How Does Page Rank Affect User Choice in Online Search?[*†]

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

November 29, 2011

Abstract

We investigate whether a website's rank on a search results page affects likelihood of a click to that site, using a dataset of user search behavior. A strong correlation between ranking and click-through rates partly reflects the search engine accurately predicting the relevance of websites to users' needs. Controlling for relevance using domain and time fixed effects significantly reduces the coefficients on high rank, but they remain highly significant, statistically and economically, and larger for high-ranking domains. This implies that rank effects are complementary to branding effects, and therefore that rank works via conspicuousness on the page, not via reputation.

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

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 domain names, search result rankings, and click-through rates for results presented in response to user search queries on Bing.com. Supplementary data for additional domain 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 and improving the rigor and robustness of our estimation strategy

[‡]Authors' names are presented in alphabetical order, and their affiliations are, respectively: University of Utah and Keystone Strategy, Keystone Strategy, Keystone Strategy, and Toulouse School of Economics

1 Introduction

This paper seeks to estimate the extent to which search engines can influence traffic to a website by modifying the rank of that website on its results page. This is a challenging phenomenon to investigate, since it is far from obvious whether the well-known correlation between websites' rankings and the traffic they receive occur because ranking *determines* traffic or because ranking accurately *predicts* traffic[6] [8]. But the question is important since the answer affects our view about whether search engines merely reduce the transaction costs of establishing commercial relationships or whether they affect the entire competitive landscape by determining which websites do and do not attract enough traffic to establish economic viability. Recent anti-trust investigations of the internet search market in the US and Europe have considered whether search engines have the ability to manipulate rankings to influence the distribution of traffic among sites (whether they also have an incentive to do so is a separate question) [9].The influence of ranking on traffic may be limited by the fact that consumers have potential counter strategies, including the avoidance of less relevant links at the top of the search engine results page (hereafter SERP) by scanning down and clicking on other sites that are more relevant [7] and switching search engines, assuming alternative search engines have sufficient scale and capability to deliver the most relevant results.

If the correlation between click-through rates (CTRs) and ranking is simply the result of the search engine correctly predicting the relevance of a site to consumers, modification of ranking should have little or no impact on site traffic. What matters is how large is the effect of ranking on traffic after we have separated out the causal influence running in the other direction.

Our paper addresses this question by controlling for a site's relevance to consumers independent of its rank on the SERP. We study econometrically the determinants of a domain's click-through rate and find that position (rank) on the search engine results page is economically very important and a highly statistically significant determinant of traffic, even accounting for the independent determinants of a website's relevance to the needs of users. This allows us to estimate the impact of algorithm modification on a CTR, illustrating that it is an extremely effective method for influencing (positively or negatively) the amount of traffic to a domain.

Indeed, we find that a website in rank 1 has, on average, a CTR anywhere between 5.3 and 26.1 percentage points higher than it would have had on average outside of the top 3 ranks, taking into account its intrinsic relevance to users' needs. To see what a large impact this is, note that in our dataset, websites that rank below the first four positions on the SERP have average CTRs of between around 0.2% and 3.7%, depending on the rank and the query, and promotion from rank below 3rd to the top rank would increase their traffic on average by between 6 and nearly 30 times. Conversely, our estimates imply that if a domain were moved from rank 1 to a random rank below 3rd, its CTR would decrease by between 17% and 87 %. There is considerable variation in the magnitude of these effects depending upon the type of query undertaken, but the broad qualitative findings appear robust to this variation. We conclude that a search engine does have the ability to influence very substantially the traffic to destination websites by modification of its search algorithm. To the extent that such traffic may in turn determine the economic viability of rival websites, this ability enables the search engine, should it so choose and should it attract a large enough share of overall searches, to influence the entire landscape of competition in complementary web-based markets.

We proceed here as follows. In section 2, we review the existing literature. In section 3, we sketch some simple principles guiding the hypotheses we wish to test. In section

4, we outline the data we use for our study. In section 5, we describe and implement our estimation methodology. In section 6, we interpret the results in terms of the implied impact of algorithmic modification on CTRs.

2 Literature Review

The relationship between rank and click through rate is non-coincidental and has received a fair amount of attention in the existing literature. The algorithms that assign rank to a domain (PageRank and similar algorithms) have been developed and studied in the computer science literature. This literature has long asserted the hypothesis that click through rate should be one of the determinants of the domain rank on SERP. However, to our knowledge there have not been direct studies of the reverse effect, namely the effect of rank on domain popularity, although the possibility of such effects has been acknowledged.

The increasing market power and influence of search engines have made scholars of law pay close attention to the dynamics of internet search. A number of law review articles have mentioned the relationship between domain rank and its CTR. For example, Goldman [6] mentions that result ranking impacts CTRs, and Devine [4] goes further to suggest that Google may manipulate ranking, and that such manipulation can adversely affect the domain CTR.

The economic theory literature has investigated the properties of auction mechanisms in the on-line advertiser market (Edelman et al. [5]). These models are based on the assumption that higher ranking spots mean more clicks and visibility and thus are more valuable. The vast majority of relevant economic empirical literature is based on on-line advertiser datasets, which describe the placement of ads on the search page and the ads' click-through-rates. This literature acknowledges that the domain rank on SERP is one of the determinants of the domain CTR. Dummy variables are used to control for the effect of rank, when the user choices to click or not on a given domain are investigated. However the endogeneity of such dummy variables is also implied and the direct effect of rank on CTR has never been estimated. For examples of recent empirical papers see Athey and M. Cohen Meidan [1] and Athey and Nekipelov [2].

Our paper closes a gap in the existing literature, as our data collection method allows us to estimate directly the effect of the domain rank on the consumer decision to click on that domain.

3 Search and the scarcity of user attention

We begin with a discussion of how general web search engines generate results on their SERPS when users conduct a search. The Internet is a network of interrelated web pages containing information useful to different users at different times. Search engines analyze these interconnected page links and apply complicated probability distribution algorithms to assign weighted importance to, and thus measure the relevance to users of, each of the web pages they have indexed.

When a user conducts a query using a search engine, the search engine instantaneously considers the web pages in its index for inclusion on its SERP. The algorithm used by a search engine has two components: choosing the domains that match the query term and ranking them. The first component uses natural language techniques, which are keyword-focused measures that look at the placement and count of the query term words in a domain name, anchor text[1] and domain content. Once the set of domains to be displayed in the SERP is determined, these domains are ranked using natural language techniques, static rank[2] and user behavior data.

According to Google research, users evaluate the SERP "so quickly that they make most of their [click] decisions unconsciously.[3]" Users "start from the first result and continue down the list until they find a result they consider helpful and click it — or until they decide to refine their query." [4] The click-through rates that we observe may be driven by some combination of the observed relevance of the site (the users recognize the site brand name or the caption text as being most appropriate to their search intent) and the site's rank on the SERP (the user assumes the search engine may know best based on page rank scores).

Internet search is an economically important activity because the time and attention of the searcher are scarce. It is obvious and well-known that the time taken to load a search engine results page may create a cost for users in looking beyond the first page of results. Less obvious, but no less important, is the possibility that certain places on the SERP are more conspicuous than others and therefore more likely to attract the attention of searchers. In principle, there could be reasons for this rooted in the psychology of the user: for instance, many users who are familiar with the Roman alphabet read from left to right and from top to bottom on a page, though readers of Arabic, Hebrew or Chinese do not. This may in turn increase the conspicuousness of results located near the top or at the left of the page. Alternatively, there could be reasons rooted in the

[1] Anchor text is the text in a hyperlink that leads to the domain.

[2] 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 domain. This score represents the probability that a person starting at a random page and randomly clicking on links will arrive at the domain in question.

[3] http://googleblog.blogspot.com/2009/02/eye-tracking-studies-more-than-meets.html.

[4] http://googleblog.blogspot.com/2009/02/eye-tracking-studies-more-than-meets.html.

experience of the user with previous searches: for instance, if higher-ranked results have generally been more relevant to the user in the past, the user may systematically look at higher-ranked results first.

It seems likely that the characteristics that contribute to the probability that a user clicks on a search result fall roughly into two types: the capacity to catch the user's initial attention and the capacityto hold the user's attention and convert that attention into an intent to click on a search result. We call the first of these the "catching effect" and the second the "holding effect". In the first category will be characteristics that contribute to conspicuousness (size, color and location on the page), while in the second will be characteristics that contribute to a sense of reliability or the stimulation of curiosity. The former characteristics are chosen by the search engine, while the latter will depend both on the search engine (if its results in a particular place are known to be particularly reliable or interesting) and also on the characteristics of the website (its brand recognition, the catchiness of its name, its reputation for reliability or value, and so on). We can therefore speak more precisely of a "brand holding effect" if the holding effect comes about due to the visible characteristics of the website, and a "rank holding effect" if it comes about through the page rank.

It seems reasonable to draw two main empirical conclusions. First, because the capacity to hold users' attention is reliant upon the prior ability to catch the users' attention, even websites with a highly attractive set of characteristics will still be dependent, to a significant extent, on conspicuousness, and not all ranks on the SERP are equally conspicuous. Thus we expect that some page ranks (probably, but not necessarily, those near the top of the page) will have a statistically and economically positive significant impact on CTR for all websites. Second, we expect the page ranks that increase CTR to be positively correlated with intrinsic relevance, since that is an important reason why users will be influenced by those page ranks in the first place. Accordingly, controlling for intrinsic relevance should be expected to reduce the correlation between page rank and CTR.

A further empirical question concerns the interaction between the influence of rank and that of the website's intrinsic characteristics (including any reputation or brand loyalty it may attract); are they substitutes or complements? Although it might seem more plausible that they would be substitutes, the answer is likely to depend on whether page rank contributes more to the catching effect or the holding effect. Even a website with strong brand loyalty will be able to transform its users' loyalty into clicks only if it first succeeds in being noticed by users. Therefore, if page rank works mainly via conspicuousness, the rank and brand effects would instead be complements. To see this more precisely, consider a single page view. In equation 1 we express the probability that a website in rank i and with visible intrinsic characteristics r is clicked on during that view as $C(i,r)$, where

$$C(i,r) = p(i)(q(r) + (1 - q(r))t(i)) \tag{1}$$

where

$p(i)$ denotes the probability that the website catches the user's attention, which is a function of rank only – this is the catching effect;

$q(r)$ denotes the probability that the website that has caught the user's attention manages to hold it through its visible characteristics – this is the "brand holding effect" and

$t(i)$ denotes the probability that the website that has caught the user's attention but has not managed to hold it through its visible characteristics is able nevertheless to hold it through its page rank. This is the "rank holding effect". Note that this will be strictly less than one for all i.

We can then take the derivative of $C(i, r)$ with respect to i as

$$\partial C(i, r)/\partial i = p'(i)q(r) + (1 - q(r))[p'(i)t(i)) + p(i)t'(i)] \quad (2)$$

and the cross-derivative of $C(i, r)$ with respect to i and r as

$$\partial^2 C(i, r)/\partial i \partial r = q'(r)[p'(i)(1 - t(i)) - p(i)t'(i)] \quad (3)$$

Whether this is positive or negative, and therefore whether the rank and brand effects are complements or substitutes, depends on the relative strength of the catching effect of rank and the holding effect of rank. Consider what happens if catching the user's attention is easy (so that there is no catching effect of rank), while holding it is not. Then all sites on a page will catch the user's attention with probability one, so that $p(i) = 1$ and $p'(i) = 0$. This means that equation (3) simplifies to

$$\partial^2 C(i, r)/\partial i \partial r = -q'(r)t'(i) \quad (4)$$

which will be strictly negative if both the brand holding effect and the rank holding effect are strictly positive.

Alternatively, consider what happens if the catching effect is strong but there is no rank holding effect. Then equation (3) simplifies to

$$\partial^2 C(i,r)/\partial i \partial r = q'(r)p'(i)(1-t(i)) \quad (5)$$

which will be strictly positive if both the catching effect and the brand holding effect are strictly positive. This implies that those websites that have relatively attractive intrinsic characteristics will have a higher influence of page rank on CTR if the catching effect of rank dominates the holding effect of rank, and a lower influence of page rank on CTR if the holding effect of rank dominates the catching effect of rank.

We can therefore summarize our empirical predictions as follows:

1) The effect of page rank on CTR should be systematically greater for some subset of page ranks than for others;

2) Controlling for intrinsic relevance should reduce the correlation between page rank and CTR;

3) A positive correlation between intrinsic relevance and the influence of page rank on CTR indicates that rank and brand effects are complements, and therefore that the catching effect of rank dominates the holding effect of rank. A negative correlation between intrinsic relevance and the influence of page rank on CTR indicates that rank and brand effects are substitutes and therefore that the holding effect of rank dominates the catching effect of rank.

4 Data

In order to conduct our study, we have marshaled a data set taken from the log files from Microsoft's Bing search engine for four query terms over a three-month time period. All search engines store data from user search sessions in detailed logs. The logs that we use contain recorded observations for each of the millions of Bing user queries, including for each query: a record of the date and time; all domains that were displayed on SERPs generated from the search; each domain's position on the SERPs; and which domains were clicked. This means that we know, for each domain that appeared in a set of search results, at what rank it appeared in each view and whether it was clicked on during that view.

Tractability requires that we select a limited number of query terms. Here, we have selected four query terms whose search results appeared likely to be reasonably free of extraneous influences. For example, a highly monetized or super-fresh query term would contain noise and would be less suitable for our study. This is because the relevance of results for such queries would fluctuate frequently for reasons that are difficult to control. Instead, we have chosen query terms that are (a) non-navigational (that is, not directed towards a particular domain), (b) of a relatively general topic, (c) not "super fresh", and (d) not heavily monetized, and (e) common enough to produce sufficient observations and variance for a meaningful econometric analysis. Using these criteria we chose the query terms "fun games", "phone numbers", "free movies" and "sports".[5]

We identify blended search results 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 on the same page, and count two or more clicks to the same domain on the same SERP

[5]Among the queries for which we collected data, only these four queries met the criteria a) through e); we did not arbitrarily select queries and report only those that produced significant results

as one click (one choice). We also omit SERPs where the same domain was observed twice in the top three ranks: after that, if the domain is displayed twice or more on the SERP, only the highest-ranking observation is kept. The data span the time period from 11/01/2010 to 1/31/2011.

The sample consists of those domains 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 domains, "fun games" for 158, "phone numbers" for 322, and "sports" for 996. However, not all domains had views in all ten positions. The tables in the Appendix show the top five domains (as determined by the total number of views for the time period analyzed) for each of the four queries. Table 1 shows the average CTR by page rank for domains in each of rank 1 to 10 for the four queries. It can be seen that domains that are more highly ranked have substantially greater CTRs on average. However, the effect of rank on CTR varies substantially by query: in particular, for the query "free movies", domains ranked in the last four ranks still have a CTR of over 1%, whereas for the other queries, the last four ranks have CTRs of less than 1%. The purpose of our econometric estimation is to examine the extent to which the strong association between high rank and CTR is a causal relationship as opposed to a mere correlation.

5 Econometric estimation

Our analysis begins from calculating click-through rates for each domain and for each hour during the time period studied. We then show how click-through rates depend systematically on the rank in which a domain appears. We present results for the "fun games" query first. Table 2 displays, in Regression 1, a simple OLS regression of hourly domain average click-through rate on the proportion of time that the domain spends in each page rank for the first three ranks. We focus on these three ranks for simplicity, though a similar account could be given for a larger number of ranks.. Each rank coefficient can be interpreted, for an average domain, as the number of additional percentage points of CTR associated with being in the rank in question, compared to being somewhere at random in the seven lower ranks.

Not surprisingly, the coefficients on the rank variables in Regression 1 are somewhat similar to the average click-through rates in Table 1. Regression 2 reports the effect of controlling for domain fixed effects and for the fixed effects of the hour of the day. The domain fixed effects are highly significant, and decrease the magnitude of the coefficients on all of the page ranks dummies, while still leaving these coefficients highly statistically significant, as indicated in the second column of each regression. In particular, the coefficient for rank 1 drops from 41.2% to 26.1%, which is a significant reduction but still indicated a very large positive effect. We can thus see that our first two empirical predictions are strongly confirmed by these data.

Tables 3 through 5 contain the regressions for the "phone numbers, "free movies", and "sports" queries, both excluding and including domain fixed effects, as in Table 2. The magnitude of the coefficients is lower than for "fun games" and varies substantially, with the coefficient on Rank 1 in Regression 2 varying from a low of 5.3% for the query term "sports" to a high of 18.1% for the query term "fun games". It is also worth noting that the inclusion of domain and hourly fixed effects changes the coefficients for the "free movies" query very little and that the coefficients on the second and third ranks are not always statistically significant, though the coefficient on the first rank is invariably so. Nevertheless, the coefficients on Rank 1 are always statistically significant and are always largest when we control for domain and hourly effects, thereby confirming our first two empirical predictions.

It might be expected that the estimates of the effect of page rank would still contain some effect of the intrinsic relevance of each domain. After all, controlling for domain fixed effects manages to control for relevance only in so far as this remains constant over the sample period. If there are factors that affect the relevance of domains to users' needs that are changing over time in a way that also affects the domains' page-ranking, then some of the apparent correlation between page ranking and CTR might be spurious, that is an artifact of the changing relevance.

How serious a concern this is will depend on the nature of the underlying process that determines page rank. It is only if page rank varies for reasons systematically related to relevance that our estimates will be biased; random fluctuations are to be welcomed since they constitute a natural experiment that helps us to isolate the effect of rank on traffic.

Two kinds of evidence suggest that systematic variation in page rank for reasons related to relevance is not likely to be an important problem here. The first consists of evidence of the way in which the SERP rank algorithm is constructed. User behavior data are the means by which systematic time-varying relevance factors might exert an influence on rank. However, though sources differ in their exact categorization of SERP rank determinants, there is a general consensus that past CTR constitutes less than 7% of all factors considered and that past CTR is used only intermittently in the rank algorithm [6] Our avoidance of so-called "super-fresh" query terms suggests that, for these four queries, past CTR will account for even less. However, there is a second source of evidence: we can also test directly for potential systematic influences of relevance on rank by looking at the dynamic structure of the observations in our data. Table 6 reports

[6] Sources: http://ilpubs.stanford.edu:8090/361/, http://www.vaughns-1-pagers.com/internet/google-ranking-factors.htm, and http://www.seomoz.org/article/search-ranking-factors

the results of an instrumental variables regression of the first difference of the CTR on the first difference of the rank, with the latter instrumented by the lag of the rank and the lagged first difference of the rank (this is the Anderson-Hsiao estimator [3]). In three of the four equations the rank variables easily pass the test of exogeneity. In the "phone numbers" query, the exogeneity hypothesis is rejected at the 5% level, but the estimated coefficient is much larger than in the OLS regressions, rather than smaller as we would expect if the cause of endogeneity were unobserved fluctuations in domain relevance. Altogether, the results of these regressions give us reasonable confidence in concluding that our OLS estimates of the impact of page rank on CTR are not upward biased by the failure to take into account any unobserved dependence of page rank on domain relevance [7].

One disadvantage of these OLS estimates is that they aggregate CTRs over an essentially arbitrary time span, which has the feature of weighting user behavior more heavily when it takes place at periods in which there are fewer clicks on a particular domain. It also raises the question of whether different time periods of aggregation might unduly influence the parameter estimates. An important robustness check on our estimates is therefore provided by the results, reported in Tables 7 though 10, of a multinomial logit estimation of the determinants of the probability that a given domain is clicked on when it is viewed. Here the unit of observation is not the views that take place for a given domain during a given hour, but rather a single query in which that domain appears on the results page; this is therefore independent of any influence of aggregation by arbitrary time period. The parameter estimates for each rank should be interpreted as the impact of appearing in that particular rank on the probability that the page will be clicked on, compared to the probability of clicking on Rank 5, and conditional on there being a click somewhere in the first five ranks.

[7] We also tested for systematic dynamics in the rank variable observations by regressing the first difference of rank on its own lagged value, instrumenting according to the Anderson-Hsiao procedure, and found no significant effects in any of the four queries, including "phone numbers".

In addition to reporting pure rank coefficients (which are the analogue of the rank coefficients in the OLS regressions), we also allow the choice probability to be influenced by one of the key characteristics of the rank, namely the "brand" or "quality" of the domain that is in that rank at the time that the page is viewed. The way we measure "brand" is as the mean page rank of the domain over the sample period as a whole (since "high" page rank involves low absolute numbers, this variable is therefore decreasing in the strength of the brand. Thus a positive coefficient on mean page rank would imply that rank effects are stronger for domains with weak brands (that is, branding and rank effects are substitutes), while a negative coefficient would imply that they are complements [8].

The results are very clear, and rather striking. First, the effects of being in Rank 1 and Rank 2 on click probabilities are always significant, and that of being in Rank 3 is significant in all queries except "sports". Except in "sports" where Rank 2 has a marginally higher coefficient than Rank 1, the rank coefficients are always decreasing by rank.

Secondly, the coefficient on Mean Rank is always negative, very often significant (13 times out of 16), and very often substantially larger in absolute magnitude for the top two ranks than for the third and fourth. The fact that this coefficient takes the same sign in all 16 cases is very strong evidence that branding effects are complementary to rank effects. The higher-ranked websites benefit more than lower-ranked websites do from being in a high rank. This implies, as we discussed above, that the catching effect of rank dominates the holding effect of rank. Equivalently, it implies that rank influences CTRs primarily by making domains more conspicuous on the page rather than by mak-

[8] We exclude all domain/rank combinations for which there were fewer than 20 observations

ing users more likely to click on them once their attention has already been attracted to the domain.

6 Conclusion

In this paper we have used an original dataset to investigate the impact of the rank attained by websites on a Search Engine Results Page on the rate at which users click through to the website in question. We have done so in a way that explicitly controls for the potentially confounding effect of a website's intrinsic relevance to users' needs, which may influence both its rank on the page and the willingness of users to click independently of that rank. We have found that controlling for intrinsic relevance does indeed make a difference to the estimated direct impact of page rank, substantially reducing the coefficients on high rank but still leaving them positive and highly significant both statistically and economically. Furthermore, our results offer an interesting insight into the mechanism by which page rank has its effect on click through rates. We have shown that high rank influences CTRs by making sites more conspicuous on the results page, and not by increasing the willingness of users to click on them once their attention has already been attracted. This therefore underlines the extent to which the scarce attention of users, and not merely their limited information about websites' characteristics, is a limiting resource that shapes the nature of competition on the internet. Exploring the characteristics of such scarce attention, and the way in which different technologies shape and develop the allocation of scarce user attention, is an exciting subject for future research.

Table 1: Average Click-Through Rate by Page Rank

	Average Click-Through Rate			
Rank	Fun Games	Phone Numbers	Free Movies	Sports
1	0.209	0.385	0.298	0.323
2	0.107	0.108	0.143	0.121
3	0.064	0.053	0.058	0.053
4	0.031	0.016	0.011	0.016
5	0.031	0.020	0.013	0.011
6	0.029	0.009	0.007	0.005
7	0.016	0.008	0.005	0.003
8	0.013	0.005	0.004	0.003
9	0.013	0.004	0.003	0.002
10	0.012	0.003	0.001	0.002
Number of User Serches in Sample	24,964	104,362	31,848	116,114

Table 2: Regressions of Hourly Domain Click-Through Rate on Page Rank for "Fun Games"

	Regression 1		Regression2	
Hourly Domain CTR	Coefficient	t-statistic	Coefficient	t-statistic
Rank 1	0.412 ***	94.82	0.261 ***	9.77
Rank 2	0.081 ***	33.78	0.037 ***	5.02
Rank 3	0.035 ***	17.73	0.015 ***	4.68
Constant	0.006 ***	20.47	0.024 ***	11.01
Domain FE	No		Yes	
Hour FE	No		Yes	
Fraction of Variance Due to Domain Fixed Effects			0.369	
Adjusted R^2	0.648		0.044	
Number of Observations	28,525		28,525	
Number of User				

Robust standard errors: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Regressions of Hourly Domain Click-Through Rate on Page Rank for "Phone Numbers"

	Regression 1		Regression2	
Hourly Domain CTR	Coefficient	t-statistic	Coefficient	t-statistic
Rank 1	0.278 ***	56.65	0.156 ***	8.67
Rank 2	0.147 ***	37.78	0.033	1.91
Rank 3	0.048 ***	18	0.009	0.63
Constant	0.004 ***	15.22	0.021 ***	7.29
Domain FE	No		Yes	
Hour FE	No		Yes	
Fraction of Variance Due to Domain Fixed Effects			0.213	
Adjusted R^2	0.429		0.06	
Number of Observations	33,146		33,146	
Number of User				

Robust standard errors: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regressions of Hourly Domain Click-Through Rate on Page Rank for "Free Movies"

	Regression 1		Regression2	
Hourly Domain CTR	Coefficient	t-statistic	Coefficient	t-statistic
Rank 1	0.195 ***	26.12	0.181 ***	13.33
Rank 2	0.088 ***	15.83	0.092***	7.51
Rank 3	0.045 ***	10.67	0.043***	3.97
Constant	0.015 ***	21.9	0.016 ***	7.79
Domain FE	No		Yes	
Hour FE	No		Yes	
Fraction of Variance Due to Domain Fixed Effects			0.159	
Adjusted R^2	0.153		0.047	
Number of Observations	19,011		19,011	
Number of User				
Robust standard errors: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$				

Table 5: Regressions of Hourly Domain Click-Through Rate on Page Rank for "Sports"

	Regression 1		Regression2	
Hourly Domain CTR	Coefficient	t-statistic	Coefficient	t-statistic
Rank 1	0.225 ***	92.56	0.053 ***	6.56
Rank 2	0.132 ***	54.52	0.005	0.78
Rank 3	0.040 ***	29.91	0.006 **	2.84
Constant	0.003 ***	23.94	0.012 ***	33.65
Domain FE	No		Yes	
Hour FE	No		Yes	
Fraction of Variance Due to Domain Fixed Effects			0.56	
Adjusted R^2	0.344		0.009	
Number of Observations	115,009		115,009	
Number of User				
Robust standard errors: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$				

Table 6: IV Regressions of First Difference of the Hourly Domain Click-Through Rate on First Difference of Rank for All Four Queries

Hourly Domain Rank CTR	Fun Games		Phone numbers		Free movies		Sports	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
Rank 1	0.195*	2.15	0.656***	3.22	0.129 ***	1.88	0.217**	2.77
Rank 2	0.074 *	1.97	0.242	1.23	-0.005	-0.09	0.040	0.85
Rank 3	0.001	0.04	0.403	1.94	0.022	0.73	-0.003	-0.29
Constant	0.00	0.08	-0.000	-0.21	-0.001	-0.04	0.000	0.85
P-value for Exogeniety F-test (Robust standard errors) for	0.63		0.03		0.48		0.78	
Adjusted R^2	0.02		0.005		0.01		0.004	
Number of Observations	22,757		18,397		11,060		44,031	

Instruments: lagged rank, lagged first difference of rank. All variables defined as differences with respect to their hourly mean
Robust standard errors: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Multinomial Logistic Regression of Click on Rank and Mean rank for "Fun Games"

Effect compared to Rank 5	Coefficient	Robust Standard Error	Z	$P > z$	95% Confidence Interval	
Mean Rank of Domain in Rank 1	-6.745	2.003	-3.370	0.001	-10.670	-2.820
Effect on Rank 1	18.992	4.808	3.950	0.000	9.570	28.415
Mean Rank of Domain in Rank 2	-2.747	1.260	-2.180	0.029	-5.216	-0.277
Effect on Rank 2	11.326	4.213	2.690	0.007	3.069	19.583
Mean Rank of 3 Domain in Rank 3	-1.511	0.589	-2.570	0.010	-2.665	-0.358
Effect on Rank 3	7.477	2.606	2.870	0.004	2.370	12.584
Mean Rank of Domain in Rank 4	-0.376	0.235	-1.600	0.109	-0.836	0.083
Effect on Rank 4	2.108	1.339	1.570	0.115	-0.516	4.732
Number of Observations	60,986					
Pseudo R^2	0.5326					

Table 8: Multinomial Logistic Regression of Click on Rank and Mean rank for "Phone Numbrs"

Effect compared to Rank 5	Coefficient	Robust Standard Error	Z	$P > z$	95% Confidence Interval	
Mean Rank of Domain in Rank 1	-12.453	2.443	-5.100	0.000	-17.664	-7.664
Effect on Rank 1	43.583	8.405	5.190	0.000	27.110	60.057
Mean Rank of Domain in Rank 2	-11.102	2.322	-4.780	0.000	-15.653	-6.551
Effect on Rank 2	40.870	8.238	4.960	0.000	24.723	57.016
Mean Rank of 3 Domain in Rank 3	-7.387	2.314	-3.190	0.001	-11.923	-2.851
Effect on Rank 3	32.028	8.318	3.850	0.000	15.726	48.330
Mean Rank of Domain in Rank 4	-1.759	0.857	-2.050	0.040	-0.836	0.080
Effect on Rank 4	11.017	5.135	2.150	0.032	0.953	21.082
Number of Observations	16,986					
Pseudo R^2	0.3535					

Table 9: Multinomial Logistic Regression of Click on Rank and Mean rank for "Free Movies"

Effect compared to Rank 5	Coefficient	Robust Standard Error	Z	$P > z$	95% Confidence Interval	
Mean Rank of Domain in Rank 1	-2.417	0.985	-2.450	0.014	-4.348	-0.485
Effect on Rank 1	11.601	2.557	4.540	0.000	6.589	16.612
Mean Rank of Domain in Rank 2	-1.234	0.489	-2.520	0.012	-2.193	-0.274
Effect on Rank 2	8.216	2.058	3.990	0.000	4.183	12.250
Mean Rank of 3 Domain in Rank 3	-1.328	0.464	-2.860	0.004	-2.237	-0.419
Effect on Rank 3	8.054	2.054	3.920	0.000	4.028	12.080
Mean Rank of Domain in Rank 4	-0.693	0.282	-2.460	0.014	-1.246	-0.140
Effect on Rank 4	4.471	1.901	2.350	0.019	0.745	8.197
Number of Observations	10,211					
Pseudo R^2	0.1859					

Table 10: Multinomial Logistic Regression of Click on Rank and Mean rank for "Sports"

Effect compared to Rank 5	Coefficient	Robust Standard Error	Z	$P > z$	95% Confidence Interval	
Mean Rank of Domain in Rank 1	-7.985	3.627	-2.200	0.028	-15.094	-0.876
Effect on Rank 1	26.790	12.236	2.190	0.029	2.808	50.773
Mean Rank of Domain in Rank 2	-8.378	3.513	-2.380	0.017	-15.264	-1.492
Effect on Rank 2	26.929	12.022	2.240	0.025	3.367	50.492
Mean Rank of 3 Domain in Rank 3	-4.012	3.131	-1.280	0.200	-10.149	2.124
Effect on Rank 3	16.797	11.200	1.500	0.134	-5.156	38.749
Mean Rank of Domain in Rank 4	-0.861	0.570	-1.510	0.131	-1.979	0.256
Effect on Rank 4	4.431	3.754	1.180	0.238	-2.925	11.788
Number of Observations	61,21,					
Pseudo R^2	0.4102					

Table 11: Top Five Domains for "Fun Games"

Domain	Position	Views	Clicks	CTR
bumarcade.com	1	1,401	352	0.251
	2	16,438	1,192	0.073
	3	80,228	3,062	0.038
	4	963	27	0.028
	5	2,410	40	0.017
	6	1,884	23	0.012
	7	913	7	0.008
	8	25	0	0.000
	9	9	0	0.000
	10	5	0	0.000
addictinggames.com	1	2,418	874	0.361
	2	86,142	8,186	0.095
	3	14,400	1,032	0.072
	4	936	34	0.036
	5	110	6	0.055
	6	43	1	0.023
	7	21	0	0
	8	9	0	0
	9	5	0	0
	10	1	0	0
funny-games.biz	1	201	97	0.483
	2	85	10	0.118
	3	7,909	293	0.037
	4	27,979	713	0.025
	5	33,570	582	0.017
	6	25,842	327	0.013
	7	8,011	87	0.011
	8	258	2	0.008
	9	51	1	0.02
	10	59	0	0
mostfungames.com	1	99,894	43,237	0.433
	2	1,198	237	0.198
	3	669	105	0.157
	4	58	8	0.138
	5	308	35	0.114
	6	190	17	0.089
	7	3	0	0
	8	19	0	0
	9	1	0	0
	10	-	-	-
bored.com	1	83	3	0.036
	2	1	0	0
	3	25	1	0.04
	4	83	0	0
	5	10,572	134	0.013
	6	8,424	102	0.012
	7	19,471	109	0.006
	8	27,549	173	0.006
	9	22,809	129	0.006
	10	11,272	54	0.005

Table 12: Top Five Domains for "Phone Numbers"

Domain	Position	Views	Clicks	CTR
phonenumber.com	1	17,075	5,045	0.295
	2	13,315	2,241	0.168
	3	1,417	142	0.1
	4	9	0	0
	5	6	0	0
	6	5	0	0
	7	1	0	0
	8	-	-	-
	9	3	1	0.333
	10	1	0	0
whitepages.com	1	14,652	4,010	0.274
	2	16,558	2,549	0.154
	3	580	56	0.097
	4	1	0	0
	5	10	1	0.1
	6	5	0	0
	7	-	-	-
	8	-	-	-
	9	7	0	0
	10	13	0	0
en.wikipedia.org	1	1	0	0
	2	12	0	0
	3	8	0	0
	4	229	0	0
	5	4,055	6	0.001
	6	21,648	31	0.001
	7	4,142	5	0.001
	8	1,288	4	0.003
	9	349	0	0
	10	74	0	0
switchboard.com	1	80	50	0.625
	2	1,893	186	0.098
	3	29,734	1,600	0.054
	4	36	4	0.111
	5	19	3	0.158
	6	22	2	0.091
	7	6	0	0
	8	4	0	0
	9	1	0	0
	10	3	0	0
anywho.com	1	5	4	0.8
	2	-	-	-
	3	15	1	0.067
	4	8,645	244	0.028
	5	6,185	129	0.021
	6	1,933	28	0.014
	7	9,653	146	0.015
	8	3,650	45	0.012
	9	1,428	13	0.009
	10	201	0	0

Table 13: Top Five Domains for “Free Movies” Query

Domain	Position	Views	Clicks	CTR
hulu.com	1	440	57	0.13
	2	13,031	1,335	0.102
	3	10,364	808	0.078
	4	107	8	0.075
	5	98	2	0.02
	6	117	8	0.068
	7	78	3	0.038
	8	37	1	0.027
	9	15	1	0.067
	10	1	0	0
fancast.com	1	20,613	4,396	0.213
	2	2,866	333	0.116
	3	158	12	0.076
	4	45	1	0.022
	5	18	0	0
	6	32	4	0.125
	7	56	1	0.018
	8	134	3	0.022
	9	87	2	0.023
	10	94	0	0
free-new-movies.com	1	-	-	-
	2	374	47	0.126
	3	41	3	0.073
	4	68	3	0.044
	5	57	3	0.053
	6	166	5	0.03
	7	2,319	64	0.028
	8	8,738	208	0.024
	9	5,638	121	0.021
	10	5,470	130	0.024
freemoviescinema.com	1	3,231	701	0.217
	2	7,879	873	0.111
	3	321	33	0.103
	4	40	7	0.175
	5	325	15	0.046
	6	4,866	184	0.038
	7	4,385	145	0.033
	8	1,215	26	0.021
	9	214	7	0.033
	10	216	5	0.023
ovguide.com	1	5	3	0.6
	2	73	11	0.151
	3	53	1	0.019
	4	3	0	0
	5	229	9	0.039
	6	3,594	137	0.038
	7	9,259	279	0.03
	8	3,329	87	0.026
	9	858	27	0.031
	10	66	0	0

Table 14: Top Five Domains for “Sports” Query

Domain	Position	Views	Clicks	CTR
sports.com	1	21	6	0.286
	2	343	0	0
	3	2,557	17	0.007
	4	1,717	13	0.008
	5	41,364	238	0.006
	6	38,514	138	0.004
	7	27,411	132	0.005
	8	3,549	6	0.002
	9	450	4	0.009
	10	40	0	0
espn.go.com	1	48,027	9,919	0.207
	2	65,521	8,648	0.132
	3	1,873	187	0.1
	4	15	1	0.067
	5	304	20	0.066
	6	95	8	0.084
	7	10	1	0.1
	8	7	0	0
	9	19	1	0.053
	10	18	0	0
sports.yahoo.com	1	66,744	18,228	0.273
	2	48,290	9,390	0.194
	3	488	37	0.076
	4	4	0	0
	5	40	2	0.05
	6	8	0	0
	7	6	0	0
	8	8	1	0.125
	9	13	3	0.231
	10	7	1	0.143
msn.foxsports.com	1	875	627	0.717
	2	664	86	0.13
	3	104,481	10,538	0.101
	4	284	14	0.049
	5	6,816	333	0.049
	6	421	32	0.076
	7	133	4	0.03
	8	7	0	0
	9	50	1	0.02
	10	77	4	0.052
sportsillustrated.cnn.com	1	166	110	0.663
	2	1	1	1
	3	131	4	0.031
	4	261	8	0.031
	5	38,243	872	0.023
	6	14,065	269	0.019
	7	7,527	107	0.014
	8	14,548	199	0.014
	9	19,488	306	0.016
	10	9,470	152	0.016

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] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.