choice in online search?[*,†]

Mark Glick, Greg Richards, Margarita Sapozhnikov, Paul Seabright[‡]

October 11, 2013

Abstract

This paper investigates whether a search engine's ordering of algorithmic results has an important impact on website traffic. A website's ranking on a search engine results page is positively correlated with the clicks it receives. This could result from the search engine 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 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.

Keywords: internet search, page rank, click-through rates, scarce attention.

JEL codes: D03,D12,D83

[*]The analysis for this paper was performed in conjunction with ongoing work for Microsoft Corporation; Keystone and Seabright acknowledge financial support from Microsoft. The primary data used in the analysis come from log files from Microsoft's Bing search engine and include the website names, search result rankings, and click-through rates for results presented in response to user search queries on Bing.com. Supplementary data for additional website click-through rates were derived from additional Microsoft opt-in consumer panel data containing online behavior

[†]The authors are grateful to Scott Gingold, who was involved with the project from the beginning and gave us invaluable feedback, ideas, encouragement and support; and to Thierry Magnac, who gave us invaluable econometric expertise and spent much valuable time helping us to understand the data.

[‡]Authors' names are presented in alphabetical order, and their affiliations are, respectively: University of Utah and Keystone Strategy; Keystone Strategy; Charles River Associates; and Toulouse School of Economics/Institute for Advanced Study in Toulouse.

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 of websites[1] is positively correlated with click-through rate (CTR)[2]. If this correlation reflected a causal impact of ranking on CTR, 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 CTR 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 CTR) 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 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 undertaken. We concede that no one statistical method completely eliminates endogeneity concerns;

[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 associated with that hyperlink.

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

however, our results are robust and all evidence points to very high economic significance of the algorithmic rank[3].

We proceed as follows. In section 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 section 3 we demonstrate econometrically the impact 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. 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 con-

[3] There are no studies to our knowledge in the economic literature that estimate the effect of rank in the algorithmic search results. Susan Athey [2] [1] are important papers on the analysis of user behavior in the paid results.

stant 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 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 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 there is 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.

Finally, query terms which arise from non-uniform intent across users are also unsuitable. One example is the query “eclipse”. Based on the websites 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 known as Eclipse, and to search for one of the Twilight Saga books with this title (a teenage vampire romance novel).

Thus, we manually sorted through an extensive list of queries, and found only four query terms that were suitable for our purposes. In alphabetical order these are: “free movies”, “fun games”, “phone numbers” and “sports”. [4]

2.2 Algorithm

Algorithms are typically 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 by Brin and Page that introduced Google [4] states that “Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems.” The fundamentals of a Search Engine algorithm are techniques based on natural language, and data based on the topology of the internet.

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

[4] 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.

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

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 CTR that determines ranking rather than ranking determining future CTR. 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 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 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

[6] Static Rank is computed based on the ontological map of all webpages, consisting of nodes and links between them. Given these interconnections, Static Rank assigns a score to each website. This score represents a 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 on a cluster, for example.

of views for the time period analyzed) for the query term "Phone numbers'.

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

In addition, the frequency with which the top three websites appear in the top rank is also 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), and while one 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 every rank compared to whitepages.com. This is consistent with the observation that phonenumber.com is observed in rank 1 more often.

Tables 2 - 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, Figure 1 has the time series of the daily CTR (blue line) and daily percent of views in Rank 1 (red 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 do not explain the switch between the ranks.

This is confirmed by Granger causality tests that were run for both websites, the results of which are reported in Table 5. 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 CTR: the F-statistic for the exclusion of the percentage of time spent in Rank 1 from the equation for CTR is significant at 1% for one domain and 0.1% for the other. On the other hand, we utterly failing 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.

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 well known multinomial logit model 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. This therefore allows us to abstract from the many factors that can affect click through rates, such as time of day, since these factors do not vary between alternatives presented to the user in a given page view.

The results are presented in Tables 6 through 9 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 a positive effect.

There is a large variation between query terms in the magnitude of the rank effects, but the broad qualitative findings are remarkably similar. Specification (1) in each table gives the effect of rank without controlling for website relevance. 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 specification (2) in each table, 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, reported in specification (3), is to use a dummy variable we call "Brand" for any website that appears more than 75 times in rank 1 during the sample period. While evidently somewhat arbitrary as a definition, it nevertheless captures the idea that such websites are likely to be perceived as more relevant. Adding this variable to the specification including Mean Rank reduces further the odds ration for rank 1, by between 10 per cent for "Free movies" and 70 per cent for "Phone numbers".

As an control we use separate fixed effects for each of the "Brand" websites instead of a single dummy variable, as reported in specification (4). This has a slightly smaller effect on the odds of being in rank 1 than does the "Brand" dummy.

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 compared to rank 10. Even in the most conservative specification (number 3), the odds ratios vary from around 9 (for "Free movies") to over 30 (for "Fun games"), and this effect is at least 60% higher and sometimes more than twice as high as the effect of being in rank2. 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 the website may have in its own right. Websites with strong positive reputations will require less assistance from the reputation of the search engine.

Similarly, being in a high rank will be less valuable for a website that has a strong reputation in its own right. This helps us to look for interaction relationships between our rank variable and our separate measures of website relevance, Mean Rank and Brand.

Thus, 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, the ability of the search engine to increase the conspicuousness of a website on the results page should be complementary to the website's own reputation. Higher reputation websites 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.

Therefore, if the positive impact of being in a high rank 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.

Tables 10 through 13 explore this question by interacting our relevance measures with page rank. For both Mean Rank and Brand, we include an interaction term for the variable for the first five ranks only. 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 depend on the measure used. The effect of Mean Rank is always lower in the top five ranks than the bottom five, giving clear support to the hypothesis that page rank impacts click probabilities principally via the reputation mechanism. However, the effect of Brand is significantly higher in the top five ranks for two of the four query terms, and insignificantly lower for the other two query terms.

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 suggesting that rank influences CTRs primarily 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, implying that one of the assets that search engines deploy is access to the scarce attention of users.

References

[1] Athey, S., Cohen Meidan, M., *Exchange Rate Fluctuations, Consumer Demand, and Advertising: The Case of Internet Search*, 2011.

[2] Athey, S. and Nekipelov, D., *A Structural Model of Sponsored Search Advertising Auctions*, 2011.

[3] Arellano, M., *Panel Data Econometrics.* Cambridge University Press, 2004.

[4] Brin, S. and Page, L., *The Anatomy of a Large-Scale Hypertextual Web Search Engine* Seventh International World-Wide Web Conference (WWW 1998), 1998.

[5] Devine, K. L., *Preserving Competition in Multisided Innovative Markets: How Do You Solve a Problem Like Google?*, North Carolina Journal of Law and Technology, vol. 10, Issue 1, 2008.

[6] Edelman, B, Ostrovsky, M and Schwartz, M., *Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords* American Economic Review, 2007, vol.97, pp 242-259.

[7] Goldman,E., *Search Engine Bias and the Demise of Search Engine Utopianism.* Yale Journal of Law and Technology, 2006, vol. 14, pp. 121-133.

[8] Litvak, N., Scheinhardt, W.R.W., Volkovich, Y., *In-Degree and PageRank of Web pages: Why do they follow similar power laws?*, Memorandum 1807 Department of Applied Mathematics, University of Twente, Enschede, 2006.

[9] Manne, G. A., Wright, J.D., *Google and the Limits of Antitrust: The Case Against the Antitrust Case Against Google.* Harvard Journal of Law and Public Policy, 2011, vol. 34, No.1.

[10] Padney, S., Roy, S., Olston, C., Cho, J., Chakrabarti, S., *Shuffling a Stacked Deck: The Case for Partially Randomized Ranking of Search Engine Results*, Proceedings of the 31st International Conference on Very Large Data Bases, 2005, pp. 781-792.

TABLES

Table 1: Top Five websites for "Phone Numbers"

Website/ Rank	1	2	3	4	5	6	7	8	9	10
phonenumber.com										
Views	17,075	13,315	1,417	9	6	5	1	0	3	1
CTR	0.295	0.168	0.1	0	0	0	0	–	0.33	0
whitepages.com										
Views	14,652	16,558	580	1	10	5	0	0	7	13
CTR	0.274	0.154	0.097	0	0.1	0	–	–	0	0
en.wikipedia.org										
Views	1	12	8	229	4,055	21,648	4,142	1,288	349	74
CTR	0	0	0	0	0.001	0.001	0.001	0.003	0	0
switchboard.com										
Views	80	1,893	29,734	36	19	22	6	4	1	3
CTR	0.625	0.098	0.054	0.111	0.158	0.091	0	0	0	0
anywho.com										
Views	5	0	15	8,645	6,185	1,933	9,653	3,650	1,428	201
CTR	0.8	–	0.067	0.028	0.021	0.014	0.015	0.012	0.009	0

Table 2: Top Five websites for "Free Movies"

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

Table 3: Top Five websites for "Fun Games"

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

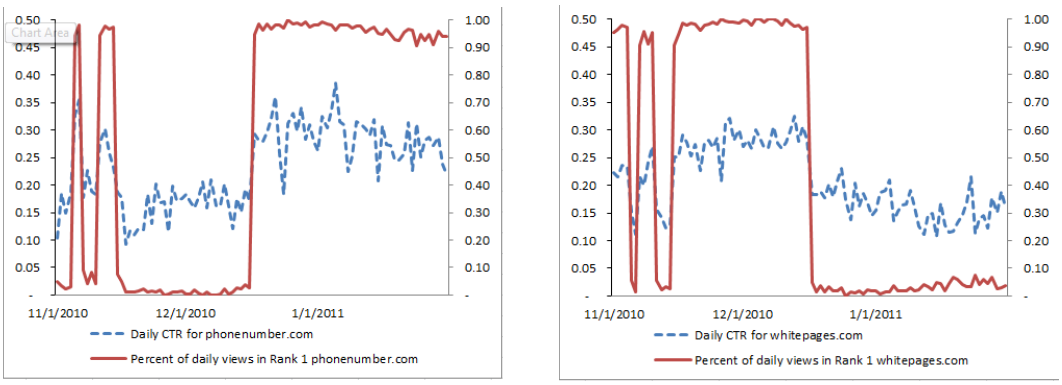
Table 4: Top Five websites for "Sports"

Website/ Rank	1	2	3	4	5	6	7	8	9	10
sports.com										
Views	21	343	2,557	1,717	41,364	38,514	27,411	3,549	450	40
CTR	0.286	0	0.007	0.008	0.006	0.004	0.005	0.002	0.009	0
espn.go.com										
Views	48,027	65,521	1,873	15	304	95	10	7	19	18
CTR	0.207	0.132	0.1	0.067	0.066	0.084	0.1	0	0.053	0
sports.yahoo.com										
Views	66,744	48,290	448	4	40	8	6	8	13	7
CTR	0.273	0.194	0.076	0	0.05	0	0	0.125	0.231	0.143
msn.foxsports.com										
Views	875	664	104,481	284	6,816	421	133	7	50	77
CTR	0.717	0.13	0.101	0.049	0.049	0.076	0.03	0	0.02	0.052
sportsillustrated.cnn.com										
Views	166	1	131	261	38,243	14,065	7,527	14,548	19,488	9,470
CTR	0.663	1	0.031	0.031	0.023	0.019	0.014	0.014	0.016	0.016

Table 5: **GRANGER CAUSALITY: DAILY CTR AND % IN RANK 1**

		Predict CTR Exclude % Rank 1	Predict % Rank 1 Exclude CTR
Query Term	**Domain**	**F-Statistics**	
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

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



(a) phonenumber.com (b) whitepages.com

Figure 1: CTR and % of views in Rank 1 daily. Query "Phone Numbers"

Table 6: Page rank and domain reputation as determinants of click odds: "Free movies"

	(1) Rank only	(2) Rank and Mean Rank	(3) Rank and Brand	(4) Domain fixed effects
Rank 1				
	11.30***	10.95***	8.925***	9.979***
	(50.07)	(39.41)	(32.67)	(30.08)
Rank 2				
	5.657***	5.574***	4.473***	5.538***
	(34.43)	(32.23)	(24.33)	(23.48)
Rank 3				
	3.653***	3.607***	3.282***	4.478***
	(24.75)	(23.56)	(21.10)	(19.19)
Rank 4				
	1.187*	1.180*	1.190**	1.469***
	(2.57)	(2.46)	(2.59)	(4.92)
Rank 5				
	1.649***	1.639***	1.625***	1.791***
	(5.72)	(5.64)	(5.53)	(6.53)
Rank 6				
	1.926***	1.919***	1.759***	1.888***
	(10.34)	(10.25)	(8.70)	(9.42)
Rank 7				
	1.376***	1.373***	1.311***	1.386***
	(5.23)	(5.19)	(4.40)	(5.19)
Rank 8				
	1.085	1.083	1.072	1.105
	(1.26)	(1.24)	(1.08)	(1.55)
Rank 9				
	0.951	0.951	0.951	0.957
	(-0.76)	(-0.76)	(-0.75)	(-0.65)
Mean Rank		1.055	0.976	0.191***
		(0.85)	(-0.39)	(-5.18)
Brand			1.291***	
			(7.08)	
Domain Name 1				3.495***
				(6.77)
Domain Name 2				1.405***
				(7.88)
Domain Name 3				1.543***
				(7.13)
Domain Name 4				2.123***
				(6.32)
Observations	111161	111161	111161	111161

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Page rank and domain reputation as determinants of click odds: "Fun games"

	(1) Rank only	(2) Rank and Mean Rank	(3) Rank and Brand	(4) Domain fixed effects
Rank 1				
	145.0***	50.23***	33.83***	36.01***
	(84.02)	(55.34)	(48.88)	(42.30)
Rank 2				
	31.30***	20.09***	13.25***	12.11***
	(57.48)	(48.29)	(40.42)	(29.96)
Rank 3				
	14.67***	10.79***	7.109***	7.702***
	(44.11)	(38.40)	(30.74)	(24.91)
Rank 4				
	4.059***	3.470***	3.562***	4.659***
	(21.39)	(18.92)	(19.31)	(18.17)
Rank 5				
	4.836***	4.419***	3.396***	3.448***
	(23.50)	(22.12)	(17.97)	(16.16)
Rank 6				
	3.282***	3.065***	2.690***	2.681***
	(17.19)	(16.19)	(14.25)	(13.08)
Rank 7				
	2.263***	2.198***	2.171***	2.183***
	(11.38)	(10.98)	(10.80)	(10.42)
Rank 8				
	1.691***	1.666***	1.681***	1.720***
	(7.07)	(6.87)	(6.99)	(7.05)
Rank 9				
	1.280**	1.268**	1.256**	1.287**
	(3.11)	(2.99)	(2.88)	(3.14)
Mean Rank		3.703***	3.532***	0.00878***
		(27.40)	(26.64)	(-8.01)
Brand			2.034***	
			(23.64)	
Domain Name 1				15.63***
				(15.16)
Domain Name 2				1.422***
				(7.69)
Domain Name 3				5.426***
				(15.43)
Domain Name 4				2.842***
				(27.01)
Domain Name 5				236.3***
				(11.75)
Domain Name 6				1.198*
				(2.01)
Observations	577590	577590	577590	577590

Table 8: Page rank and domain reputation as determinants of click odds: "Phone numbers"

	(1) Rank only	(2) Rank and Mean Rank	(3) Rank and Brand	(4) Domain fixed effects
Rank 1				
	220.5***	101.2***	28.37***	55.47***
	(26.93)	(21.27)	(9.99)	(11.75)
Rank 2				
	120.6***	56.68***	15.86***	31.22***
	(23.89)	(18.65)	(8.26)	(10.07)
Rank 3				
	43.58***	30.28***	8.102***	16.47***
	(18.73)	(16.60)	(6.26)	(8.17)
Rank 4				
	8.212***	6.982***	7.143***	25.57***
	(10.15)	(9.34)	(9.45)	(13.81)
Rank 5				
	10.71***	10.00***	10.02***	16.74***
	(11.21)	(10.88)	(10.88)	(12.98)
Rank 6				
	2.733***	2.566***	2.589***	4.422***
	(4.51)	(4.23)	(4.27)	(6.51)
Rank 7				
	6.579***	6.231***	6.312***	9.901***
	(8.84)	(8.58)	(8.64)	(10.57)
Rank 8				
	3.628***	3.529***	3.569***	4.447***
	(5.79)	(5.66)	(5.71)	(6.68)
Rank 9				
	2.424***	2.384***	2.381***	2.680***
	(3.69)	(3.62)	(3.61)	(4.10)
Mean Rank		4.295***	3.613***	0.0000183***
		(9.33)	(8.00)	(-9.55)
Brand			3.950***	
			(4.99)	
Domain Name 1				1650.0***
				(11.69)
Domain Name 2				30.43***
				(10.00)
Domain Name 3				1188.2***
				(11.59)
Observations	134907	134907	134907	134907

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Page rank and domain reputation as determinants of click odds: "Sports"

	(1) Rank only	(2) Rank and Mean Rank	(3) Rank and Brand	(4) Domain fixed effects
Rank 1				
	105.8***	39.02***	13.72***	22.60***
	(72.79)	(48.35)	(33.68)	(33.98)
Rank 2				
	66.45***	25.02***	8.804***	15.03***
	(65.36)	(42.64)	(28.05)	(29.43)
Rank 3				
	39.73***	26.99***	7.874***	4.978***
	(57.10)	(49.63)	(30.06)	(18.97)
Rank 4				
	7.049***	5.512***	9.242***	7.508***
	(28.84)	(24.94)	(31.82)	(23.25)
Rank 5				
	5.860***	5.211***	2.604***	2.291***
	(25.77)	(24.00)	(13.64)	(11.37)
Rank 6				
	2.324***	2.137***	2.015***	1.894***
	(11.22)	(10.10)	(9.28)	(8.32)
Rank 7				
	1.629***	1.545***	1.752***	1.681***
	(6.10)	(5.44)	(6.98)	(6.44)
Rank 8				
	1.697***	1.648***	1.467***	1.442***
	(6.67)	(6.30)	(4.81)	(4.59)
Rank 9				
	1.708***	1.674***	1.237**	1.223*
	(6.71)	(6.45)	(2.64)	(2.50)
Mean Rank		6.165***	2.880***	13.60***
		(24.71)	(13.72)	(5.73)
Brand			6.717***	
			(59.36)	
Domain Name 1				1.599*
				(2.47)
Domain Name 2				7.951***
				(30.09)
Domain Name 3				1.782*
				(2.56)
Domain Name 4				6.600***
				(52.64)
Observations	619528	619528	619528	619528

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Interaction of page rank and reputation: "Free movies"

	(1) Mean Rank in top ranks	(2) Brand in top ranks
Mean Rank	2.017*	
	(2.14)	
Mean Rank in top ranks	0.511*	
	(-2.01)	
Brand		1.167*
		(2.55)
Brand in top ranks		1.172*
		(2.04)
Rank 1		
	12.02***	8.291***
	(32.24)	(31.63)
Rank 2		
	6.075***	4.194***
	(26.21)	(21.29)
Rank 3		
	3.927***	3.166***
	(19.71)	(20.17)
Rank 4		
	1.282**	1.181*
	(3.14)	(2.49)
Rank 5		
	1.775***	1.616***
	(5.95)	(5.49)
Rank 6		
	1.847***	1.819***
	(9.25)	(8.95)
Rank 7		
	1.337***	1.337***
	(4.65)	(4.67)
Rank 8		
	1.068	1.077
	(1.01)	(1.15)
Rank 9		
	0.948	0.951
	(-0.79)	(-0.75)
Observations	111161	111161

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Interaction of page rank and reputation: fun_games

	(1) Mean Rank in top ranks	(2) Brand in top ranks
Mean Rank	14.31***	
	(9.73)	
Mean Rank in top ranks	0.248***	
	(-4.97)	
Brand		1.507***
		(10.06)
Brand in top ranks		2.080***
		(11.72)
Rank 1		
	61.27***	53.54***
	(50.03)	(51.41)
Rank 2		
	24.03***	11.57***
	(43.97)	(31.42)
Rank 3		
	12.85***	5.448***
	(35.68)	(21.56)
Rank 4		
	4.112***	2.905***
	(19.04)	(14.72)
Rank 5		
	5.223***	2.288***
	(21.97)	(10.32)
Rank 6		
	2.832***	3.065***
	(14.58)	(16.12)
Rank 7		
	2.132***	2.256***
	(10.51)	(11.34)
Rank 8		
	1.639***	1.699***
	(6.64)	(7.13)
Rank 9		
	1.256**	1.275**
	(2.87)	(3.06)
Observations	577590	577590

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Interaction of page rank and reputation: phone_numbers

	(1) Mean Rank in top ranks	(2) Brand in top ranks
Mean Rank	3107.1***	
	(8.56)	
Mean Rank intopranks	0.00124***	
	(-6.99)	
Brand		7.109**
		(3.15)
Brand in top ranks		0.836
		(-0.26)
Rank 1		
	243.2***	37.48***
	(20.98)	(10.24)
Rank 2		
	136.0***	20.52***
	(18.81)	(8.54)
Rank 3		
	70.65***	7.423***
	(17.14)	(5.66)
Rank 4		
	16.04***	8.234***
	(11.16)	(10.16)
Rank 5		
	22.81***	10.64***
	(12.46)	(11.16)
Rank 6		
	1.984**	2.731***
	(3.04)	(4.51)
Rank 7		
	5.032***	6.620***
	(7.54)	(8.87)
Rank 8		
	3.236***	3.660***
	(5.26)	(5.82)
Rank 9		
	2.246***	2.417***
	(3.36)	(3.68)
Observations	134907	134907

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Interaction of page rank and reputation: sports

	(1) Mean Rank in top ranks	(2) Brand in top ranks
rank		
Mean Rank	22304.9***	
	(21.76)	
Mean Rank intopranks	0.000234***	
	(-17.87)	
brand		7.289***
		(45.45)
brandintopranks		0.923
		(-1.25)
Rank 1		
	113.0***	25.01***
	(48.14)	(39.73)
Rank 2		
	72.34***	15.88***
	(43.73)	(34.10)
Rank 3		
	73.94***	10.10***
	(48.66)	(28.55)
Rank 4		
	14.90***	10.90***
	(30.24)	(33.91)
Rank 5		
	13.92***	2.859***
	(29.58)	(13.07)
Rank 6		
	1.426***	2.075***
	(4.47)	(9.66)
Rank 7		
	1.215*	1.801***
	(2.39)	(7.33)
Rank 8		
	1.458***	1.471***
	(4.73)	(4.85)
Rank 9		
	1.529***	1.226*
	(5.29)	(2.53)
Observations	619528	619528

Exponentiated coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$