Search Engine Optimization: What Drives Organic Traffic to Retail Sites?

Michael R. Baye

Department of Business Economics and Public Policy
Kelley School of Business
Indiana University
Bloomington IN 47405
mbaye@indiana.edu

Babur De los Santos

Department of Economics
Clemson University
Clemson SC 29634
babur@clemson.edu

Matthijs R. Wildenbeest

Department of Business Economics and Public Policy
Kelley School of Business
Indiana University
Bloomington IN 47405
mwildenb@indiana.edu

The lion's share of retail traffic through search engines originates from organic (natural) rather than sponsored (paid) links. We use a dataset constructed from over 12,000 search terms and 2 million users to identify drivers of the organic clicks that the top 759 retailers received from search engines in August 2012. Our results are potentially important for search engine optimization (SEO). We find that a retailer's investments in factors such as the quality and brand awareness of its site increases organic clicks through both a direct and an indirect effect. The direct effect stems purely from consumer behavior: The higher the quality of an online retailer, the greater the number of consumers who click its link rather than a competitor in the list of organic results. The indirect effect stems from our finding that search engines tend to place higher quality sites in better positions, which results in additional clicks because consumers tend to click links in more favorable positions. We also find that consumers who are older, wealthier, conduct searches from work, use fewer words, or include a brand name product in their search are more likely to click a retailer's organic link following a product search. Finally, the quality of a retailer's site appears to be especially important in attracting organic traffic from individuals with higher incomes. The beneficial direct and indirect effects of an online retailer's brand equity on organic clicks, coupled

with the spillover effects on traffic through other online and traditional channels, leads us to conclude that investments in the quality and brand awareness of a site should be included as part of an SEO strategy.

1. Introduction

Search engines are an important way of obtaining information on the Internet. According to Alexa Traffic Rank, Google.com is the most popular web site in the United States as well as in the world, and in May 2011, it was the first web site to achieve one billion monthly unique visitors.[1] Many people use search engines as a starting point for navigating the Web, making search engines a crucial link in connecting content providers and users. This has spurred a sizable literature on search marketing that studies clicking behavior at search engines. To date, most of this literature has concentrated on the sponsored links that are typically displayed alongside organic links when consumers conduct searches.

Although most of the economics and marketing literature on search engines has focused on paid clicks, the bulk of the traffic retailers receive through search engines is actually through unpaid clicks on organic links (Jerath et al., 2014).[2] For this reason, more advertisers engage in search engine optimization (SEO) to improve organic clicks than purchase sponsored links to get paid clicks (Berman and Katona, 2013). To the best of our knowledge, the present paper is the first to provide search marketers with information on drivers of organic clicks to aid in SEO.

Existing studies of sponsored search are typically based on a modest number of search terms and the corresponding number of paid clicks received by a single retailer. Our research complements these studies by focusing on the organic clicks that 759 retail sites received from more than 12,000 search terms. There is considerable cross-sectional variation in our data: It includes Web-only retailers as well as traditional retailers and covers 15 different retail segments including apparel, electronics, and mass merchants. For each of these search terms, we observe which retail sites received organic clicks as well as the number of clicks. We also obtained data from the first five pages of search results on Google and Bing for each search term, and this ultimately permits us to quantify the impact on organic clicks of a site's rank (position) in the search results. Our data also include several different measures of the accumulated brand equity of online retailers. These data allow us to determine whether consumers are more likely to click the link of a retailer who is perceived to operate a high-quality site (as a result of the retailer's current and past investments in advertising, the depth and breadth of offerings, secure payments, one-click purchases, returns policies, and so on). Ultimately, this permits us to quantify the benefits of SEO strategies that attempt to gain traffic by improving a retailer's rank in organic search results, versus gaining traffic by improving the quality and brand awareness of a site.

Not surprisingly, we find that a retailer's rank on a results page is an important driver of its organic clicks: Exclusion from the first five pages of results for a search leads to a 90% reduction in organic clicks. For retailers that are listed on the first five pages of results, a 1% improvement in rank leads to 1.3% more organic clicks for that search.

---

1. Alexa Traffic Rank is calculated by combining a web site's average number of daily visitors and page views over the past month.
2. Our data are consistent with this finding.

Importantly, however, we also find that the brand equity of an online retailer is an important driver of organic clicks and that it is easy for search marketers to overlook the benefits of including investments in the quality and brand awareness of its site as part of an SEO strategy. The direct benefit of these and other investments in brand equity is an increase in the number of consumers clicking one retailer's link instead of a competing link on results pages. Moreover, investments in brand equity have an additional indirect effect: Search engines tend to place retailers with stronger brands in better positions, which results in additional increases in organic clicks. Finally, estimated effects of rank on organic clicks are keyword specific, whereas improvements in the brand equity of an online retailer increases clicks associated with all relevant searches. Taking all of these effects into account, we find that brand equity is as important as rank in determining organic clicks.

We also point out that investments that improve site quality and consumer awareness (and more broadly, that enhance an online retailer's brand equity) are likely to have spillover benefits in other channels that are not accounted for in this or other studies of organic and sponsored search. These benefits include increases in clicks through other online channels (such as price comparison sites), increases in the number of direct visits to a retailer's web site, increases in visits through navigational searches at search engines, and increases in traffic at the retailer's physical stores. These considerations—coupled with the fact that position is a zero-sum game and thus a retailer is unlikely to obtain a sustainable advantage through direct efforts to improve its ranking—lead us to conclude that brand equity is one of the more important components of retailers' SEO strategies.

We also find that a retail site's brand equity is especially important in attracting organic traffic from individuals with higher incomes. Our results indicate that consumers who are older, wealthier, conduct searches from work, use fewer words, or include a brand-name product in their search are more likely to click a retailer's organic link following a product search.

The remainder of this section provides an overview of SEO and the related literature. Section 2 discusses our data and describes the econometric methodology underlying our analysis. Section 3 presents our empirical findings, whereas Section 4 provides robustness checks and some additional results. Finally, we conclude in Section 5 with some additional managerial implications of our findings for SEO.

1.1. Search Engine Optimization

Figure 1 highlights the avenues that retailers have for gaining traffic through search engines. This screenshot shows the search results that appear following a search for "shoes online" using Google Search. In this particular example, three different types of links appear: top ads, side ads, and organic results.[3] The top ads (marked by the red box in Figure 1), if any, are the highest listed search results and appear against a yellow background. For this particular search there are three top ads; the maximum number of top ads that may be displayed is four. The organic results (marked by the blue box) are listed below the top ads. Up to ten organic results can appear on a search result page. Finally, the side ads (purple box) appear on the right-hand side of the screen; Google allows for up to eight side ads to be shown on a result page.

---

3. Depending on the search term, up to four bottom ads may appear as well. For the search term "shoes online" no bottom ads were shown.

FIGURE 1. SEARCH RESULTS

One way retailers obtain traffic is through the paid links that appear in top or side ads. Unlike organic links, retailers can directly influence the position of ads, which are displayed and ranked according to the results of an auction that is run in real time. Retailers identify keywords they want to bid on and specify how much they are willing to spend. Google determines the ad rank using a site's maximum bid specified for the keyword and a quality score, which includes factors like click-through rates and relevance. Advertisers only pay when the link is clicked; the cost per click is equal to the minimum amount needed to get a specific position (generalized second-price auction mechanism). There is an extensive literature (discussed below) examining this avenue for obtaining clicks.

A second way retailers obtain traffic through the search engine channel is through clicks on organic results, and this is the focus of our analysis.[4] A site's position in Google's organic search results depends on the site's relevance to a given search term. The exact

---

4. Although our main focus is on organic (non-paid) links, we do take the presence of sponsored links (ads) into account, because they may affect organic clicking behavior.

algorithm that Google uses to determine a site's ranking is proprietary; according to Google, it depends on thousands of factors.[5]

Although the goal of SEO is to optimize the organic traffic a retailer receives through product searches on search engines, the ultimate goal of retailers is presumably to maximize their profits. One of the initial steps in this optimization process is identifying the benefits and costs of different strategies for increasing traffic.[6] Our paper represents a first attempt to examine the benefits side of the ledger, and in particular, to quantify the drivers of retailers' organic clicks.

The first, and most common, SEO strategy is to tweak a site in an attempt to increase the rank of a retailer's organic link on the results pages for a given search term. The presumption is that higher ranks result in more organic clicks, but SEO requires quantifying the effects of rank on clicks. This is one objective of our paper.

One myopic tactic for improving position, known as a "black-hat" strategy, is designed to "trick" search engines into elevating a retailer's rank in the results. Search engines are themselves players, and have incentives to adapt algorithms to ensure that search engine users receive relevant results. Consumers are players too, and may favor links of retailers they know and trust: SEO strategies that focus exclusively on rank (such as spamming links or hiding keywords) might improve the position of a retailer's link but not impact its clicks. For this reason, SEO strategies based on "tricking" or "spamming" engines are unlikely to yield sustainable improvements in rankings, may not result in additional clicks, and can even backfire as a result of negative effects on reputation. Furthermore, it is important to recognize that rankings are effectively a zero-sum game: One retailer can move up on a particular results page only by pushing down the link of another retailer. Thus, although it makes sense for online retailers to ensure that their sites include page titles that accurately describe content, make use of head tags, are free of dead links, and so on, these efforts alone are unlikely to give a particular retailer a sustainable rank advantage because other retailers have incentives to engage in these strategies as well.

A second and more costly SEO strategy—but one that is more likely to yield sustainable improvements in a retailer's organic traffic from search engines—focuses on improving site quality and brand awareness, or more broadly on enhancing the online retailer's brand equity (which embodies current and past investments in advertising, service and return policies, depth and breadth of offerings, prices, etc.). This strategy recognizes that consumers tend to click retailers that are more recognized, trusted, have reputations for providing value (in terms of prices, product depth or breadth), service (well-designed web sites, return policies, secure payment systems), and so on. This SEO strategy is alluded to by Google, which advises businesses to base "…optimization decisions first and foremost on what's best for the visitors of your site. They're the main consumers of your content and are using search engines to find your work. Focusing too hard on specific tweaks to gain ranking in the organic results of search engines may not deliver the desired results." [7]

Although it may be tempting to dichotomize rank-improving and brand-building SEO strategies, these two strategies are not mutually exclusive. Indeed, brand-building

---

5. See http://www.google.com/explanation.html. One of these factors is Google PageRank, which is an algorithm that uses the number of incoming links to measure the relative importance of a web site.

6. SEO is based on the premise that more clicks translates into more sales, thus making organic clicks a useful intermediate metric.

7. See Google's *Search Engine Optimization Starter Guide*, 2013, p. 2. A link to this guide is available online at http://support.google.com/webmasters/bin/answer.py?hl=en&answer=35291

activities may indirectly result in better positions. Furthermore, unlike direct investments in position, brand-building investments are not a perfect zero-sum game because consumers always have an option not to click on any organic search results. For example, some consumers using the search phrase "tennis shoes" might not recognize any retailers in the list of search results, and hence not click on any links. If these retailers all invest in the brand awareness of their sites, they each may receive additional clicks because consumers are less likely to exercise the outside option.

1.2. Related Literature

Our paper is connected to several different literatures, including a handful of academic papers on SEO which provide important theoretical insights into SEO (Sen, 2005; Xing and Lin, 2006; Berman and Katona, 2013). These papers highlight several features of the equilibrium interaction between web sites and search engines that we take into account in our empirical analysis, including the endogeneity of the rank of organic links and the position of sponsored links in search results. To the best of our knowledge, there is no antecedent empirical research on SEO.

There is, however, a sizeable theoretical and empirical literature on search engines that focuses on the sponsored links that appear alongside the organic results. The theoretical literature has in particular focused on the auction mechanism behind these paid results (e.g., Edelman et al., 2007; Varian, 2007). Earlier studies took user behavior as given; more recent work by Chen and He (2011) as well as Athey and Ellison (2012) take into account that users search optimally. White (2013) and Xu et al. (2012) focus on trade-offs between organic and sponsored search results.

The empirical research on search engines has mostly focused on sponsored search as well. Yao and Mela (2011) develop a dynamic structural model of keyword advertising that takes optimal consumer behavior into account. Animesh et al. (2010) focus on quality uncertainty in sponsored search markets and find some evidence of adverse selection, but only for unregulated sponsored search markets. Ghose and Yang (2009) focus on ad placement and its effects on profitability and find a negative relationship between position and click-through rate as well as conversion rates. Agarwal et al. (2011) also find a negative relationship between position and click-through rates but find a positive relationship with conversion rates, which means that the top position is not necessarily the most profitable.[8]

Our paper is also related to three recent papers that focus on the relationship between sponsored and organic search results. Yang and Ghose (2010) find organic clicks to be positively related to the presence of sponsored links, and vice versa. However, the presence of an organic link increases the utility of a sponsored listing more than the other way around. Similarly, Agarwal et al. (2015) find the presence of a link in the organic search results to be positively related to the click-through rate for sponsored links, but negatively related to conversions. A third paper by Jerath et al. (2014) uses clicks data based on 120 keywords to examine how the "popularity" of different keywords impacts clicking behavior. Their results suggest that less popular keywords are "more targetable" for sponsored search advertising than more popular keywords.

---

8. Other contributions to the literature on sponsored search include Jeziorski and Segal (2015) and Blake et al. (2015), among others.

Finally, our research is related to a very large literature documenting the importance of screen position and a seller's reputation or brand equity [9] for retailers selling through other online channels including price comparison sites, shopbots, and auction sites (Brynjolfsson and Smith, 2000; Melnik and Alm, 2002; Dellarocas, 2003; Baye and Morgan, 2009; Baye et al., 2009; De los Santos et al., 2012). Although the broad message is that branding, screen position, consumer attributes, and retailer characteristics are all important determinants of click-through behavior in these channels, to date, little is known about their impact on organic clicks through search engines.

2. Data Description and Econometric Model

2.1. Data Description

Our main dataset is assembled using data from comScore Search Planner and contains information on the number of organic clicks web sites received for search terms and phrases entered at main search engines (i.e., Google, Bing, Yahoo, AOL, and ASK) during August 2012. Search Planner uses the comScore panel, which contains all online browsing activity of around two million U.S. users. Because our goal is to analyze the drivers of organic traffic following product searches, we restrict our sample to only include web sites that are Internet retailers. For this we make use of Internet Retailer's Top 500 Guide, which contains a ranking of North America's 500 largest e-retailers based on annual Web sales. Although not all of these retailers appeared in the comScore Search Planner database, some e-retailers (e.g., Amazon) operate multiple web sites (e.g., Amazon.com and Zappos.com), resulting in a total of 759 retail sites for which we have click-through data. For each of these 759 retailers, we used Search Planner to identify all search terms that generated traffic from Google to the retailer. There is some overlap in search terms: as shown in Figure 1, Onlineshoes.com as well as Zappos.com appear relatively high in the organic results for the term "shoes online," which means that for both retailers this term is part of the set of search terms that generated traffic from Google. In total we end up with 12,184 distinct search terms that led users to the 759 online retailers. The third dataset we use contains all the links that appeared on the first five search results pages on Google Search and Bing Search for each of the 12,184 search phrases. We collected these data using a scraper written in Java; the data contain organic search results as well as paid links.[10]

Not all 759 online retailers are relevant for each of the search terms in our data. For instance, Best Buy is not relevant for individuals searching for shoes and is therefore

9. Aaker (1991) defines brand equity as "a set of brand assets and liabilities linked to a brand, its name and symbol, that add to or subtract from the value provided by a product or service to a firm and/or to that firm's customers." Brand equity is based on factors like brand loyalty, name awareness, and quality. See Keller and Lehmann (2006) for a recent survey of the branding literature.

10. Given the time required to query Bing and Google with the 12,184 search terms, our Java program retrieved position results only once. Given the intervening time between clicks and the collection of rank data, two potential issues arise: (a) endogeneity of position, which we address in Section 2.2, and (b) the stability of ranks over time. The limited data we have suggest that ranks for product searches are likely to be fairly stable over the relatively short window of time that is at issue here. For example, an examination of rank data we collected for two different projects reveals that, for the 3,599 overlapping observations in these two datasets, the (pooled) correlation between each firm's rank for searches conducted in July 2012 with its rank for the searches conducted in September 2012 is 0.80. Although we do not have more frequent (e.g., daily or weekly) data on ranks, the relatively high correlation in ranks over this 2-month period suggests that ranks are likely to be reasonably stable within the span of a single month.

unlikely to show up in the search results for "shoes online." [11] Given that a retail site must be listed on the search results pages to receive organic clicks, we only include a retailer as an observation for a specific search term if we observe the retail site in our search results data. Because we only captured the first five pages of search results, this does not rule out that a site that did not appear in these search results did in fact get clicks; we therefore also include a retail site if the site received organic clicks for the search term according to the Search Planner database.

Our measure of a retail site's brand equity is based on the methodology developed in Baye et al. (2012) to overcome challenges in measuring the "prominence" of online retailers' names. These authors point out that the standard approach, which uses historical advertising data to measure accumulated brand equity and the strength of a firm's name, is not useful in the case of online retailers. Among other things, many online retailers are privately held and do not disclose advertising expenditures; the parent companies of publically traded online retailers do not typically report monthly advertising expenditures at the URL level. They further argue that the number of product searches on Google (or Bing) that include the retailer's name or URL in the search query—*navigational searches*, in industry parlance—is a useful measure of a retailer's brand equity. Intuitively, the inclusion of "Amazon" in a product search indicates the consumer recognizes the company and can recall its name. This may be the result of past advertising campaigns, recommendations from friends, knowledge of Amazon's pricing practices, product depth, return policies, and so on. And in contrast to advertising expenditures, it is possible to use data from comScore to measure the number of navigational searches Amazon received in a given month.

Based on a revealed preference argument, Baye et al. conclude that the number of navigational searches for a retailer in a given month embodies more than name recognition or recall; the inclusion of a retailer's name in a product search indicates that the consumer values the attributes associated with its name (e.g., its reputation, product breadth, product depth, and service quality), and that search results with links to that particular retailer are desirable. Baye et al. provide evidence that this measure works well in both retail and education contexts.[12] For example, they show that there is a strong relationship between navigational searches for universities and external rankings of their quality: Universities with stronger brand names (e.g., Harvard University) receive significantly more navigational searches than universities with weaker brand names (e.g., Indiana University). Moreover, they provide evidence that the prestige of a university—not its size—is the main driver of the navigational searches it receives.

For these reasons, we use navigational searches to measure a firm's brand equity. Navigational searches are essentially a shortcut for typing in the URL of a specific retailer and then searching its site. Thus, our measure of brand equity is the total number of organic clicks a particular retailer received in August 2012 from searchers who navigated

---

11. Indeed, Best Buy did not show up in at least the first 30 pages of search results on Google for the search term "shoes online" (checked on February 26, 2013).

12. Although many of the general branding principles carry over to retailers, the measurement of retail brand equity provides some unique challenges; Ailawadi and Keller (2004) identify some unique issues to the measurement of retail brand equity. For our purposes, the key is that this measure is related to the overall image of an online retailer and its attributes, which includes factors like name recognition, product breadth and depth, shopping experience, and reputation (prices, quality, shipping, return policy, etc.).

to its site by including the retailer's name as a search term, including misspellings; examples include "Amazon," "Amazn," and "Buy camera at Amazon."[13]

Although we use navigational searches to construct our measure of brand equity, our dependent variable excludes organic clicks from navigational searches. We seek to understand why searchers choose to click on Amazon (or some other link) following a non-navigational search like "Levis Jeans," and not why they click on an Amazon link following a navigational search like "Amazon." Thus, in our econometric analysis of organic clicks, we exclude all search terms (and hence organic clicks based on searches) containing the name of one of the 759 retail sites. This results in 40,117 observations, where each observation is the number of clicks for each search term-retailer.

In addition to the number of organic clicks per retail site, the Search Planner data also contain information on the demographics of searchers using each of the search terms, including the percentage of searchers by age, income, and location (home or work). We also used data from Internet Retailer's Top 500 Guide to identify each retailer's retail segment (e.g., mass merchant, apparel and accessories, sporting goods, etc.), whether the retailer has a presence on social media (Facebook or Twitter), the year in which the retail site began operating online, and whether the retailer has a brick-and-mortar presence. Table I provides summary statistics of these variables, as well as the other variables we use in our analysis.

Finally, we analyzed each search term and constructed search-term specific variables based on the content of the search term. The first variable is the *number of words* in the search term. The second variable, denoted *branded search term*, is an indicator variable for whether the search terms include the brand name of a *product* (e.g., Nike or Adidas) in the product search. Note that, in our sample, this is different from the brand associated with a particular retailer's site (e.g., Zappos or Amazon). These two search-term specific variables may tell us something about the intent of search. For instance, an individual searching for "Nike running shoes" is more specific in what she is looking for than someone searching for "shoes online," and this may affect clicking behavior.

2.2. Econometric Model

Our main objective is to study the drivers of organic clicks arising from searches for products on search engines.[14] Let $Clicks_{ik}$ denote the total number of organic clicks retailer i received from individuals searching for search term k. Because of the presence of substantial positive skewness in organic clicks data, we use a log-normal regression model to analyze the relationship between organic clicks and the explanatory variables, that is,

$$\ln(Clicks_{ik}) = \gamma_0 + \gamma_1 (RNO_{ik}) + \gamma_2 \ln(Rank_{ik}) + \gamma_3 Sponsored_{ik} + \gamma_4 \ln(BE_i)$$
$$+ \gamma_5 X_{ik} + \varepsilon_{ik}, \tag{1}$$

---

13. Baye et al. (2012) show that the utility of this measure is invariant to whether one uses a narrow definition of a navigational search (e.g., "Amazon") or this broader measure that includes both the retailer's name and product-related (e.g., "Buy camera at Amazon").

14. As a referee notes, clicks for some search terms may be more valuable (in terms of conversions, margins, or other performance measures) than others. Our focus here is on clicks; unfortunately, our data do not permit us to examine the differential effects of rank and other variables on these alternative measures of retailer performance.

<div align="center">

TABLE I.
DESCRIPTIVE STATISTICS ($N = 40,117$)

</div>

Variable	Mean	Std. Dev.	Min	Max
Google Related Data				
Organic clicks	2,039	10,809	0	752,296
Brand Equity (thousands)	3,412	5,746	0	18,000
Rank	20.480	17.564	1	52
Rank not observed	0.114	0.317	0	1
Sponsored Link on first page	0.311	0.463	0	1
Bing Related Data				
Rank	35.895	20.652	1	52
Rank not observed	0.574	0.494	0	1
Sponsored Link on first page	0.064	0.244	0	1
Alexa Related Data				
Sites Linking In (thousands)	186	395	0	1,183
Demographics				
Age 18–24	0.097	0.166	0	1
Age 25–34	0.176	0.220	0	1
Age 35–44	0.213	0.235	0	1
Age 45–54	0.252	0.249	0	1
Age 55–64	0.157	0.219	0	1
Age 65+	0.105	0.186	0	1
Income <25k	0.168	0.218	0	1
Income 25–50k	0.217	0.232	0	1
Income 50–75k	0.211	0.231	0	1
Income 75–100k	0.153	0.207	0	1
Income >100k	0.251	0.271	0	1
Home	0.792	0.278	0	1
Children	0.649	0.291	0	1
Search Term Characteristics				
Branded Search Term	0.141	0.348	0	1
Number of words	2.367	1.025	1	12
Retailer Characteristics				
Social network presence	0.944	0.230	0	1
Site age	13.710	2.827	2	24
Web only retailer	0.323	0.468	0	1
Retail Segment				
Apparel/accessories	0.217	0.412	0	1
Automotive parts/accessories	0.006	0.075	0	1
Books/music/video	0.020	0.139	0	1
Computers/electronics	0.128	0.334	0	1
Flowers/gifts	0.005	0.070	0	1
Food/drug	0.012	0.110	0	1
Hardware/home improvement	0.042	0.201	0	1
Health/beauty	0.022	0.148	0	1
Housewares/home furnishings	0.032	0.176	0	1
Jewelry	0.005	0.074	0	1
Mass merchant	0.363	0.481	0	1
Office supplies	0.019	0.136	0	1
Specialty/non-apparel	0.054	0.225	0	1
Sporting Goods	0.047	0.211	0	1
Toys/hobbies	0.028	0.166	0	1

where RNO_{ik} (short for *rank not observed*) is a dummy variable that equals 1 if retailer i is not observed on the first five pages of search results for search term k, $Rank_{ik}$ is the rank (or position) of retailer i on the first five pages of search results for search term k, $Sponsored_{ik}$ is a dummy for whether the retailer had a sponsored link on the first results page for search term k, BE_i a measure of retailer i's *brand equity*, and X_{ik} is a vector of other controls including demographic variables, search term specific variables, retailer characteristics as well as retail segment fixed effects.[15]

There are two primary concerns with estimating this equation: (i) it is likely that some of the explanatory variables are endogenous (correlated with ε_{ij}); and (ii) owing to the nature of the Search Planner data, we only observe the dependent variable in equation (1) when clicks exceed a certain threshold. Below we discuss how we deal with these concerns.

2.2.1. ENDOGENEITY

Google continuously updates its rankings of search results to generate the most relevant search results, which means that our rank variable will depend on past clicks. It is therefore likely that rank is correlated with the error term and thus endogenous. A similar effect may be at work for the ads variable: Ad positions are based on the outcome of a second-price auction that takes the relevance of the bidder with respect to the search term into account, again making it likely that ad positions are based on past clicking behavior on Google.

The standard approach in the literature on clicks at platforms (e.g., clicks at price comparison sites or sponsored clicks at search engines) is to assume that such positions are exogenous. Using the Wu–Hausman test for endogeneity, however, we reject the hypothesis that rank and ad positions are exogenous in our data ($p = 0.0023$ and $p = 0.0116$, respectively). To account for the potential endogeneity of these variables, we use information about rank and ads on Bing as instruments. These instruments are correlated with the endogenous regressors, but are unlikely to be correlated with the error term, because Bing's decisions on search result rankings and ad positions are not based on past clicks on Google. Indeed, using the Sargan test for overidentifying restrictions, we cannot reject the hypothesis that these are valid instruments for rank and ad positions on Google ($p = 0.3795$).

One might also worry that our measure of brand equity is correlated with the error in equation (1). For example, if users begin searching for products with generic search phrases and end their searches with a navigational search, navigational searches may be driven by organic search results, and thus endogenous. Based on the Wu–Hausman test, however, we cannot reject the hypothesis that our measure of brand equity is exogenous.[16] Our main results thus treat only position and ads as endogenous. Section 4 shows that our results are robust to the use of three alternative measures of brand equity that are also unlikely to be correlated with the errors.

2.2.2. SAMPLE SELECTION

As we explained in Section 2.1, a retail site is included as an observation if it appears on the first five pages of the Google search result page for a specific search term, independent

---

15. Retail segment fixed effects control for systematic differences in clicks across, for example, mass merchants who may receive many clicks owing to product breadth effects and specialty retailers who receive fewer clicks. Although this specification assumes the marginal impact of brand equity and rank is identical across retail segments, we show in Section 4 that the results are similar when one excludes mass merchants such as Amazon and Walmart.

16. The *p*-value for this test, which uses navigational searches on Bing as an instrument, is $p = 0.1692$.

of whether the retailer received organic clicks according to Search Planner. Complicating matters, Search Planner only reports the number of organic clicks if those clicks exceed a certain threshold, which means we do not know whether sites receiving zero organic clicks according to Search Planner really received no click-throughs for the search term in question or whether they were censored.

What makes our setting different from a standard censoring environment is that the selection rule depends on total clicks (including paid clicks) rather than just organic clicks. This means that a different probability mechanism generates both the zero clicks and the positive clicks, and this cannot be captured by a standard Tobit censoring model. For this reason, we estimate a Heckman-type selection model. As we argued in the previous subsection, endogeneity is likely to be important in our data, so we allow for endogenous explanatory variables. Estimation of the model consists of two stages. In the first stage we regress a dummy for having positive clicks on all exogenous variables (including instruments) z. Here, it is important to include at least one more instrument than is necessary for dealing with the endogeneity problem (otherwise identification is purely based on the parametric form of the inverse Mills ratio). This additional exclusion restriction should relate to the probability of observing positive organic clicks. Because this probability relates to the number of total clicks, we use additional variables in the selection equation that are important for getting paid clicks: We add dummies for whether a sponsored link was displayed on each of pages 2 through 5 in the Bing search results. We obtain the inverse Mills ratio, given by $\hat{\lambda} = \lambda(z\hat{\delta}) = \phi(z\hat{\delta})/\Phi(z\hat{\delta})$ from the selection equation, and add this to the second stage to obtain

$$\ln(Clicks_{ik}) = \gamma_0 + \gamma_1 (RNO_{ik}) + \gamma_2 \ln(Rank_{ik}) + \gamma_3 Sponsored_{ik} + \gamma_4 \ln(BE_i) + \gamma_5 X_{ik}$$
$$+ \gamma_6 \hat{\lambda} + \varepsilon_{ik} \tag{2}$$

We estimate this equation using the selected subsample (for which we observe organic clicks), that is, by two-stage least squares using instruments $(z, \hat{\lambda})$ for the endogenous variables RNO_{ik}, $Rank_{ik}$, and $Sponsored_{ik}$.

3. Results

Table II provides results for the specification in equation (2), which regresses the logarithm of organic clicks on explanatory variables that account for the impact on clicks of rank, brand equity, retailer characteristics, as well as characteristics including searcher demographics and the nature of search terms. Recall that these results control for potential endogeneity as well as censoring, and include a constant and retail segment fixed effects to account for potential differences in clicks across the 15 retail segments identified in Table I. All statistical tests are based on the reported robust standard errors, which account for potential heteroskedacity.

The estimated coefficient for the inverse Mills ratio is significantly different from zero at the 1% level, which indicates that it is indeed appropriate to control for censoring of the data. We discuss the other estimated parameters of the model below.

3.1. Rank

As discussed earlier, one potential goal of SEO is to increase the ranking (or position) of a retailer's links in organic search results. But just how important is position in driving a retailer's organic clicks following a product search? The first two estimated coefficients

Dependent variable: log of organic clicks on Google	
ln(Rank)	−1.347 (0.252)***
Rank not observed	−2.335 (0.616)***
Sponsored link on first page	0.317 (0.150)**
ln(Brand Equity)	0.084 (0.030)***
Age 18–24	−0.109 (0.124)
Age 25–34	0.088 (0.108)
Age 35–44	0.141 (0.103)
Age 55–64	0.223 (0.107)**
Age 65+	0.181 (0.132)
Income <25k	−0.708 (0.107)***
Income 25–50k	−0.211 (0.100)**
Income 75–100k	0.337 (0.111)***
Income >100k	0.773 (0.122)***
Home	−1.451 (0.149)***
Children	−0.191 (0.074)**
Branded search term	0.101 (0.057)*
Number of words	−0.138 (0.035)***
Social network presence	0.056 (0.144)
Site age	0.012 (0.009)
Web only retailer	0.127 (0.053)**
Inverse Mills ratio	0.851 (0.286)***
Observations	40,117
Censored	5,681
Uncensored	34,436
Selection	Yes
Endogeneity	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. All specifications include a constant and retail segment fixed effects.

in Table II provide an answer. The estimated coefficient for *rank not observed* captures the effect of a retailer's link not being included on the first five pages of search results for a given search term. The estimated coefficient of −2.335 is significant in both an economic and statistical sense, and implies that a firm not appearing on the first five pages receives 90% fewer clicks for a given search term. For a retailer whose link is observed on the first 5 pages, the estimated coefficient of −1.347 for *ln(Rank)* implies that a 1% decline in rank induces a 1.3% reduction in organic clicks for a given search term. For example, a retailer moving from the fifth to the sixth position in a search for "jeans" experiences a 27% reduction in organic clicks for that search term, whereas moving from the sixth to the seventh position results in a 22% decline.[17]

Although these results indicate that rank is a very important driver of organic clicks following product searches, it is important to stress that the unit of observation underlying these results is $Rank_{i,k}$; that is, retailer i's position in the results for search term k. Thus, these rank coefficients measure the effect of improving a retailer's position for a *single* search term. Consequently, SEO efforts that are term specific (e.g., designed to elevate a retailer's rank following a search for "jeans" but that have no effect on positions following other product searches), will result in a much smaller percentage improvement

17. We also ran specifications with position bins rather than the log-linear specification and the results were qualitatively similar.

in that retailer's *total* organic clicks. By way of example, the average retailer in our sample was relevant for about 60 search terms, so the corresponding effect on *total* organic clicks is about $1/60^{\text{th}}$ of the rank coefficients in Table II. For example, the estimated coefficient of -1.347 for *ln(Rank)* implies that a 1% improvement in rank following a given keyword search results in a 0.02% ($= 1.347/60$) increase in *total* organic clicks.

These results indicate that the returns to term-specific SEO critically depend on the breadth and depth of a retailer's product offerings and hence the number of search terms in which its link is relevant.

3.2. Retailer Brand Equity

Table II also reports estimates of the direct effect of a retailer's brand equity on the clicks it receives following a product search. The estimated coefficient for the logarithm of brand equity is positive and very precisely estimated, indicating that the direct effect of brand equity of a retailer's site is an important determinant of the organic clicks it receives following a product search. It is important to note that, unlike rank, brand equity is not search-term specific. As such, the estimated impact of brand equity in Table II captures the impact on a retailer's *total* organic clicks: Holding rank and the other factors influencing clicks constant, a 1% increase in a site's brand equity results in a 0.084% increase in a retailer's *total* organic clicks. These results, coupled with those discussed above for rank, indicate that a marginal improvement in a retailer's brand equity has a larger direct effect on its organic clicks than SEO efforts resulting in a marginal improvement in its position associated with a particular search term.

Notice that the estimated coefficient for brand equity in Table II, which corresponds to γ_4 in equation (2), measures the *direct* effect of retailer i's brand equity on its organic clicks. However, because search engines' algorithms determine rankings or positions of listings, in part, on past clicking behavior, there is also an *indirect* effect of brand equity on clicks: Retailers with greater brand equity and stronger brand names enjoy more clicks, which results in better future ranks. Figure 2, which graphs the average number of times retail sites appear on the first page of search results for different sextiles of brand equity, shows that online retailers with stronger brands in our data tend to have better ranks on results pages following non-navigational searches. The total effect of brand equity on clicks includes the direct effect identified in Table II as well as the indirect effect through rank.

We identify the total effect of brand equity on clicks by using a standard two-step procedure (Alwin and Hauser, 1975). In the first step we regress the rank variables on the logarithm of brand equity to obtain brand-equity adjusted ranks to determine the impact of brand equity on rank. In the second step, we proceed as in equation (2) but using these brand equity adjusted ranks. This regression yields an estimate of the total effect of brand equity on organic clicks, including both the direct effect shown in Table II as well as the indirect effect stemming from the impact of brand equity on position. As shown in Figure 3, the indirect effect of brand equity on clicks (through its impact on rank) is slightly larger than the direct effect, resulting in a total effect on organic clicks of 0.185—roughly twice the direct effect.[18]

---

18. Note that this procedure does not impact any of the other parameter estimates in Table II, nor does it impact the overall fit of the model. The estimated total effect is significant at the 1% level; the robust standard error for the point estimate of 0.185 is 0.047.
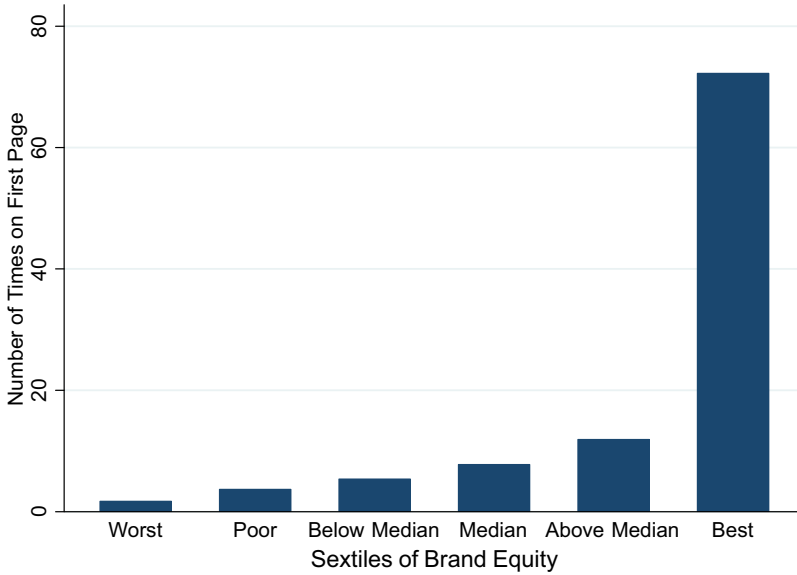
FIGURE 2. *PAGE RANK (POSITION) AND BRAND EQUITY*
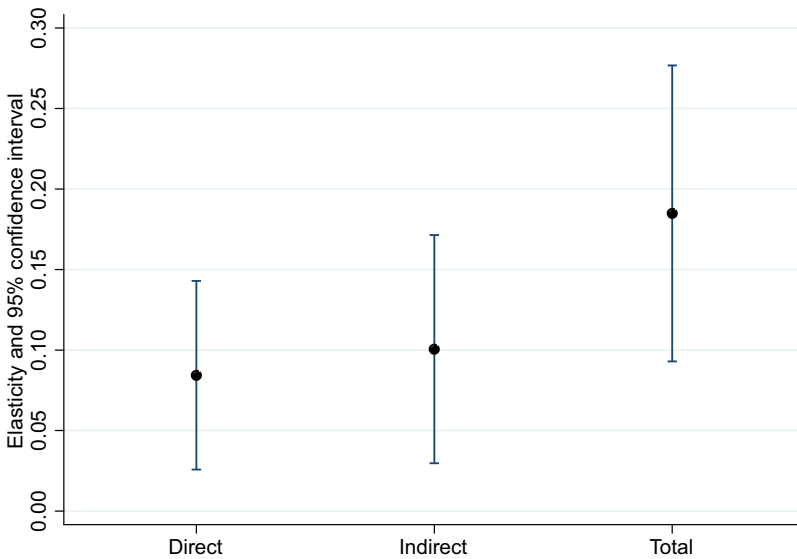
FIGURE 3. *ELASTICITY OF CLICKS WITH RESPECT TO BRAND EQUITY*

From the standpoint of SEO, these results highlight an important interaction be-
tween brand equity and rank. A 1% improvement in a retailer's brand equity directly
increases its *total* organic clicks by 0.084%, owing to the fact that consumers more fre-
quently click on its link instead of a competing one in the list of search results. Ultimately,
this induces search engines to elevate the firm's position in all relevant searches, which
results in an additional 0.101% increase in clicks. The total effect of a 1% improvement in
a retailer's brand equity is therefore a 0.185% increase in *total* organic clicks. Unlike the

impact of rank, this percentage increase applies to a retailer's overall clicks rather than the clicks stemming from a single search phrase. To illustrate, suppose a relatively unknown retailer gets 1,000 organic clicks from navigational searches and 30,000 through non-navigational searches.[19] If the retailer makes an investment in brand equity sufficient to increase its navigational searches by 10 clicks (1%), it gains an additional 55 (= 30, 000 × 0.185%) organic clicks from non-navigational searches.

3.3. Consumer Characteristics

The specification in Table II indicates that consumer characteristics systematically influence organic clicks on search engines. These results are of potential interest to retailers engaging in SEO to attract customers within particular demographic groups.

Notice that all of the income categories are statistically significant at the 5% or better levels: Consumers with higher incomes tend to more frequently click an organic link following a product search than do consumers with lower incomes. Although not all of the age categories are statistically significant, the general pattern suggests that younger individuals are less likely to click organic links than older individuals. Interestingly, the results also indicate that consumers searching from home are less likely to click following an organic search than individuals conducting a product search from the workplace. These patterns may stem from differences in search behavior across consumers with different demographic characteristics. For example, consumers with greater incomes may be more likely to conduct product searches on platforms such as Amazon rather than a search engine; individuals with lower incomes may be more likely to search using price comparison sites.

3.4. Keyword-Specific Effects

One might worry that the demographic effects documented above stem from differences in the sophistication of searchers with different demographic characteristics. To account for this, we include two controls for the nature of the keyword search: (1) *branded search term*, which is an indicator for whether the search phrase includes a brand-name product (e.g., "Levi's Jeans"), and (2) *number of words*, which is simply a count of the number of words included in the search. The results indicate that searchers who include specific brands of products in their terms, or who use fewer words in their search, are more likely to click an organic link following a product search. These findings are consistent with Yang and Ghose (2010), who find a positive relationship between branded searches and paid clicks as well as a positive relationship between keyword length and paid clicks. Our results are consistent with longer search phrases resulting in organic results that contain less relevant links, which would result in fewer organic clicks but more paid clicks.

3.5. Other Retailer Characteristics

In addition to retail segment fixed effects, the results in Table II include controls for other retailer characteristics that might impact SEO. We discuss each of these in turn.

---

19. To put this hypothetical in perspective, electricgeneratorsdirect.com received 1,208 organic clicks from navigational searches and 31,955 organic clicks through non-navigational searches in August of 2012.

First, note that retailers with a sponsored link on the first page of organic results receive 37% more organic clicks after controlling for rank, brand equity, and other drivers of clicks. This positive relationship is consistent with findings of Yang and Ghose (2010) and suggests that these sponsored links may provide searchers information about the retailer that increases the perceived value of clicking its organic link. For instance, such a link might lead searchers to conclude that the corresponding organic listing is relevant; alternatively, the sponsored link might have value as an advertisement that increases the brand equity of the retailer, making consumers more likely to click on organic as well as sponsored links. As with rank effects, however, *Ads* is a keyword specific variable so this 37% increase applies to the base of clicks from that keyword; it does not imply a 37% increase in overall organic clicks.

Second, the results in Table II indicate that Web-only retailers receive about 13% more total organic clicks than their bricks-and-clicks counterparts. This highlights that drivers of organic clicks through search engines may differ from those through other channels, such as price comparison sites. For example, Baye et al (2009) find that bricks-and-clicks retailers selling on a leading price comparison site receive over 25% more clicks than their Web-only counterparts.

Finally, notice that the specification in Table II also includes additional controls designed to capture other variables that potentially influence organic clicks. These include site age (a potential proxy for cumulative brand equity) and whether the site has a presence on social networks (Facebook and Twitter). Although the coefficients for these two controls have a positive effect on organic clicks, they are relatively small and not statistically significant at conventional levels. On balance, we view this as evidence that the effects discussed above are not the result of spurious correlation with excluded drivers of organic clicks.

4. Robustness Checks and Additional Results

In this section we demonstrate that our results are robust to a variety of alternative specifications, and offer some additional results that are of potential interest for SEO related to generating traffic from consumers in different income classes.

4.1. Results Based on Alternative Measures of Retailer Brand Equity

One may worry that our results stem from endogeneity issues related to our measure of brand equity. Although the Wu–Hausman test did not trigger any formal concerns about our measure of brand equity being correlated with the error in equation (2), one may wonder whether our results are sensitive to this particular measure of brand equity. Table III shows that our main findings are robust to using three alternative measures of brand equity that are unlikely to suffer from endogeneity concerns.

The first specification in Table III uses navigational searches on Google from June rather than August to construct the measure of retailer brand equity. Because navigational searches in June were predetermined at the time searchers made their August click decisions, this lagged measure of brand equity mitigates concerns that an unobserved factor drives both navigational and non-navigational clicks in the August clicks data. As shown in column (1), all parameter estimates, including the brand equity coefficient, increase slightly in magnitude but are qualitatively similar to those reported in Table II.

<div align="center">

TABLE III.
ROBUSTNESS: ALTERNATIVE MEASURES OF BRAND

</div>

	Dependent variable: log of organic clicks on Google		
	Lagged Brand (1)	Bing (2)	Brand Equity on Alexa (3)
ln(Rank)	−1.463 (0.263)***	−1.162 (0.227)***	−1.282 (0.237)***
Rank not observed	−2.569 (0.655)***	−1.910 (0.554)***	−2.154 (0.594)***
Sponsored Link on first page	0.422 (0.165)**	0.245 (0.138)*	0.404 (0.160)**
ln(Brand Equity)	0.091 (0.029)***	0.035 (0.017)**	0.124 (0.032)***
Age 18–24	−0.068 (0.131)	−0.151 (0.116)	−0.116 (0.121)
Age 25–34	0.117 (0.114)	0.066 (0.102)	0.084 (0.105)
Age 35–44	0.162 (0.107)	0.128 (0.097)	0.138 (0.101)
Age 55–64	0.209 (0.112)*	0.231 (0.101)**	0.229 (0.106)**
Age 65+	0.189 (0.140)	0.180 (0.124)	0.179 (0.131)
Income <25k	−0.728 (0.113)***	−0.674 (0.099)***	−0.689 (0.104)***
Income 25–50k	−0.231 (0.105)**	−0.199 (0.093)**	−0.205 (0.098)**
Income 75–100k	0.339 (0.116)***	0.331 (0.105)***	0.342 (0.110)***
Income >100k	0.805 (0.130)***	0.721 (0.114)***	0.757 (0.119)***
Home	−1.529 (0.159)***	−1.363 (0.138)***	−1.432 (0.147)***
Children	−0.187 (0.077)**	−0.206 (0.069)***	−0.201 (0.072)***
Branded search term	0.114 (0.060)*	0.093 (0.054)*	0.077 (0.057)
Number of words	−0.151 (0.037)***	−0.115 (0.032)***	−0.133 (0.034)***
Social network presence	0.132 (0.155)	0.030 (0.135)	−0.042 (0.141)
Site age	0.015 (0.009)	0.015 (0.009)*	0.000 (0.009)
Web only retailer	0.143 (0.056)**	0.117 (0.050)**	−0.041 (0.054)
Inverse Mills ratio	1.011 (0.308)***	0.626 (0.251)**	0.817 (0.278)***
Observations	40,117	40,117	40,117
Censored	5,681	5,681	5,681
Uncensored	34,436	34,436	34,436
Selection	Yes	Yes	Yes
Endogeneity	Yes	Yes	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. All specifications include a constant and retail segment fixed effects.

The second specification in Table III uses navigational searches from Bing rather than Google to measure brand equity. Because Bing has a different population of users and employs a different algorithm for returning search results, it is unlikely that unobserved factors that affect clicks on Google are correlated with this measure of brand equity based on navigational searches on Bing. The results in column (2) show that our findings are robust to using this alternative measure of brand equity.

The final specification in Table III is based on an alternative measure of brand equity pioneered by Animesh et al. (2010). This measure is constructed from data produced by the Web traffic reporting firm, Alexa, and measures the "Sites Linking In." It is based on the number of links to a web site from sites that are visited by individuals on Alexa's Web traffic panel.[20] Animesh et al. use these data to measure seller quality, noting that links pointing to a web site can be viewed as a positive recommendation from the referring site. As shown in column (3) of Table III, our results are also robust to using this alternative measure of brand equity—as well as to interpreting the brand equity effect identified in our earlier results as purely capturing "seller quality."

20. According to Alexa.com, "Links that were not seen by users in the Alexa traffic panel are not counted. Multiple links from the same site are only counted once." See also Alexa.com.

TABLE IV.
ROBUSTNESS: ALTERNATIVE CENSORING MODELS

	Dependent variable: log of organic clicks on Google	
	Tobit with IV	Tobit
ln(Rank)	−2.887 (0.118)***	−2.800 (0.040)***
Rank not observed	−5.544 (0.788)***	−4.199 (0.101)***
Sponsored Link on first page	0.381 (0.296)	−1.021 (0.064)***
ln(Brand Equity)	0.232 (0.019)***	0.275 (0.017)***
Age 18–24	0.775 (0.190)***	0.701 (0.190)***
Age 25–34	0.335 (0.156)**	0.332 (0.156)**
Age 35–44	−0.041 (0.149)	−0.001 (0.149)
Age 55–64	−0.106 (0.159)	−0.108 (0.159)
Age 65+	−0.498 (0.186)***	−0.456 (0.186)**
Income <25k	−0.606 (0.171)***	−0.620 (0.171)***
Income 25–50k	−0.319 (0.157)**	−0.281 (0.158)*
Income 75–100k	0.197 (0.168)	0.200 (0.168)
Income >100k	0.765 (0.143)***	0.750 (0.143)***
Home	−1.397 (0.110)***	−1.314 (0.110)***
Children	0.166 (0.104)	0.115 (0.104)
Branded search term	0.041 (0.085)	0.168 (0.083)**
Number of words	−0.202 (0.029)***	−0.175 (0.028)***
Social network presence	0.232 (0.173)	0.185 (0.169)
Site age	0.038 (0.013)***	0.034 (0.012)***
Web only retailer	0.457 (0.070)***	0.518 (0.069)***
Observations	40,117	40,117
Censored	5,681	5,681
Uncensored	34,436	34,436
Selection	No	No
Endogeneity	Yes	No

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. All specifications include a constant and retail segment fixed effects.

4.2. ALTERNATIVE CENSORING MODELS

Table IV shows that our main results are robust to using a Tobit censoring model rather than the Heckman selection model used in our main specification. The Tobit model can be interpreted as a constrained version of the selection model, with the selection and outcome equations being equivalent while not allowing for any selection bias. Column (1) reports results controlling for both selection and endogeneity, as in our main specification, whereas column (2) simply controls for selection. Comparing the parameter estimates to those in column (1) of Table II, most parameters increase in magnitude and are largely consistent with those reported in our main specification in Table II.

4.3. BRAND EQUITY AND CONSUMER INCOME

Our main specification in Table II assumes that the coefficients for the drivers of organic clicks are identical across consumers in different income groups. We conclude by showing that our qualitative results are not an artifact of pooling across searchers with different incomes.[21] These results are potentially of independent interest, because different retailers may use SEO to target consumers in different income groups.

---

21. We also ran specifications that did not pool across searchers of different ages or conducting searches from different locations, but those results did not materially differ from those presented in Table II.

Table V.
Unpooled Results by Income

	Dependent variable: log of organic clicks on Google (for income group)				
	Income <25k	Income 25-50k	Income 50-75k	Income 75-100k	Income >100k
ln(Rank)	−1.143 (0.212)***	−0.963 (0.243)***	−1.235 (0.238)***	−1.426 (0.222)***	−1.619 (0.244)***
Rank not observed	−1.887 (0.511)***	−1.166 (0.563)**	−1.818 (0.531)***	−1.877 (0.527)***	−2.526 (0.567)***
Sponsored Link on first page	0.309 (0.174)*	0.257 (0.187)	0.283 (0.195)	0.397 (0.210)*	0.589 (0.245)**
ln(Brand Equity)	0.071 (0.033)**	0.046 (0.037)	0.061 (0.036)*	0.100 (0.044)**	0.130 (0.047)***
Age 18–24	0.321 (0.165)*	0.139 (0.164)	0.090 (0.188)	−0.365 (0.234)	−1.240 (0.271)***
Age 25–34	0.381 (0.151)**	0.441 (0.143)***	0.192 (0.173)	−0.148 (0.212)	−0.596 (0.224)***
Age 35–44	0.305 (0.149)**	0.079 (0.144)	0.268 (0.165)	0.007 (0.206)	−0.026 (0.185)
Age 55–64	0.057 (0.158)	0.175 (0.153)	0.306 (0.172)*	−0.011 (0.203)	0.115 (0.202)
Age 65+	0.383 (0.183)**	0.083 (0.183)	0.174 (0.194)	−0.471 (0.237)**	0.013 (0.221)
Home	−0.030 (0.189)	−0.460 (0.215)**	−0.979 (0.216)***	−0.729 (0.259)***	−1.337 (0.267)***
Children	−0.182 (0.105)*	−0.221 (0.107)**	−0.220 (0.120)*	−0.542 (0.139)***	−0.414 (0.143)***
Branded search term	0.067 (0.070)	0.076 (0.067)	0.177 (0.073)**	0.182 (0.084)**	0.203 (0.089)**
Number of words	−0.157 (0.053)***	−0.085 (0.055)	−0.123 (0.056)**	−0.139 (0.070)**	−0.208 (0.079)***
Social network presence	0.092 (0.157)	0.023 (0.165)	−0.072 (0.179)	−0.059 (0.197)	0.062 (0.225)
Site age	0.010 (0.010)	0.013 (0.010)	0.011 (0.012)	0.005 (0.013)	0.012 (0.014)
Web only retailer	0.107 (0.064)*	0.079 (0.065)	0.104 (0.071)	0.224 (0.084)***	0.291 (0.096)***
Inverse Mills ratio	0.820 (0.290)***	0.456 (0.332)	0.730 (0.325)**	0.848 (0.364)**	1.323 (0.423)***
Observations	40,117	40,117	40,117	40,117	40,117
Censored	4,558	4,602	4,409	3,800	3,773
Uncensored	35,559	35,515	35,708	36,317	36,344
Selection	Yes	Yes	Yes	Yes	Yes
Endogeneity	Yes	Yes	Yes	Yes	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. All specifications include a constant and retail segment fixed effects.

Table V reports the results of estimation by income group, and shows that our main qualitative findings hold in the absence of pooling. Interestingly, however, these results suggest that brand equity has differential effects across individuals in different income classes.

For the three lowest income groups, the elasticity of organic clicks with respect to brand equity is smaller than the 0.084 reported in Table II based on pooled data, whereas for the top two income groups the elasticity is greater. Although one of the brand equity coefficients is not estimated precisely enough to infer that it is significantly different from the excluded ($51 to $74 thousand income) category, the results on balance indicate that brand equity is a more important driver of organic clicks for richer than poorer searchers. This result is illustrated in Figure 4, where the dots represent the point estimates for the elasticity of organic clicks with respect to brand equity for the five income groups, and the lines represent the corresponding 95% confidence intervals.

4.4. Relevance of Results for Specialty Retailers

Finally, one might worry that our main results are driven by the fact that large mass merchants (such as Amazon) have high levels of brand equity because of the breadth and
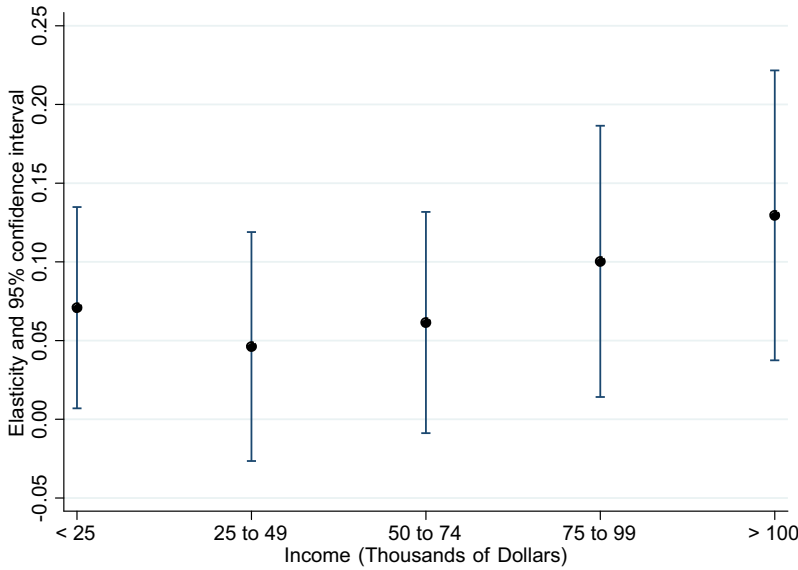
FIGURE 4. ELASTICITY OF CLICKS WITH RESPECT TO BRAND EQUITY BY INCOME OF SEARCHER

depth of their offerings, and also tend to receive large numbers of clicks because of this; that is, our regressions could simply be picking up "large retailer" effects. Although this concern is mitigated to some extent by the fact that all of our specifications include retail segment fixed effects, our specifications assume that the coefficients for brand equity, rank, and other variables are the same for mass merchants and specialty retailers.

      Table VI reports our main results (column 1) along with the results obtained when we exclude Amazon (column 2), Amazon–Walmart–Target (column 3), and all mass merchants (column 4) from the data. Notice that the estimated coefficients are remarkably similar across these different samples.

5. Managerial Implications for SEO

Our results are intuitive: When confronted with a list of potentially "relevant" search results, consumers are more likely to click the link of the retailer with the greatest brand equity. That is, holding other drivers of clicks constant, consumers tend to click retailers that are more recognized, trusted, have reputations for providing value (in terms of prices, product depth, or breadth), service (well-designed web sites, return policies, secure payment systems), and so on. Unlike price comparison sites and other online channels where signals of these attributes may be separately observed (through displays that include user feedback ratings, third-party certification, prices, shipping costs, etc.), the only signals consumers observe in organic product search results are sites' names (which embody their brand equity) and their "relevance" (as proxied by the rank or position that the search engine's algorithm assigns each organic link). We also showed that our findings are robust to several alternative specifications and, importantly, to controls for censoring as well as the endogeneity of a retailer's rank or position in the list of organic results. We conclude by discussing implications of our analysis for SEO, and by providing a few caveats regarding their implementation.

<div align="center">

TABLE VI.

ROBUSTNESS: MAIN SPECIFICATION FOR DIFFERENT RETAILER SAMPLES

</div>

	Dependent variable: log of organic clicks on Google			
	All Retailers (1)	Excluding Amazon (2)	Excluding Amazon, Walmart & Target (3)	Excluding All Mass Merchants (4)
ln(Rank)	−1.347 (0.252)***	−1.359 (0.289)***	−1.267 (0.293)***	−1.330 (0.290)***
Rank not observed	−2.335 (0.616)***	−2.489 (0.777)***	−2.449 (0.880)***	−2.491 (0.900)***
Sponsored link on first page	0.317 (0.150)**	0.189 (0.145)	0.106 (0.137)	0.029 (0.140)
ln(Brand Equity)	0.084 (0.030)***	0.074 (0.032)**	0.067 (0.032)**	0.061 (0.031)*
Age 18–24	−0.109 (0.124)	−0.129 (0.133)	−0.124 (0.133)	−0.041 (0.141)
Age 25–34	0.088 (0.108)	0.044 (0.121)	0.047 (0.122)	0.123 (0.129)
Age 35–44	0.141 (0.103)	0.136 (0.113)	0.159 (0.112)	0.148 (0.114)
Age 55–64	0.223 (0.107)**	0.261 (0.117)**	0.268 (0.118)**	0.401 (0.124)***
Age 65+	0.181 (0.132)	0.367 (0.147)**	0.395 (0.151)***	0.582 (0.163)***
Income <25k	−0.708 (0.107)***	−0.734 (0.118)***	−0.724 (0.118)***	−0.700 (0.118)***
Income 25–50k	−0.211 (0.100)**	−0.311 (0.112)***	−0.296 (0.110)***	−0.292 (0.117)**
Income 75-100k	0.337 (0.111)***	0.329 (0.121)***	0.304 (0.120)**	0.347 (0.127)***
Income>100k	0.773 (0.122)***	0.681 (0.135)***	0.709 (0.138)***	0.716 (0.142)***
Home	−1.451 (0.149)***	−1.410 (0.167)***	−1.325 (0.157)***	−1.343 (0.162)***
Children	−0.191 (0.074)**	−0.175 (0.079)**	−0.220 (0.080)***	−0.208 (0.085)**
Branded search term	0.101 (0.057)*	0.100 (0.064)	0.085 (0.066)	0.079 (0.068)
Number of words	−0.138 (0.035)***	−0.141 (0.041)***	−0.128 (0.041)***	−0.133 (0.040)***
Social network	0.056 (0.144)	0.029 (0.142)	0.017 (0.138)	−0.008 (0.140)
Site age	0.012 (0.009)	0.015 (0.010)	0.014 (0.010)	0.016 (0.011)
Web only retailer	0.127 (0.053)**	0.108 (0.062)*	0.094 (0.062)	0.145 (0.068)**
Inverse Mills ratio	0.851 (0.286)***	0.726 (0.305)**	0.622 (0.313)**	0.634 (0.305)**
Observations	40,117	35,384	31,513	25,572
Censored	5,681	4,371	3,957	3,569
Uncensored	34,436	31,013	27,556	22,003
Selection	Yes	Yes	Yes	Yes
Endogeneity	Yes	Yes	Yes	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Allspecifications include a constant and retail segment fixed effects.

5.1. RANK OR POSITION ON RESULTS PAGES

Our results indicate that rank is an important determinant of clicks; it is hard for a retailer to get organic clicks from a specific product search if its link is not observed on the first five pages of results for that search. For retailer's above this virtual "fold," the elasticity of clicks with respect to rank is about unity: a 1% improvement in rank leads to a 1% increase in organic clicks (for that search). Our results thus suggest that there are returns to SEO efforts that make it easier for search engines to determine a site's relevance for a particular product search. This includes making effective use of anchor texts, descriptive headings and meta tags, robot.txt files, and using accurate and unique page titles.

However, although these sorts of strategies for SEO are necessary and important, our analysis suggests that it would be a mistake to make them the exclusive focus of SEO. Rankings are a zero-sum game, and other retailers also have strong incentives to ensure that their sites contain the information needed to be properly indexed by search engines. In light of best responses by other sites, these efforts may prevent a retailer from losing ranks due to miscommunication with search engines, but are unlikely to result

in improvements in the equilibrium ranks of a particular retailer's link. Additionally, strategies designed to improve rank need not result in a long-run increase in organic clicks. This is particularly true of efforts to "trick" search engines into viewing a site to be relevant when it is not.

It is also important for managers to recognize that estimates of the impact of rank on organic clicks are keyword specific, and that apples-to-apples comparisons of the benefits of rank versus other drivers of clicks require an adjustment for the importance of that search in generating organic clicks relative to all relevant searches. Among other things, this means that the returns to focusing on improving rank may be larger for a niche retailer (which sells a single product on a site with a single page) than a mass merchant with thousands of products and pages.

5.2. Site Branding

The benefits of including brand equity as part of an SEO strategy are high. Such investments include increasing consumer awareness (through traditional as well as online advertising), making the site more user-friendly (easier to navigate), providing quality content and service (such as one-click purchases, easy return policies, and using a secure payment system), and more generally, enhancing the value of the brand that underlies the retailer's link. A number of retailers—including both Amazon and Walmart—have successfully used these strategies.

Investments in branding have both direct and indirect effects on organic clicks. The direct effect stems from our finding that consumers are more likely to click on links they know and trust—a finding that is consistent with evidence from other channels, including price comparison and auction sites. But brand equity has an equally sizeable indirect effect: Search engines want to provide users with relevant links, and the brand equity of a site is correlated with the relevance of links, which leads to better ranks and positions. Importantly, the brand equity of a site impacts organic clicks for all relevant keywords, not just those related to a particular search, so there is an amplification effect of SEO strategies targeted to improve the branding of a site.

In addition to spillovers on organic clicks related to searches for other keywords, investments that enhance brand equity are likely to lead to benefits in other channels. These benefits are not accounted for in our estimates, nor in the benefits that other papers document regarding the impact of brand and reputation on sponsored clicks. For example, our analysis focuses exclusively on drivers of non-navigational searches at search engines, so the regression coefficients in Table II do not include the benefits of increases in brand awareness or site quality on organic traffic from navigational searches at search engines. Likewise, improvements in a site's brand equity are likely to result in more direct visits to a retailer's site, as well as more clicks at other platforms including price comparison and auction sites. Finally, for retailers operating both online and physical stores, some investments (such as advertising) may result in positive spillovers into the physical channel.

Our analysis indicates not only that investments in brand equity lead to significantly more organic clicks, but also that these investments are more likely to be sustainable than SEO efforts focused entirely on rank. Additionally, such investments have spillover benefits in other channels as well, as has already been documented in extensive research on other online markets as well as traditional retail channels. For all of these reasons, we conclude that site quality, brand awareness, and other investments that

enhance the brand equity of an online retailer are important components of an overall SEO strategy.

5.3. Search Term Considerations

Some search terms and phrases are more likely to generate clicks than other keywords, even when one accounts for differences in the brand equity of different retailers and their ranks in search results. This is potentially relevant for SEO as well. For example, our finding that searchers including the specific brand of a product (e.g., "Levi's Jeans") in a search are more likely to click an organic link suggests that retailers selling branded products benefit by ensuring that their sites present information about the brands in their portfolio in a way that allows search engines to properly index them. Likewise, searchers using longer keywords are less likely to click an organic link, so parsimony in this regard is also important for SEO.

5.4. Demographic Considerations

Our findings that individuals that are older, have higher incomes, or who conduct product searches at work are more likely to click organic links also have ramifications for SEO. Among other things, these results suggest that SEO is more likely to be important for retailers targeting consumers with these demographic characteristics. In addition, because the elasticity of organic clicks with respect to brand equity is higher for individuals with higher incomes, the marginal benefits of SEO efforts targeted at improving the quality and brand awareness of a site are greater for retailers targeting individuals with higher incomes. More generally, the key implication is that the benefits of SEO vary, depending on the demographic characteristics of the consumers retailers are attempting to attract through this channel.

5.5. Retailer Considerations

The relationship between sponsored links and organic clicks identified in our data highlights yet another set of spillovers that complicates the calculus of SEO. Retailers attempting to increase traffic through organic links should recognize that there are possible spillovers from paid links: Consumers are more likely to click organic links associated with sponsored links. On the surface, this might seem like a pure win for retailers, because a sponsored link that results in an organic click rather than a paid click costs nothing. This is unlikely to be part of a sustainable strategy, however. Ultimately, if consumers click a retailer's organic rather than sponsored link, its prospects for winning that sponsored link in an auction will decline, because search engines have little incentive to allocate scarce ad space to retailers that do not receive sponsored clicks.

5.6. Concluding Caveats

Our analysis has focused on the potential benefits of SEO by focusing exclusively on the drivers of organic clicks. We have not taken into account the costs of improving these drivers, such as the costs of improving the meta tags associated with a particular keyword to improve rankings or advertising through traditional media to improve the brand awareness of a site. Costs are obviously an important component of optimization, and it would be a mistake to base SEO decisions purely on the drivers documented

above. Future research documenting the costs of different SEO strategies is therefore also important for the SEO literature.

It is also important to recognize that search engines are only one of many online platforms where consumers conduct product searches. Baye et al. (2013) note that in June 2012, consumers using browsers conducted 634 million product searches at retailer sites (such as Walmart.com), 134 million product searches at price comparison sites (such as Dealtime.com), and 877 million searches at marketplace sites (such as eBay.com). They also point out that 70% of eBay's listings were for new products, and over 60% of its listings were through posted prices rather than auctions. Unlike SEO efforts designed to improve rankings at a search engine, SEO efforts to improve a retailer's brand equity can improve the clicks it receives from searches in these other channels. Because these spillover benefits are difficult to quantify, it is easy for those engaging in SEO to underestimate the benefits of investing in the quality and brand awareness of a site.

REFERENCES

Aaker, D.A., 1991, *Managing Brand Equity: Capitalizing on the Value of a Brand Name*, New York, NY: The Free Press.

Agarwal, A., K. Hosanagar, and M.D. Smith, 2011, "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets," *Journal of Marketing Research*, 48, 1057–1073.

———, ———, and ———, 2015, "Do Organic Results Help or Hurt Sponsored Search Performance?," forthcoming in *Information Systems Research*.

Ailawadi, K.L., and K.L. Keller, 2004, "Understanding Retail Branding: Conceptual Insights and Research Priorities," *Journal of Retailing*, 80, 331–342.

Alwin, D., and R.M. Hauser, 1975, "The Decomposition of Effects in Path Analysis," *American Sociological Review*, 40, 37–47.

Animesh, A., V. Ramachandran, and S. Viswanathan, 2010, "Quality Uncertainty and the Performance of Online Sponsored Search Markets: An Empirical Investigation," *Information Systems Research*, 21, 190–201.

Athey, S., and G. Ellison, 2011, "Position Auction with Consumer Search," *The Quarterly Journal of Economics*, 126, 1213–1270.

Baye, M.R., B. De los Santos, and M.R. Wildenbeest, 2012 "What's in a Name? Measuring Prominence, and its Impact on Organic Traffic from Search Engines," forthcoming, *Information Economics and Policy*.

———, ———, and ———, 2013, "The Evolution of Product Search," *Journal of Law and Economic Policy*, 9, 201–221.

———, and J. Morgan, 2009, "Brand and Price Advertising in Online Markets," *Management Science*, 55, 1139–1151.

———, J. Rupert, J. Gatti, P. Kattuman, and J. Morgan, 2009, "Clicks, Discontinuities, and Firm Demand Online," *Journal of Economics & Management Strategy*, 18, 935–975.

Berman, R., and Z. Katona, 2013, "The Role of Search Engine Optimization in Search Marketing," *Marketing Science*, 32, 644–651.

Blake, T., C. Nosko, and S. Tadelis, 2015, "Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment," *Econometrica*, 83, 155–174.

Brynjolfsson, E., and M.D. Smith, 2000, "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 46, 563–585.

Chen, Y., and C. He, 2011, "Paid Placement: Advertising and Search on the Internet," *The Economic Journal*, 121, F309–F328.

Dellarocas, C., 2003, "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49, 1407–1424.

De los, Santos, B., A. Hortaçsu, and M.R. Wildenbeest, 2012, "Testing Model of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102, 2955–2980.

Edelman, B., M. Ostrovsky, and M. Schwarz, 2007, "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, 97, 242–259.

Ghose, A., and S. Yang, 2009, "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 55, 1605–1622.

Goldfarb, A., Q. Lu, and S. Moorthy, 2009, "Measuring Brand Value in an Equilibrium Framework," *Management Science*, 28, 69–86.

Jerath, K., L. Ma, and Y.-H. Park, 2014, "Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity," *Journal of Marketing Research*, 51, 480–486.

Jeziorski, P., and H. Segal, 2015, "What Makes Them Click: Empirical Analysis of Consumer Demand for Search Advertising," *American Economic Journal: Microeconomics*, 7, 24–53.

Keller, K.L., and D.R. Lehmann, 2006, "Brands and Branding: Research Findings and Future Priorities," *Marketing Science*, 25, 740–759.

Melnik, M.I., and J. Alm, 2002, "Does a Seller's Commerce Reputation Matter? Evidence from eBay Auctions," *The Journal of Industrial Economics*, 50, 337–349.

Sen, R., 2005. "Optimal Search Engine Marketing Strategy," *International Journal of Electronic Commerce*, 10, 9–25.

Smith, M.D., and E. Brynjolfsson, 2001, "Consumer Decision-Making at an Internet Shopbot: Brand Still Matters," *The Journal of Industrial Economics*, 49, 541–558.

Varian, H., 2007. "Position Auctions," *The Journal of Industrial Economics*, 25, 1163–1178.

White, A., 2013, "Search Engines: Left Side Quality versus Right Side Profits," *International Journal of Industrial Organization*, 31, 690–701.

Xing, B., and Z. Lin, 2006, "The Impact of Search Engine Optimization on Online Advertising Market," in *Proceedings of the 8th International Conference on Electronic Commerce*, 519–529, New York, NY: ACM.

Xu, L., J. Chen, and A. Whinston, 2012, "Effects of the Presence of Organic Listings in Search Advertising," *Information Systems Research*, 23, 1284–1302.

Yang, S., and A. Ghose, 2010, "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" *Marketing Science*, 29, 602–623.

Yao, S., and C.F. Mela, 2011, "A Dynamic Model of Sponsored Search Advertising," *Marketing Science*, 30, 447–468.