What's in a name? Measuring prominence and its impact on organic traffic from search engines☆

Michael R. Baye [a], Babur De los Santos [b,*], Matthijs R. Wildenbeest [a]

[a] *Indiana University, United States*
[b] *Clemson University, Department of Economics, 238 Sirrine Hall, Clemson, SC 29634, USA*

A B S T R A C T

Organic product search results on Google and Bing do not systematically include information about seller characteristics (e.g., feedback ratings and prices). Consequently, it is often assumed that a retailer's organic traffic is driven by the prominence of its *position* in the list of search results. We propose a novel measure of the prominence of a retailer's *name*, and show that it is also an important predictor of the organic traffic retailers enjoy from product searches through Google and Bing. We also show that failure to account for the prominence of retailers' names—as well as the endogeneity of retailers' positions in the list of search results—significantly inflates the estimated impact of screen position on organic clicks.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Recent theoretical work by Arbatskaya (2007), Armstrong et al. (2009), Armstrong and Zhou (2011) emphasizes the role of *prominence* in consumer search models. The key to these models is that consumers visit more prominent firms first. In the context of online product search, this implies that more prominent retailers receive more clicks than their less prominent rivals. Prominence in online markets is sometimes interpreted as a firm's screen position in listings of search results—and for good reason: there is abundant evidence that position is an important determinant of the clicks retailers receive from searches at price comparison sites (e.g., Shopper.com and Nextag.com), marketplaces (e.g., Amazon), and auction sites (e.g., eBay).[1] Of course, position is not the only information included in search results on these platforms, and it is well documented that this other information about retailer attributes—reputational ratings, third-party certifications, prices, shipping charges, and so on—is an

---

[1] See Ansari and Mela (2003), Ellison and Ellison (2009), Baye et al. (2009), Brynjolfsson et al. (2010), and the studies cited therein. Additionally, Armstrong et al. (2009), De los Santos and Koulayev (2013), and Novarese and Wilson (2012) summarize a number of studies in both online and offline environments (including the yellow pages, voting, and academic citations) that find that position significantly impacts choice.

even more important determinant of the number of clicks retailers receive through these platforms.[2] Other things equal, firms with more prominent positions do tend to get more clicks—but other things equal, so do firms with more prominent reputations and other characteristics relevant to consumers.

In contrast to the rich information returned for searches on these platforms, product searches on Google and Bing return a list of organic results deemed "relevant" by the search engine's algorithm. Essentially, this is a list of retailer names. Consequently, one can readily control for a retailer's position in the list of results, but not its attributes. Effectively, non-position attributes that influence clicks—reputational ratings, third-party certifications, prices, shipping charges, and so on—are embodied in the retailers' names.

To be concrete, suppose consumers query Google or Bing with the phrase "buy product X online" and that an obscure website, say FlyByNight.com, is displayed in a higher position than Amazon in the organic results. On the one hand, FlyByNight has the more prominent position, and based solely on this one might predict that consumers will choose to click its link. This prediction, of course, ignores the fact that Amazon has the more prominent name (or equivalently in this context, is more prominent in terms of its name recognition, brand awareness, or brand equity). In general, one would expect both the retailer's position and name to influence clicks; a given retailer may be less prominent in one of these dimensions but more prominent in the other.

While little is known about how these factors impact the organic clicks retailers receive as a result of product searches on general search engines, recent evidence indicates that *sponsored* links in more prominent positions tend to have higher click-through rates.[3] But there is also evidence of what Jerath et al. (2011) call the *position paradox*: sponsored links appearing in *less* prominent positions sometimes receive *more* clicks than those in more prominent positions. The position paradox is consistent with our view that factors other than position are embodied in a retailer's name and also influence the number of organic clicks retailers enjoy following product searches. We wish to test this hypothesis.

The main challenge, and the primary goal of this paper, is to objectively measure the prominence of retailers' names. This is a long-standing problem in economics and marketing, and a variety of different methodologies have been used. One approach, employed by Sappington and Wernerfelt (1985) in their study of traditional liquor markets, uses historical advertising data to proxy the strength of firms' brand names. A related approach, pioneered by Goldfarb et al. (2009) in their study of ready-to-eat break-

fast cereal, uses time series advertising data in conjunction with a structural model to obtain estimates of the value of brand names. Unfortunately, these methodologies are not ideal for measuring the prominence of online retailers' names. Many online retailers are not publicly traded and do not disclose advertising expenditures, and the parent companies of publicly traded retailers do not systematically report disaggregated advertising expenditures at the URL level. For these reasons, Drèze and Zufryden (2004) use a third approach that relies on results from a consumer survey to construct an index of the "visibility" of 100 online firms. As they note, this approach is expensive, reflects visibility at a single point in time, and involves the usual caveats regarding the use of survey data.

In Section 2 we introduce a novel measure of the prominence of a firm's name. This measure, which we construct using comScore Search Planner data, is based on the number of product searches at Google (or Bing) that includes the retailer's name or URL in the search query. Intuitively, the inclusion of "amazon" as one of the terms in a product search means, at a minimum, that the searcher recognizes and can recall Amazon's name. This may be due to current or past advertising campaigns by Amazon, recommendations from friends, knowledge of Amazon's product breadth and depth, its speedy shipping practices, and so on. This and other potentially relevant information is embodied in Amazon's name; retailers with more of these "name searches" at a given point in time are deemed to have more prominent names at that point in time than retailers with fewer name searches. A revealed preference argument indicates that this measure captures more than mere recognition or recall. The inclusion of a retailer's name in a product search indicates that the retailer's reputation and attributes (in dimensions mattering to consumers, such as product breadth or service quality) are sufficiently strong that searchers find it optimal to reveal—through search queries—that results for that particular retailer are most welcomed. We also provide evidence that this measure works as advertised in an education context: More prominent universities (based on 2012 *U.S. News and World Report* rankings) enjoy more name searches than less prominent universities.

In Section 3 we provide an empirical application of the use of our measure of name prominence. The application measures the effect of name prominence on traffic that online retailers receive from organic product searches. Our data includes approximately 12,000 product search terms and phrases that led consumers to 757 different retail sites in August, 2012. We employ data from a single month, and aggregated up to the retailer level, to emphasize that our approach does not require historical time-series or consumer-level micro data. We present preliminary regression results suggesting that the prominence of retailers' names and positions are both important determinants of the links consumers click following product searches. These results also reveal that failure to account for the prominence of retailers' names substantially inflates the estimated effects of position. This suggests that companies like FlyByNight that spend thousands of dollars on search engine optimization—in an attempt to "game" their positions in search results—may be overestimating the benefits

---

[2] See, for instance, Smith and Brynjolfsson (2001), Melnik and Alm (2002), Houser and Wooders (2006), Jin and Kato (2006), Dewally and Ederington (2006), Hossain and Morgan (2006), Baye and Morgan (2009), and Baye et al. (2009).

[3] See Ghose and Yang (2009), Rutz and Bucklin (2011), and Yao and Mela (2011). Yao and Mela also find higher click-through rates for advertised *products* with greater objective brand quality (as defined by external consumer/expert ratings). As discussed below, our results are complementary to, and consistent with, these studies.

of such strategies.[4] We also show that these preliminary results are robust to the inclusion of a variety of other controls, and thus are unlikely to be an artifact of spurious correlation.

Section 3 also tackles endogeneity issues inherent in this line of research. Search engines have strong incentives to provide searchers with relevant results. Consequently, they base firms' positions on results pages, in part, on past clicks. In the above hypothetical, as more and more searchers click Amazon, an optimizing search engine has an incentive to demote FlyByNight's position and elevate Amazon toward the top of the list. Our analysis suggests that when one controls for endogeneity, the prominence of a retailer's name is a more important determinant of clicks than the prominence of its position. We find that a retailer moving from the median to the best positions gets about 80 percent more clicks, whereas moving from the median to the best levels of name prominence increases clicks by about 154 percent. In short, when one controls for endogeneity, name and position prominence both remain economically and statistically important determinants of organic clicks, but name prominence appears to have the larger impact. We also show that our main findings—that name prominence is an important determinant of clicks, and that failure to account for name prominence results in inflated estimates of position effects—also holds for Bing data. Finally, we show that these findings are also robust to alternative measures of name as well as position prominence.

Section 4 concludes with a discussion of the potential relevance of our analysis for related research, and provides some caveats and directions for future research.

2. Measuring name prominence

Consider an online retailer interested in attracting traffic to its website. It might invest in advertising through traditional (TV, radio or print) or online channels in an attempt to enhance consumer awareness and generate visits to its website. It might spend large sums to build a customer-centric website with a broad array of product offerings and an efficient network of distribution centers to create customer loyalty and word-of-mouth (or word-of-blog) advertising. Or it might use sponsored search, some other strategy, or a blend of several strategies to induce consumers to visit its website. A less costly option is to eschew such investments altogether and simply "free ride" on any traffic obtained through organic search results. The levels of these and other investments by online retailers impact the prominence of their names.

Unfortunately, measuring such investments is a challenge in online markets. As discussed in the introduction,

many online retailers are privately held and do not disclose this information; publicly traded companies do not systematically provide detailed information about the many investments they make to enhance the prominence of their online arms. Our proposed measure of the prominence of an online retailer's name in a given period is the number of name searches it obtained during that period. Here, *name searches* refers to search terms and phrases such as "cameras amazon" and "amazon" that include the retailers name or URL.

For a variety of reasons, the number of name searches is a potentially useful measure of the prominence of a retailer's name. First, it is measurable. For example, one can use comScore Search Planner or Google Trends data to calculate the number of name searches different retailers received in a given month. Second, the number of name searches captures the aggregate behavior of individual searchers who are acting on *all* of the many investments retailers made up to that point in time. Essentially, a firm's number of name searches embodies the cumulative branding efforts of the firm up to and including the instant a search is made. Third, the number of name searches in a given month takes into account the *stock* of name prominence or the accumulated brand equity of retailers. In contrast, even if data were available on the investments different retailers made on advertising and other brand-enhancing activities in a given month, such expenditures merely represent *flows* that incrementally change name prominence relative to previous periods, and therefore would not be helpful in conducting a cross-sectional analysis of the impact of name prominence on firms' organic clicks from product searches. Even with time-series data on advertising and other brand-enhancing expenditures, one would have to deal with the thorny issue of identifying the stock of brand equity from such flow data. Finally, a revealed preference argument indicates that the total number of name searches conveniently embodies branding, reputation and other attributes that shoppers associate with a retailer's name.

More formally, let A denote a set of alternative search terms, let $\succsim$ denote an individual's binary preference ordering, and $S \equiv S(A, \succsim)$ denote the search phrase actually used by the individual (e.g., his or her choice of search terms). By the weak axiom of revealed preference, $S' \in A$ implies $S \succeq S'$. To illustrate, note that $S_1 = \{shop\ online\}$, $S_2 = \{amazon\}$, and $S_3 = \{flybynight\}$ are among a plethora of alternative search terms (e.g., A) that an online shopper might use to identify a product, a seller, or both. Since each $S_i \in A$, the weak axiom of revealed preference indicates (among other things) that $S_2 \in S(A, \succsim)$ implies $S_2 \succeq S_i$ for all $i \neq 2$. For this reason, one may reasonably infer that an individual using the search phrase $S_2 = \{amazon\}$ prefers search results containing links to "amazon" over results with links to alternative retailers. Presumably, the name "amazon" embodies retailer attributes that this particular shopper prefers to the attributes he or she associates with the names of alternative retailers.

Thus, the total number of times a retailer's name is used in online searches is an intuitively appealing measure of name prominence that is strongly rooted in revealed preference theory. Note that, unlike some applications of

---

[4] Baye et al. (2014) provide an application of our measure of name prominence to search engine optimization (SEO) and find that brand equity has beneficial direct and indirect effects on organic clicks and conclude that investments in brand equity should be part of a retailer's SEO strategy. Their analysis uses organic clicks at the search term-retailer level, whereas in this paper we aggregate clicks to the retailer level. The role of the application in this paper is to emphasize that prominence affects consumer behavior at search engines, which can be accomplished at a lower level of granularity.
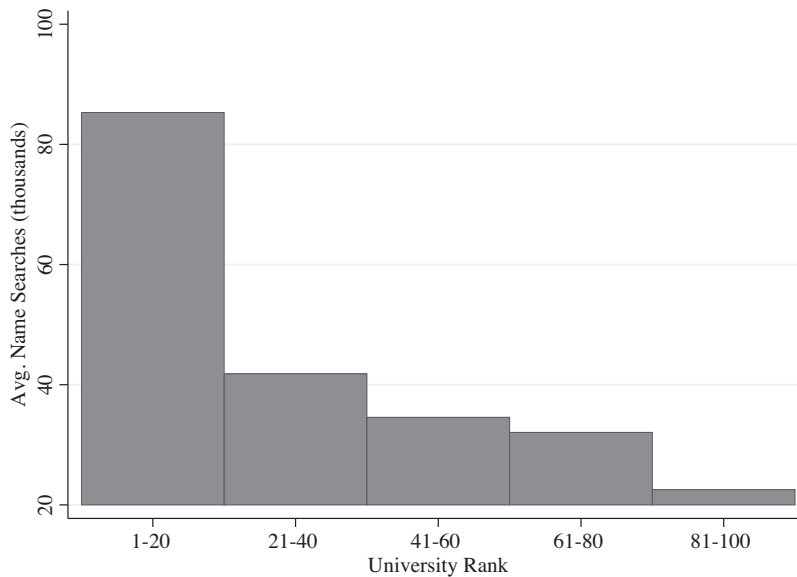
Fig. 1. Name searches and university rankings.

revealed preference, this measure of name prominence does not assume that all searchers have identical preferences. Indeed, any heterogeneities in preferences will be manifested (and thus accounted for) in differing counts for searches with different names.

In order to examine the potential promise of our proposed measure of name prominence, we obtained *U.S. News and World Report* rankings of the top 100 universities in 2012 and placed each university in one of five quintiles. Thus, the most prominent universities (which include the likes of Harvard, Stanford, Princeton, and the other usual suspects) were in the top quintile; Indiana University (a large public university) was in the 4th quintile. We then used comScore Search Planner data (employing a methodology analogous to that described in the next section) to determine the total number of name searches universities in each quintile received during February 2012. As shown in Fig. 1, more prominent universities (as measured by the *U.S. News and World Report Rankings*) received more name searches than less prominent universities. For example, universities in the top 20 averaged over 80,000 name searches, while those in the bottom 20 averaged around 22,000 name searches.

Importantly, this relationship holds even though the underlying data includes name searches by those looking for scientific studies (e.g., "harvard fluoride study") as well as name searches by students merely wishing to login to university email accounts (e.g., "purdue email login"). Name searches are a good predictor of university prominence despite the fact that less prominent universities tend to have more students conducting name searches to merely login to various university accounts.

Table 1 presents results from simple regressions of the logarithm of university rank (with 1 representing the best and 100 the worst) on the logarithm of the number of name searches (*Name*) and various controls. The negative and statistically significant coefficient for ln(*Name*) in

Table 1
University rank and name prominence.

Dependent variable: logarithm of university rank				
Variable	(1)	(2)	(3)	(4)
Ln(Name)	−0.418	−0.562	−0.436	−0.543
	(0.078)*	(0.069)*	(0.068)*	(0.069)*
Ln(Enrollment)		0.666		0.556
		(0.108)*		(0.146)*
Public university			0.853	0.228
			(0.148)*	(0.176)
Constant	7.923	3.178	7.734	3.910
	(0.777)*	(0.928)*	(0.683)*	(1.177)*
Observations	94	94	94	94
R^2	0.17	0.46	0.37	0.46

Notes: Standard errors in parentheses; *denotes significant at 5%.

the baseline specification in column 1 is consistent with the relationship displayed graphically in Fig. 1: More prominent universities receive more name searches than less prominent universities. The specification in column 2 shows that this finding is robust to controlling for the size of the university (the logarithm of enrollments). Indeed, this specification results in an even stronger, negative relation between the number of name searches and university rank. This conclusion is also robust to the inclusion of controls for whether the university is public or private. In particular, the specification in column 3 includes a dummy variable that equals one if the university is a public institution (45 percent of the sample), while the specification in column 4 controls for both enrollments and whether the school is public or private. In all specifications, there is a statistically significant and negative relationship between the number of name searches and the rank of a university. In summary, the regressions in Table 1 demonstrate that the graphical results displayed in Fig. 1 are hardly the result of spurious correlation between university rankings
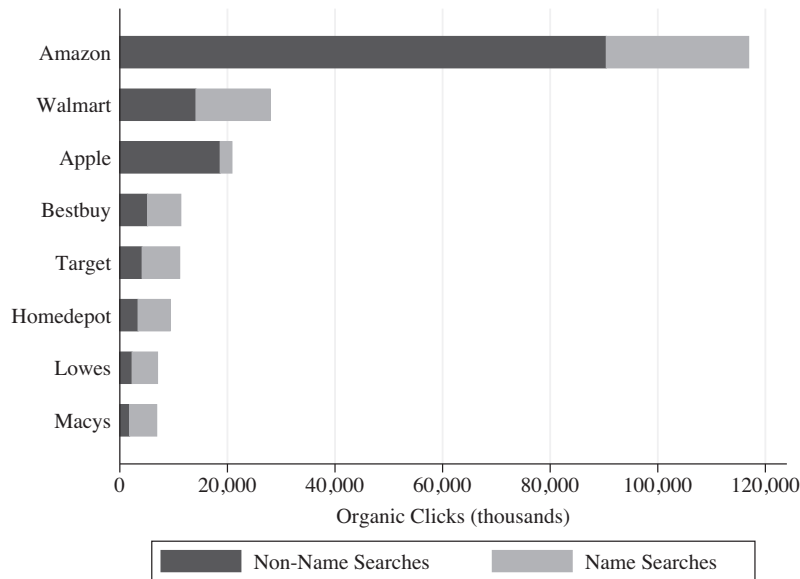
Fig. 2. Organic traffic from name searches and non-name searches.

and the size of enrollment or whether a school is public or private.

To further illustrate the importance of name prominence, we used comScore Search Planner data to obtain some of the top search terms and phrases on Google that led searchers to click on organic links directing them to Amazon.com, along with the number organic clicks associated with each term. Table 2 illustrates a striking feature: A very large proportion of organic traffic from search engines to Amazon.com stems from name searches—terms and phrases such as "amazon," "amazon.com," "amazon books," and "amazon kindle" —that searchers use as a substitute for directly navigating to the Amazon.com website.[5] In contrast, other searches (like "Panasonic TV" or "buy jeans") result in a much smaller amount of organic traffic from Google to Amazon. Fig. 2 shows that several other retail sites also receive a substantial amount of traffic from organic name searches, and that some sites receive a greater proportion of such traffic than Amazon.

For these reasons, we believe that the number of name searches is a promising way to control for the prominence of retailers' names. Next, we provide an empirical illustration of the importance of the measure of name prominence on traffic that online retailers receive from organic product searches.

3. Empirical application

3.1. Data

Our analysis is based on three datasets. We assembled two of these using data from third-party providers that specialize in electronic commerce marketing data (com-

Table 2
Selected search terms that led users from Google to Amazon in August, 2012.

Rank	Search phrase	# Organic clicks
1	amazon	10,282,361
2	amazon.com	1,447,802
3	www.amazon.com	267,650
4	amazon books	137,463
5	amazon prime	84,103
6	amazon seller central	75,942
7	amazon seller	73,208
8	amzon	60,396
9	amazon india	55,035
10	com	50,417
11	google	50,172
12	amazon .com	50,006
13	amazon publishing	47,558
14	amaozn	45,638
15	amazon kindle	44,994
16	amazon customer service	42,579
17	kindle	41,834
18	gone girl	40,566
19	aazon	38,905
20	ebay	38,687
30	cake pops	30,631
40	nendoroid	25,472
50	ncaa football 13	21,001
60	portable dvd player	17,096
70	fire extinguisher	15,154
80	bluetooth headset	12,515
90	amazon customer service number	11,283
100	skin tight	9,811
	All Search Phrases	82,963,096

Notes: comScore Search Planner data from August 2012. Search phrases are ranked by the total number of organic clicks on Google. Excludes search phrases that comScore does not disclose for privacy reasons.

Score and Internet Retailer) and created the third dataset using a web scraper written in Java.

The comScore dataset consists of monthly Search Planner data for August 2012. These data are based on the online browsing activity of two million users in the U.S.

---

[5] In industry parlance, name searches are sometimes called "navigational searches." We use "name search" to emphasize that these searches contain the name or URL of a particular retailer or site.

It provides a list of search terms and phrases that users entered at search engines (e.g., Google and Bing), along with the number of organic clicks that different websites received based on the results pages generated by each search term.

The Internet Retailer dataset provides a list of the top 500 online retailers, along with the retail segment in which each retailer operates (e.g., apparel and accessories, housewares and home furnishings, computers and electronics, and so on). The data indicate whether the retailers have a presence on Facebook or Twitter, the year in which the retail site began its online operations, and whether the firm is a web-only retailer (as is the case with Amazon) or also has a brick-and-mortar presence (as is the case with Walmart).

Since our goal is to examine *product* search on general search engines, the first step in our analysis was to link the 500 retailers in the Internet Retailer data with the comScore Search Planner data. In particular, we examined the Search Planner data and identified all of the properties owned by these 500 retailers that were tracked by comScore. Owing to the fact that some retailers own and operate websites with different domain names, (e.g., Amazon operates both the Amazon and Zappos sites; Sears operates the Sears site as well as a Kenmore and Kmart site), we ended up with a sample of 757 retail sites.[6]

Next, we extracted the comScore Site Profile for each of these 757 retail websites. Each site profile provides a list of the search terms and phrases (for Google and Bing separately) that resulted in organic clicks from results pages for that search term to a particular retail site. For each search term or phrase, it also indicates the total number of organic clicks each retail site received from results pages on each search engine. For example, across the 757 retail sites, comScore identified a total of 11,836 search terms and phrases that led consumers from Google to one or more of these 757 sites.[7]

As we showed in the previous section, a very large proportion of the organic traffic to retailers stems from name searches. It is hardly surprising that searches including "amazon" in a Google search phrase result in clicks on Amazon.com links on results pages. We are therefore interested in explaining clicks stemming from searches that are *not* name searches. For purposes of our analysis, a *name search* is defined as a search that includes the retailer/site name and misspellings (e.g., "amazon," "www.amazon.com," and "amazon.com") as well as *phrases* containing such terms (e.g., "buy camera at amazon.com" or "buy tv at amzon").[8] An examination of the 11,836

search terms and phrases revealed that 7,518 were *not* name searches. For each of the 757 websites, we computed the number of organic clicks received from these 7,518 searches that were not name searches. Thus, when we refer to clicks, we are referring to non-name organic clicks—a retailer's total number of organic clicks *minus* the organic clicks it received from searches that included its name (name searches).

The third dataset was obtained by capturing search results. We wrote a Java program that queried Google and Bing in September 2012 to capture the positions of the retailers in our sample on the first five pages of results for each of these 7,518 terms and phrases. As discussed in more detail below, this permits us to control for the positions of different retailers on results pages for different queries, as well as to construct controls for ads on results pages that may influence searchers' decisions to click on organic links.

Our analysis is based on variables constructed from the datasets described above; basic descriptive statistics are reported in Table 3.

Clicks. Our dependent variable is $\ln(Clicks_i)$. As discussed above, $Clicks_i$ represents the number of *organic* clicks retailer i received from searches that did *not* include retailer i's name in the search. As shown in Table 3, retail sites in our sample received an average of 272,000 of these clicks through Google. There is substantial cross-sectional variation, however.

Position. Based on the data obtained by querying the Google and Bing search engines using the 7,518 non-name search terms and phrases, we calculated the average screen position for each of the 757 retailers in our sample.[9] Based on these average positions, we categorized the position variables as follows. Sites that never appeared on the first five pages for this sample of search terms were placed in a position category labeled "worst." While we do not know these sites' actual average screen positions, we know these sites had the worst positions of any sites in our sample. Remaining retailers were assigned to position categories associated with the quintile in which their average screen position fell, ranked from "poor" for those retailers with an average screen position that was in the lowest quintile of those in which at least one screen position was observed, to "best" for the retailers with an average screen position in the highest quintile.[10] Thus, our primary measure of screen position is a dummy variable that equals one when a site's average screen position is one of these categories (and is zero otherwise).

---

[6] Data on the age, web-only status, and retail segments of these additional URLs, as well as data missing in the Internet Retailer database, were collected by hand using information from company websites, the WHOIS database, and the Internet Archive. All of these URLs inherit the social network status of the parent company, as reported in the Internet Retailer database.

[7] We first cleaned the data to eliminate entries such as "∗∗∗" that comScore inserted for search terms that raised potential privacy concerns.

[8] In the Appendix we show that our results are robust to a more narrow definition of a name search that includes site names (amazon.com) and misspellings (amazn.com) but excludes *phrases* with such terms ("buy camera at amazon.com").

[9] Observed positions ranged from 1 to 51. Since we only obtained data for the first five pages of search results, positions outside of this range are not observed and assigned a value of 52.

[10] The intermediate categories are "below median" (second lowest quintile), "median" (middle quintile), and "above median" (next to highest quintile). A retail site in the "below median" category has an average screen position that was in the second lowest quintile of those in which at least one screen position was observed, while a retail site in the "median" category has an average screen position that was in the middle quintile of those in which at least one screen position was observed. The "above median" category contains all those retail sites that have an average search results position that was in the next to highest quintile.

Table 3
Descriptive statistics ($N = 757$).

Variable	Mean	Std. Dev.
Clicks on Google (thousands)	271.96	2648.99
First Page	24.43	141.65
Ads	18.32	87.81
Position on Google		
Worst	0.20	0.40
Poor	0.16	0.37
Below median	0.16	0.37
Median	0.16	0.37
Above median	0.16	0.37
Best	0.16	0.37
Name on Google		
Worst	0.32	0.47
Poor	0.14	0.34
Below median	0.14	0.34
Median	0.14	0.34
Above median	0.14	0.34
Best	0.13	0.34
Name on Bing		
Worst	0.56	0.50
Poor	0.09	0.28
Below median	0.09	0.28
Median	0.09	0.28
Above median	0.09	0.28
Best	0.09	0.28
Social network presence	0.89	0.31
Site age	13.04	3.41
Web only retailer	0.35	0.48
Retail Segment		
Apparel/accessories	0.28	0.45
Automotive parts/accessories	0.01	0.09
Books/music/video	0.03	0.18
Computers/electronics	0.08	0.27
Flowers/gifts	0.04	0.19
Food/drug	0.04	0.20
Hardware/home improvement	0.09	0.29
Health/beauty	0.04	0.19
Housewares/home furnishings	0.06	0.24
Jewelry	0.02	0.12
Mass merchant	0.07	0.26
Office supplies	0.02	0.14
Specialty/non-apparel	0.14	0.35
Sporting goods	0.07	0.26
Toys/hobbies	0.02	0.13

First page. Some of our robustness checks use an alternative measure of position that represents the number of times a retailer's link appears on the first page of Google search results. As shown in Table 3, the average site appears on the first page about 24 times, but the standard deviation for this variable is substantial. Indeed, one retail site appeared on the first page for 3,306 of the search terms in our sample.

Name. As discussed above, our control for the prominence of a retailer's name is its number of name searches, and we create two measures: One is based on name searches at Google and the other is based on name searches at Bing. Unfortunately, comScore only records searches for terms and phrases that exceed an unknown threshold, and as a result, 32 percent of the retail sites on Google and 56 percent of those on Bing had so few name searches that comScore did not report them. Retail sites in this category were coded as having the "worst" name prominence. The remaining retailers—those in which the number of name searches is observed—were categorized

into five quintiles based on their total number of name searches, ranked from "poor" (number of name searches is in the lowest quintile of observed name searches) to "best" (top quintile of observed name searches).[11] Thus, our primary measure of name prominence is a dummy variable that equals one when the site's number of name searches is in one of these categories (and zero otherwise).

Ads. In addition to displaying organic results, search engines also display paid (or sponsored) results. Paid results are essentially advertisements that expose users to the names of retailers, and sometimes contain other information (such as shipping charges or feedback ratings) that may impact a site's organic clicks. Based on the data collected by querying Google and Bing, we computed a variable called *Ads*. This represents the number of times each retailer's ads were displayed on the first page of search results. As shown in Table 3, the average number of ads a retailer in our sample had on the first page was about 18. Again, there is considerable variation; some retailers had no ads on the first page, while one retailer had 1,716 ads on the first page.

Social network presence. Sites that have a presence on Facebook or Twitter get additional exposure to potential searchers, and this might affect a site's organic clicks. For each retail site, we created a dummy that equals 1 if its parent company has a presence on Facebook or Twitter, and zero otherwise. As shown in Table 3, 89 percent of sites in our sample have a social network presence.

Site age. One might speculate that firms that have been online for a longer period are better known or have had more time to build customer-centric features into their systems. These sorts of considerations are captured in *Site age.* As shown in Table 3, retailer sites in our sample have been around for an average of 13 years. The youngest site in our sample is aged two, while the oldest retail site has been around for 23 years.[12]

Web-only retailer. To control for potential differences in organic clicks for pure-play online retailers (such as Amazon) and online retailers that also have a brick-and-mortar presence (such as Walmart), we constructed a dummy variable that equals one if the retail site is web only and zero otherwise. As shown in Table 3, 35 percent of the retail sites in our sample do not have a brick-and-mortar presence.

Retail segment fixed effects. Finally, to control for systematic differences in organic clicks across different retail

---

[11] Similar to position categorization, the intermediate categories are "below median" (second lowest quintile), "median" (middle quintile), and "above median" (next to highest quintile). A retail site in the "below median" category has a number of name searches that is in the second lowest quintile of observed name searches, while a retail site in the "median" category has a number of name searches that is in the middle quintile of observed name searches. The "above median" category contains all those retail sites that have a number of name searches that is in the second highest quintile of observed name searches.

[12] Based on the Internet Retailer data, Peapod is the oldest online retailer in our sample. We were initially suspicious of the 23 year age reported by Internet Retailer, and subsequently visited Peapod's website to conduct an audit. According to its site, Peapod began taking orders in 1990; the orders "...were placed online, just not over the Internet. Peapod would provide software to customers and even sell the modems customers would need to dial in directly to Peapod."

segments, each retail site was assigned to one of the 15 retail segments identified in the Internet Retailer dataset. Table 3 shows these segments and the percentage of sites in our data within each segment. All of our econometric specifications include retail segment fixed effects.

3.2. Baseline estimates

As noted above, one limitation of the Search Planner data is that, for each search term or phrase, comScore only records the number of searches when the number is above an unspecified threshold. An unobserved number of searches (or missing search term or phrase) does not mean the site did not receive any traffic from a given search term or phrase. It simply means that the number of searches was below this threshold. Our methodology attempts to mitigate this concern in two ways.

First, we use categories rather than levels to measure name searches. Thus, while the actual number of name searches is not observed for some sites, such sites necessarily have fewer name searches than those for which comScore does report the number of searches. So long as sites within the "worst" name category do not have heterogeneous name prominence effects, the ability to observe the number of name searches of these firms would not impact our analysis. Similarly, our use of page position categories accounts for the fact that we do observe the actual position of retailers appearing beyond page 5 of search results. Again, such firms are included in the "worst" position category.

Second, we use quantile regressions to mitigate problems stemming from the fact that comScore does not disclose the number of organic clicks stemming from non-name searches when the number of clicks is below an unspecified threshold, T. We initially explored two extremes to account for this issue via OLS. In the first, we set T equal to the minimum number of clicks observed in our sample for each search term or phrase and assumed that sites with unobserved clicks received $T-1$ clicks. In the second, we assumed that sites with missing numbers of clicks received only 1 click. Results based on these two extremes were qualitatively similar—and similar to the results reported below—but the magnitude of the estimates were sensitive to these two extremes. In contrast, the results reported below, based on quantile regressions, are robust to these two extremes.[13]

We consider specifications of the form

$$\ln(Clicks_i)$$
$$= a + \sum_{b=1}^{5} \alpha_b Position_{i,b} + \sum_{b=1}^{5} \beta_b Name_{i,b} + \gamma X_i + \varepsilon_i \qquad (1)$$

where *Position* and *Name* are dummy variables corresponding to the categories of the position and name variables discussed above, and X is a vector of other potential controls, including retail segment fixed effects. The omitted categories are the "worst" position and "worst" name categories. The coefficients for the position and name dum-

mies have the usual interpretation: A firm moving from position category b' to position category b'' experiences a $[\exp(\alpha_{b''} - \alpha_{b'}) - 1] \times 100$ percentage change in organic clicks, and likewise for the name category.

Baseline quantile regression results are presented in Table 4. All specifications include retail segment fixed effects, so the only other control in specification (1) is *Position*. Consistent with the studies of other platforms highlighted in the introduction, sites with better positions on Google results pages obtain significantly more organic clicks than sites with inferior positions.

Specification (2) adds controls for the prominence of retailer names. Two aspects of this specification are noteworthy. First, adding controls for name prominence reduces the magnitude of the estimated position effects, although all of the position coefficients remain statistically significant at the 5 percent level. Second, while the impact of name prominence on clicks is not perfectly monotonic in this simple specification, name prominence has a positive and statistically significant effect on clicks, and firms with more prominent names get more organic clicks than firms in the "worst" (the omitted) category. On balance, these results suggest that name prominence is a potentially important determinant of clicks.

The third column adds controls for ads appearing on the first page of organic search results. Consistent with Yang and Ghose (2010), the coefficient is positive and statistically significant.[14] This is consistent with exposure to ads increasing the prominence of the firm's name or link, or providing other information about retailer characteristics that increases the firm's organic clicks. Adding this control, however, does little to the estimated effects of position and name prominence on organic clicks.

If the effects of name prominence identified in specifications (2) and (3) were purely the result of omitted variables or spurious correlation with better measures of firms' efforts to enhance brand awareness and clicks, the results would not be robust to the inclusion of other controls. Specification (4) adds a control for whether retailers have a social network presence. The coefficient is positive and statistically significant, but does not materially change the estimated effects of name prominence. Likewise, one might speculate that site age is a useful proxy for the prominence of retail sites, since sites that have been around longer are more likely to be better known than newer sites. The results in column 5 indicate that this variable adds little explanatory power over and above our primary measure of name prominence, and in any event does not affect the estimated effects of name prominence. Finally one might worry that the previous results are driven by differences between web-only and bricks-and-clicks retailers. Column 6 shows that the results are robust to these controls as well.

These results suggest that the prominence of a site's name and position are both important determinants of the

---

[13] The reported results assume each retailer received at least one organic click.

---

[14] Notice that this coefficient does not capture the impact of a relative price change, and therefore does not mean that organic and paid links are "complements" in the usual economic sense. Establishing such a relationship would require a natural experiment along the lines of Goldfarb and Tucker (2011), who show that online and offline ads are substitutes.

Table 4
Baseline model.

Dependent variable: logarithm of clicks on Google							
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Position on Google							
Poor	1.544	1.213	1.077	1.034	0.992	1.092	1.409
	(0.245)*	(0.233)*	(0.224)*	(0.221)*	(0.218)*	(0.215)*	(0.252)*
Below median	2.490	2.111	1.927	2.068	2.000	2.009	2.434
	(0.245)*	(0.241)*	(0.234)*	(0.230)*	(0.228)*	(0.225)*	(0.255)*
Median	3.282	2.858	2.597	2.617	2.501	2.538	3.022
	(0.244)*	(0.257)*	(0.257)*	(0.254)*	(0.251)*	(0.248)*	(0.272)*
Above median	4.580	3.969	3.562	3.652	3.569	3.533	4.211
	(0.246)*	(0.283)*	(0.290)*	(0.287)*	(0.285)*	(0.281)*	(0.294)*
Best	5.762	4.892	4.276	4.266	4.184	4.059	5.076
	(0.247)*	(0.327)*	(0.363)*	(0.357)*	(0.354)*	(0.350)*	(0.358)*
Name on Google							
Poor		0.578	0.444	0.423	0.450	0.568	
		(0.227)*	(0.217)*	(0.214)*	(0.212)*	(0.210)*	
Below median		0.781	0.643	0.743	0.670	0.828	
		(0.242)*	(0.233)*	(0.229)*	(0.227)*	(0.226)*	
Median		1.109	0.834	0.808	0.866	1.049	
		(0.254)*	(0.246)*	(0.242)*	(0.240)*	(0.242)*	
Above median		0.622	0.503	0.518	0.538	0.840	
		(0.290)*	(0.278)	(0.274)	(0.271)*	(0.274)*	
Best		1.357	1.244	1.103	1.218	1.660	
		(0.322)*	(0.314)*	(0.309)*	(0.306)*	(0.313)*	
ln(Ads on Google)		0.237	0.245	0.256	0.209	0.300	
		(0.072)*	(0.071)*	(0.070)*	(0.069)*	(0.080)*	
Social network presence				0.475	0.492	0.601	0.634
				(0.212)*	(0.210)*	(0.208)*	(0.247)*
Site age					0.022	0.038	0.016
					(0.019)	(0.019)	(0.023)
Web only retailer						0.437	0.140
						(0.150)*	(0.170)
Observations	757	757	757	757	757	757	757
Pseudo R^2	0.41	0.43	0.43	0.44	0.44	0.44	0.42

Notes: Standard errors in parentheses; *denotes significant at 5%. All specifications include a constant and retail segment fixed effects.

organic traffic retailers receive following product searches on general search engines. Comparisons of the position effects in columns 1 and 2, as well as columns 6 and 7, reveal that failure to account for the prominence of a site's name results in inflated estimates of the position effects.

3.3. Endogeneity

Following much of the literature that examines the impact of position and retailer characteristics on clicks at other platforms, the baseline results discussed above assume that the explanatory variables are not correlated with the errors in Eq. (1). Recall that the dependent variable in these regressions is the logarithm of non-name organic clicks on Google, and the controls for *Name, Position*, and *Ads* are based on name searches, positions and sponsored ads on Google. One might reasonably worry that these controls are correlated with the error term (or errors) in the regression, which captures latent factors influencing the organic clicks that retailers get through product searches at Google.

As an initial matter, if one used *total* organic clicks at Google (including organic clicks based on name searches) as the dependent variable, it would hardly be surprising to find that the number of name searches at Google is positively related to a retailer's total organic clicks. We avoid this issue by using organic clicks stemming from non-

name searches as the dependent variable. Despite this, one might worry that factors that are unobserved and unrelated to name prominence, but influence name searches on Google, might also impact organic clicks on Google. If this is the case, our control for name prominence will be correlated with the error term in the regression, potentially biasing the results.

The results reported below attempt to mitigate this first concern by using name searches at Bing rather than Google to control for name prominence. Our earlier discussion of the merits of using name searches as a control for name prominence did not rely on the identity of the search engine, so one would also expect the number of name searches on Bing to be a useful summary statistic for the prominence of a retailer's name. For example, a retailer running a TV advertising campaign would presumably get more organic clicks on Google, as well as more name searches on Bing. We may thus use name searches on Bing rather than Google to control for factors influencing the prominence of retailer's names. Given differences in the Google and Bing algorithms, and differences in their populations of users, it seems reasonable to assume that unobserved factors subsumed in the error of the regression—but that influence *organic clicks* at Google—are independent of *name searches* on Bing.

There are, of course, scenarios where this assumption might not hold. For example, a power outage that

only impacts Amazon customers will adversely affect the organic clicks it receives through Google, and also reduce the number of name searches for Amazon on Bing. In this case, the reduction in name searches on Bing is unrelated to a change in name prominence, and name searches on Bing will be correlated with the error in the regression. Notice that the concern in this scenario is mitigated when the analysis is based on aggregate data from a single month and includes retail segment and web-only fixed effects.

A more serious concern, in our judgment, is the potential endogeneity of two of our other controls, *Position* and *Ads*. First, consider *Position*. If firms' positions on Google are predetermined at the time consumers make their click decisions, it would be unnecessary to adjust for endogeneity. One might argue this is true for our data, since we are using cross-sectional data rather than a time series of data to identify position effects. However, our data come from an entire month and, in practice, search engines continually refine and optimize their algorithms in an attempt to present searchers with the most relevant organic results. From the standpoint of estimation, this means that a site's position in Google's list of organic results depends on its past clicks at Google. To further complicate matters, past organic clicks on Google depend on past positions on Google and past name prominence. In short, there are good reasons to worry that our Google position variable is correlated with the error in the regression model.

A related worry is the potential endogeneity of *Ads*—another control in our specifications. Search engines make money when users click on ads, and thus Google has an incentive to take into account the likelihood that a retailer's ad will be clicked when deciding whether to display it on the first page of Google results. Again, this means that Google's decision to display a retailer's ad depends on its past clicks on Google. Thus, there is also reason to believe that another variable in our analysis—*Ads*—may be endogenous.

Our strategy for dealing with the potential endogeneity of *Position* and *Ads* is to use information about position and ads on Bing as instruments for position and ads on Google. Since Google's position and ad decisions are based on past clicks at Google—not Bing—these instruments would seem to satisfy the requirements of valid instruments. Again, there are scenarios in which the validity of these instruments might fail. For instance, a demand shock for a particular product—say, a tablet PC—might lead to more organic and paid clicks for retailers selling that product. This would elevate their positions on both Google and Bing. This shock might also lead to more name searches on both Google and Bing for retailers selling tablet PCs. In this case, the instruments will be correlated with the error in the regression. This sort of concern is mitigated, however, if the specifications include retail segment fixed effects and estimation is based on aggregate clicks (rather than product-specific clicks), as well as data from a single month.

In summary, we address these three endogeneity concerns by using name searches on Bing to control for name prominence, and position and ads on Bing as instruments for positions and ads on Google. For the reasons

discussed above, all specifications include retail segment fixed effects and we continue to identify effects using cross-retailer variation in a single month. The standard two-stage procedure is used to control for the endogeneity of *Ads*. For positions, we use a use a two-stage ordered probit approach in order to facilitate comparisons with the baseline results. We first estimate an ordered probit model using categorized positions on Google as the dependent variable, with categorized positions on Bing and the other exogenous covariates as explanatory variables. We then use the predicted values from this first stage to create instrumented position categories, which we use in place of the original position variables. In the second stage, we estimate this equation using OLS.

Results controlling for endogeneity are displayed in the first column of Table 5. Comparison of these results with the most general specification in the baseline model (column 6 of Table 4) reveals that the results are qualitatively similar. Importantly, however, controlling for endogeneity increases the quantitative importance of name prominence relative to position prominence.

The results in Table 5 reveal that retailers with "poor" average positions do not receive numbers clicks that are statistically different from those with the "worst" positions (the omitted category). However, sites with more prominent screen positions get significantly more clicks, in both the economic and statistical sense. Other things equal, a retailer moving from the "median" to the "best" screen position enjoys an 80 percent[15] increase in organic clicks. On balance, the position effects are in line with what one would expect based on studies documenting the impact of position on clicks at other platforms.

The first column of Table 5 also shows that name prominence remains an economically and statistically important determinant of clicks across all categories. Other things equal, a firm that moves from the "median" level of name prominence to the "best" obtains 154 percent more clicks.[16] In short, when one controls for endogeneity, name prominence appears to have a greater impact on clicks than position prominence.

The estimated impact of *Ads* on organic clicks in the first column of Table 5 is almost three times larger than the corresponding estimates in the baseline model. Controlling for endogeneity of *Ads*, a 1 percent increase in a retailer's ads on the first page of Google increases organic clicks by about 0.758 percent, other things equal. It is reassuring that our finding that ads actually enhance organic clicks is consistent with results by Yang and Ghose (2010), who use an entirely different methodology and primarily focus on sponsored search.

The results in Table 5 indicate that retailers with a presence on a social network (Twitter or Facebook) get more organic clicks, but the effect is not statistically significant. On the other hand, *Site Age* is statistically significant: Retailers with older sites tend to receive more clicks. At the mean of the data, the implied elasticity of a firm's organic

---

[15] Since $(\exp(1.723 - 1.135) - 1) \times 100$ is about 80 percent.

[16] Since $(\exp(2.754 - 1.820) - 1) \times 100$ is about 154 percent.

Table 5
Specifications controlling for endogeneity.

Dependent variable: logarithm of clicks on Google		
Variable	(1) Two-stage ordered probit	(2) Two-stage ordered probit
Position on Google		
Poor	−0.341	
	(0.350)	
Below median	0.769	
	(0.300)*	
Median	1.135	
	(0.276)*	
Above median	1.473	
	(0.288)*	
Best	1.723	
	(0.296)*	
First page		
Poor		−0.182
		(0.348)
Below median		0.158
		(0.338)
Median		1.040
		(0.277)*
Above median		1.234
		(0.260)*
Best		1.932
		(0.268)*
Name on Bing		
Poor	1.377	1.358
	(0.205)*	(0.207)*
Below median	2.042	1.981
	(0.224)*	(0.234)*
Median	1.820	1.835
	(0.241)*	(0.250)*
Above median	2.146	2.223
	(0.231)*	(0.250)*
Best	2.754	2.776
	(0.308)*	(0.310)*
ln(Ads on Google)	0.758	0.736
	(0.072)*	(0.071)*
Social network presence	0.180	0.180
	(0.268)	(0.275)
Site age	0.090	0.082
	(0.028)*	(0.028)*
Web Only Retailer	0.490	0.450
	(0.214)*	(0.214)*
Observations	757	757
Pseudo R^2	0.51	0.52

Notes: Standard errors in parentheses; *denotes significant at 5%. All specifications include a constant and retail segment fixed effects.

clicks with respect to the age of its site is about 1.2.[17] This is consistent with retailers with older sites getting more clicks because they have had time to work out kinks. But it is also possible that this control is accounting for some elements of name prominence that are not captured in our measure.

Finally, notice that the coefficient associated with the web-only retailer control is positive and statistically significant; other things equal, web-only retailers enjoy more organic clicks from product searches on Google than bricks-and-clicks retailers. While we are unaware of any other study that has examined this issue in the

context of organic product search, studies based on much earlier data from price comparison sites generally reach the opposite conclusion. That is, during the early 2000s, bricks-and-clicks retailers that listed products on price comparison sites tended to get more clicks, other things equal, than web-only retailers. Two factors may account for the difference with our findings. First, during the early to mid 2000s, many web-only retailers were unknown entities, and consumers were leery of them. Thus, negative web-only effects during the early 2000s are consistent with web-only retailers having less prominent names during that period. Since then, consumers have become much more comfortable conducting transactions online, and have amassed additional knowledge about the business practices and reputations of online retailers—including many of those in our sample of the top 757 online retailers. We therefore do not find it surprising that, based on our more recent data, web-only retailers tend to receive more organic clicks, other things equal. Indeed, the results are consistent with the view that retailers specializing in internet sales provide better online shopping experiences than retailers attempting to operate in both online and traditional markets.

3.4. Robustness

Before concluding, we briefly summarize some additional robustness checks conducted during the course of our analysis.

One might worry that our method of constructing position categories somehow masks the importance of being included on the first page of organic search results. And despite our use of categories, one might worry that our position measure is sensitive to the positions assigned to retailers not appearing on the first five pages of search results.

To address these concerns, we also report results based on our alterative measure of position. Recall that our alternative measure, *First page*, represents the number of times a given retailer appears on the first page of organic search results. Retail sites that do not appear on the first page of any of the search results were coded as having the "worst" position. The remaining retail sites were again categorized into five quintiles, ranked from "poor" to "best." We then used the used the instrumental variables approach described above to control for endogeneity of *First page* and *Ads* using position and advertising data from Bing as instruments.

The second column of Table 5 shows the results. With this alternative measure of position, notice that retailers more frequently appearing on the first page of organic search results continue to receive significantly more clicks than those in less prominent positions. But more to the point, our findings that sites with more prominent names receive significantly more organic clicks than less prominent sites, and that name prominence has a larger impact on clicks than position when one controls for endogeneity, continue to hold when this alternative measure is used. Finally, note that using this alternative measure of position does little to the estimated effects of the other controls in Table 5.

---

[17] Since $\frac{d \ln(\text{Clicks})}{d(\text{Site age})}$ (Site age) $= (.090)(13.04)$, or about 1.2.

Table 6
Baseline model.

Dependent variable: logarithm of clicks on Bing							
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Position on Bing							
Poor	6.178	5.752	5.752	5.530	5.147	4.412	5.073
	(0.661)*	(0.627)*	(0.630)*	(0.633)*	(0.645)*	(0.599)*	(0.595)*
Below median	7.346	6.274	6.252	5.959	5.699	5.025	5.823
	(0.656)*	(0.639)*	(0.653)*	(0.656)*	(0.669)*	(0.633)*	(0.613)*
Median	8.214	7.065	7.065	7.006	6.534	5.829	6.878
	(0.658)*	(0.652)*	(0.674)*	(0.676)*	(0.689)*	(0.653)*	(0.628)*
Above median	9.113	7.708	7.708	7.507	7.171	6.151	7.561
	(0.658)*	(0.681)*	(0.732)*	(0.737)*	(0.752)*	(0.719)*	(0.678)*
Best	11.012	8.608	8.608	8.393	7.954	7.150	8.888
	(0.658)*	(0.781)*	(0.907)*	(0.910)*	(0.932)*	(0.894)*	(0.811)*
Name on Bing							
Poor		1.377	1.398	1.417	1.371	1.341	
		(0.718)	(0.722)	(0.725)	(0.738)	(0.701)	
Below median		1.627	1.627	1.626	1.642	2.077	
		(0.742)*	(0.746)*	(0.749)*	(0.763)*	(0.735)*	
Median		2.204	2.204	2.210	2.228	2.658	
		(0.739)*	(0.744)*	(0.747)*	(0.761)*	(0.729)*	
Above median		2.031	2.031	2.048	2.130	2.385	
		(0.765)*	(0.770)*	(0.774)*	(0.788)*	(0.763)*	
Best		3.218	3.239	3.316	3.337	3.179	
		(0.852)*	(0.877)*	(0.881)*	(0.897)*	(0.884)*	
ln(Ads on Bing)			0.000	-0.020	-0.009	0.089	0.242
			(0.257)	(0.258)	(0.263)	(0.250)	(0.245)
Social network presence				0.217	0.290	0.404	0.505
				(0.627)	(0.639)	(0.608)	(0.606)
Site age					0.058	0.028	0.037
					(0.058)	(0.057)	(0.057)
Web only retailer						0.313	-0.710
						(0.442)	(0.422)
Observations	757	757	757	757	757	757	757
Pseudo R^2	0.24	0.29	0.29	0.29	0.29	0.31	0.28

Notes: Standard errors in parentheses; *denotes significant at 5%. All specifications include a constant and retail segment fixed effects.

We replicated all of the above analysis using an analogous dataset constructed for Bing. We then used the methodology described above to estimate a model in which the dependent variable is the logarithm of organic clicks on Bing. These results, which are presented in Tables 6, 7 and 8, are consistent with the findings reported in the text based on the data for Google. We note that the Bing data are "thinner" than the Google data, and thus some of the estimated coefficients that were significant at the 5 percent level with the Google data are only significant at the 10 percent level with these data.

As an additional robustness check, we replicated our analysis using a narrower definition of name searches that only includes the name or URL of the site (e.g., excludes "buy camera at amazon" but includes "amazon. com" and "amazon"). Under this definition, a name search is purely a navigational search—consumers using this query are merely attempting to navigate to a particular firm's site. As shown in Table 7, the results are similar to those based on the broader definition of a name search.

4. Concluding remarks

There is considerable evidence that retailer characteristics impact click-through rates at price comparison sites, marketplaces, and auction sites. For product searches at Google and Bing, these factors are essentially embodied in retailers' names. This paper has provided evidence that name searches are a potentially useful way of controlling for differences in the prominence of different retailers' (and universities') names. Our analysis also suggests that failure to account for name prominence results in inflated estimates of the impact of position on clicks. We believe that our results are of potential interest to companies interested in search engine optimization (see Baye et al. (2014)), as well as policymakers working on issues related to online product search.

While we have focused on organic (or natural) clicks in the present paper, our proposed measure of name prominence is also potentially useful for controlling for factors other than position that influence clicks on sponsored (or paid) links. As noted in the introduction, some researchers have constructed similar controls through expensive consumer surveys. While the present paper utilizes a single month of data to demonstrate the potential utility of our measure, one can readily use the comScore data to construct time series measures of name prominence. While the comScore data is not cheap, constructing an 18 month time series from these data is far less expensive than conducting 18 monthly consumer surveys.

Table 7
Alternative measures of name recognition and position.

Dependent variable: logarithm of clicks on Bing				
Variable	(1)	(2)	(3)	(4)
	Measure of name recognition			
	Phrase contains name or domain of retailer		Only name or domain of retailer	
Position on Bing				
Poor	2.681		4.291	
	(0.611)*		(0.556)*	
Below median	2.988		4.174	
	(0.651)*		(0.583)*	
Median	3.845		5.072	
	(0.679)*		(0.619)*	
Above median	4.026		5.712	
	(0.768)*		(0.702)*	
Best	5.433		6.889	
	(0.918)*		(0.854)*	
First Page				
Poor		1.274		1.529
		(0.589)*		(0.613)*
Below median		1.696		1.703
		(0.604)*		(0.641)*
Median		2.509		2.515
		(0.748)*		(0.731)*
Above median		2.689		3.064
		(0.752)*		(0.824)*
Best		4.074		4.236
		(0.908)*		(0.972)*
Name on Google				
Poor	3.182	4.154	1.920	4.140
	(0.619)*	(0.603)*	(0.548)*	(0.628)*
Below median	4.336	5.807	2.960	5.466
	(0.658)*	(0.638)*	(0.596)*	(0.681)*
Median	4.517	5.583	3.021	5.444
	(0.697)*	(0.685)*	(0.626)*	(0.722)*
Above median	4.721	5.867	3.012	5.486
	(0.767)*	(0.750)*	(0.699)*	(0.815)*
Best	5.007	6.136	3.368	5.775
	(0.899)*	(0.891)*	(0.838)*	(0.970)*
ln(Ads on Bing)	0.039	0.180	0.189	0.296
	(0.257)	(0.249)	(0.215)	(0.248)
Social network presence	0.508	0.654	0.362	0.500
	(0.614)	(0.600)	(0.541)	(0.625)
Site age	0.051	0.047	0.049	0.053
	(0.057)	(0.056)	(0.051)	(0.058)
Web only retailer	0.132	0.153	0.079	0.074
	(0.451)	(0.439)	(0.395)	(0.454)
Observations	757	757	757	757
Pseudo R^2	0.32	0.32	0.35	0.33

Notes: Standard errors in parentheses; *denotes significant at 5%. All specifications include a constant and retail segment fixed effects.

We used comScore data in order to mitigate long-tail issues, and thus to offer results based on the largest possible cross-section of retailers and search terms. We note that it is also possible to measure name searches (for a more limited number of firms and search terms) using Google Trends, which is free. For this reason, the measure of name prominence introduced in this paper is a readily available and potentially powerful way of controlling for name prominence in environments unrelated to organic search.

In conclusion, note that we utilized across-retailer variation and reduced-form quantile regressions on data from a single month to demonstrate that this measure requires neither detailed micro or time-series data nor strong structural assumptions. Our rationale for using bins and quantile regressions to deal with long-tail issues is that this approach is "minimalist" in terms of the number of extraneous structural assumptions needed to conduct the analysis. While this is appropriate for demonstrating the potential utility of our proposed measure, more detailed modeling of the long tail—including structural modeling of consumer choice at the level of search terms—is required before more definitive estimates of the relative returns on investments in name prominence and position prominence can be made. This is the direction of our future work in the area.

Table 8
Specifications controlling for endogeneity.

Dependent variable: logarithm of clicks on Bing		
Variable	(1) Two-stage ordered probit	(2) Two-stage ordered probit
Position on Bing		
Poor	−0.111	
	(0.414)	
Below median	−0.920	
	(0.421)*	
Median	0.607	
	(0.433)	
Above median	0.606	
	(0.376)	
Best	0.700	
	(0.393)	
First Page		
Poor		−0.958
		(0.270)*
Below median		−1.390
		(0.325)*
Median		−0.071
		(0.218)
Above median		0.172
		(0.206)
Best		0.016
		(0.242)
Name on Google		
Poor	1.737	1.640
	(0.429)*	(0.278)*
Below median	3.481	2.223
	(0.405)*	(0.287)*
Median	3.867	2.778
	(0.408)*	(0.269)*
Above median	4.918	3.495
	(0.406)*	(0.265)*
Best	5.622	4.240
	(0.448)*	(0.322)*
ln(Ads on Bing)	0.543	0.524
	(0.134)*	(0.076)*
Social network oresence	0.557	0.614
	(0.362)	(0.287)*
Site age	0.081	0.066
	(0.036)*	(0.026)*
Web only retailer	0.335	0.719
	(0.281)	(0.208)*
Observations	757	757
PseudoR^2	0.47	0.52

Notes: Standard errors in parentheses; *denotes significant at 5%. All specifications include a constant and retail segment fixed effects.

References

Ansari, A., Mela, C., 2003. E-customization. J. Mark. Res. 40 (2), 131–146.
Arbatskaya, M., 2007. Ordered search. RAND J. Econ. 38 (1), 119–126.
Armstrong, M., Vickers, J., Zhou, J., 2009. Prominence and consumer search. RAND J. Econ. 40 (2), 209–233.
Armstrong, M., Zhou, J., 2011. Paying for prominence. Econ. J. 121 (556), F368–F395.
Baye, M.R., Morgan, J., 2009. Brand and price advertising in online markets. Manag. Science 55 (7), 1139–1151.
Baye, M.R., Gatti, R.J.R., Kattuman, P., Morgan, J., 2009. Clicks, discontinuities, and firm demand online. J. Econ. Manag. Strateg. 18, 935–975.
Baye, M. R., De los Santos, B., Wildenbeest, M. R., 2014. Search engine optimization: what drives organic traffic to retail sites?" Kelley School of Business Research Paper No. 2014-2047.
Brynjolfsson, E., Dick, A., Smith, M.D., 2010. A nearly perfect market? Quant. Mark. Econ. 8 (1), 1–33.
De los Santos, B., Koulayev, S., 2013. Optimizing click-through in online rankings for partially anonymous consumers. Working paper.
Dewally, M., Ederington, L., 2006. Reputation, certification, warranties, and information as remedies for seller-buyer information asymmetries: Lessons from the online comic book market. J. Bus. 79 (2), 693–729.
Drèze, X., Zufryden, F., 2004. Measurement of online visibility and its impact on internet traffic. J. Interact. Mark. 18 (1), 20–37.
Ellison, G., Ellison, S.F., 2009. Search, obfuscation and price elasticities on the internet. Econometrica 77 (2), 427–452.
Ghose, A., Yang, S., 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. Manag. Sci. 55 (10), 1605–1622.
Goldfarb, A., Tucker, C., 2011. Search engine advertising: Channel substitution when pricing *Ads* to context. Manag. Sci. 57 (3), 458–470.
Goldfarb, A., Lu, Q., Moorthy, S., 2009. Measuring brand value in an equilibrium framework. Manag. Sci. 28 (1), 69–86.
Hossain, T., Morgan, J., 2006. Plus shipping and handling: Revenue (non) equivalence in field experiments on ebay. BE J. Econ. Anal. Policy 6 (2).
Houser, D., Wooders, J., 2006. Reputation in auctions: Theory, and evidence from ebay. J. Econ. Manag. Strateg. 15 (2), 353–369.
Jerath, K., Ma, L., Park, Y., Srinivasan, K., 2011. A 'position paradox' in sponsored search auctions. Mark. Sci. 30 (4), 612–627.
Jin, G.Z., Kato, A., 2006. Price, quality, and reputation: Evidence from an online field experiment. RAND J. Econ. 37 (4), 983–1005.
Melnik, M.I., Alm, J., 2002. Does a seller's ecommerce reputation matter? evidence from ebay auctions. J.Ind. Econ. 50 (3), 337–349.
Novarese, M., Wilson, C. M., 2012. Being in the right place: a natural field experiment on list position and consumer choice, Working paper.
Rutz, O.J., Bucklin, R.E., 2011. From generic to branded: A model of spillover dynamics in paid search advertising. J. Mark. Res. 48 (1), 87–102.
Sappington, D.E.M., Wernerfelt, B., 1985. To brand or not to brand? a theoretical and empirical question. J. Bus. 58 (3), 279–293.
Smith, M.D., Brynjolfsson, E., 2001. Consumer decision-making at an internet shopbot: Brand still matters. J. Ind. Econ. 49 (4), 541–558.
Yang, S., Ghose, A., 2010. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? Mark. Sci. 29 (4), 602–623.
Yao, S., Mela, C.F., 2011. A dynamic model of sponsored search advertising. Mark. Sci. 30 (3), 447–468.