Online Advertising as Passive Search

*Current version: July 2022

Raluca M. Ursu[†] Andrey Simonov[‡] Eunkyung An[§]

Abstract

Standard search models assume that consumers actively decide on the order, identity, and number of products they search. We document that online, a large fraction of searches happen in a more passive manner, with consumers merely reacting to online advertisements that do not allow them to choose the timing or the identity of products to which they will be exposed. Using a clickstream panel data set capturing full URL addresses of websites consumers visit, we show how to detect whether a click is ad-initiated. We then report that in the apparel category ad-initiated clicks account for more than half of all website arrivals, are more concentrated early on in the consumer search process, and lead to less in-depth searches and fewer transactions, consistent with the passive nature of these searches. To account for these systematic differences between active and passive searches, we propose and estimate a simple model that accommodates both types of searches. Our results show that incorrectly treating all searches as active inflates the estimated value of brands that advertise frequently. Finally, we show that our model can more accurately recover data patterns, especially for advertising brands, and we explore two extensions of it, accounting for ad targeting and different forms of advertising.

Keywords: Sequential Search, Advertising, Online Browsing, Apparel Industry

JEL Classification: D83, L81, M31, M37

[*]Authors contributed equally and names are listed in reverse alphabetical order. We are thankful for comments from Simon Anderson, Anocha Aribarg, Yuxin Chen, Tülin Erdem, Anindya Ghose, Rafael Greminger, Yufeng Huang, Kinshuk Jerath, Pranav Jindal, Alex Moore, Eitan Muller, Oded Netzer, Stephan Seiler, Olivier Toubia, Qianyun (Poppy) Zhang, seminar participants at Berkeley, Michigan, UCL, and attendees of the IIOC 2021 conference, the 2021 Marketing Science conference, and the 2022 CEPR Applied IO conference (Berlin). We would also like to thank GfK and Siham El Kihal for helping us obtain the data. The usual disclaimer applies.

[†]Stern School of Business, New York University, rursu@stern.nyu.edu.

[‡]Columbia Business School and CEPR, as5443@gsb.columbia.edu.

[§]Stern School of Business, New York University, ean@stern.nyu.edu.

1 Introduction

Consumers increasingly use Internet platforms to search for the products they wish to buy. The ready availability of clickstream data tracking such decisions has generated an unprecedented level of interest in the study of search behavior (e.g., Kim et al., 2010; De los Santos et al., 2012; Koulayev, 2014; Bronnenberg et al., 2016; Chen and Yao, 2017; Ursu, 2018; Amano et al., 2022). These clickstream data reveal sequences of products searched, which – together with models of sequential search (e.g., Weitzman, 1979) – help researchers recover preference and search cost estimates and thereby inform marketing and economic decisions. A key assumption made in these sequential search models is that consumers *actively* decide on the order, identity, and number of all products they search.

At the same time, consumers use the Internet for a variety of other purposes, such as checking email, visiting social networking sites, or reading the news. While engaged in such activities, consumers may be exposed to advertisements from retailers. In fact, a large fraction of online searches are initiated by advertisements – U.S. companies spend more than half of their total advertising budgets ($129 billion in 2019) on online marketing strategies, such as paid search, email marketing, or display ads.[1] When exposed to ads, consumers may click or even buy the advertised products. This means that observations in typical clickstream data will contain both ad-initiated and organic clicks. Without data distinguishing between ad-initiated and organic clicks, prior empirical work modeling consumer search behavior has treated all clicks as organic and thus as resulting from an active search process, whereby consumers choose optimally which products to search and in what order (De los Santos et al., 2012; Koulayev, 2014; Chen and Yao, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020).[2] However, an important aspect of advertising is that consumers cannot choose which products they will be exposed to, but only how to react in response to ad exposure (e.g. whether to click on the ad).[3] Thus, it is natural to ask whether ad-initiated searches should be modeled in the same way as organic clicks or not (Blake et al., 2015). Different answers to this question lead to drastically different inferences about consumer preferences.

In this paper, we describe and model the nature of ad-initiated searches. We will do so by first distinguishing between organic and ad-initiated searches. Towards this end, we employ a detailed

[1] For more information, see emarketer.com/us-digital-ad-spending.

[2] Typical clickstream data only reveal the information consumers obtained once on a website (e.g. quality, price), but not how they landed on the website (actively or through an ad).

[3] Although consumers may not directly choose the types of ads they will see, ads may be targeted based on their search behavior or other observables. We account for the possibility of ad targeting in Section 8.1.

clickstream data set capturing all web traffic (8 million clicks) of a panel of 4,600 consumers in the Netherlands at the level of the exact URL address of a website visited. These data contain clicks in our focal category – apparel – as well as all other online activities consumers performed, such as checking email, visiting Facebook, or reading the news. A special feature of our data is the granularity of the URL addresses captured – we observe the entire URL, containing not only information on the website accessed (e.g. www.nike.com), but also the products viewed and the path the consumer took to land on the webpage. In particular, these exact URLs contain specific keywords that identify the advertiser, the medium of advertising (e.g. email, display, social media), and details of the ad campaign (e.g. fall/winter) in cases when the consumer landed on a webpage through an ad. Using these data, we develop a method for detecting ad-initiated clicks, and describe and separate them from searches occurring organically.

We then document the volume and describe the nature of ad-initiated searches. Product searches initiated by ads are extensive – 15% of all clicks and 53% of all website arrivals in the apparel category are a result of clicks on ads. Consumers are more likely to search through ads early on in their search process – the probability that a search is ad-initiated is 22% in the first decile of search and only 7% in the last decile. We also find that consumers rarely have a prior relationship with brands they visit through ads. Furthermore, ad-initiated clicks lead to lower quality website searches compared to organic clicks – these website visits involve searches of fewer and more expensive products, are shorter, and are less likely to result in a purchase. Finally, most ad-initiated searches occur when consumers are engaged in online activities that are unrelated to shopping (e.g. when checking email or visiting social media websites).

These patterns indicate that ad-initiated searches do not align with the active search behavior assumed by standard search models (e.g., Rothschild, 1974; Weitzman, 1979) and used frequently in empirical applications (Kim et al., 2010, 2017; Koulayev, 2014; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020). Such models assume consumers choose which options to search next, optimally searching products in decreasing order of an index (reservation utility) representing their expected utility net of search costs. In contrast, we find that ad-initiated searches occur predominantly early in the search process, but are generally of lower quality. Also, consumers frequently click on ads while engaged in shopping-unrelated activities, such as checking email or visiting social media websites, which are valuable activities *per se*. This means consumers are not actively seeking out product information in such settings. Consistent with this observation, a 2022 Bazaarvoice study of 14,000 consumers found that 61% of social media users buy products after "stumbling across" them in their newsfeed.[4] Finally, advertisers do not allow consumers to choose which products they will be exposed to. Based on this evidence, we propose that when search is ad-initiated, consumers may search in a more *passive* manner than assumed in standard search models – i.e. they choose how to react to information to which they were exposed, but do not optimally choose what information to see or in what order to see it.[5]

We propose a simple model that distinguishes between active and passive search decisions. The model builds on the canonical sequential search model of Weitzman (1979) of active search and combines it with insights from the theoretical framework proposed by Renault (2016), where all search is passive. Consistent with the Weitzman (1979) model, the consumer optimally ranks options she is aware of by their reservation utility and proceeds to search in that order, stopping to make a purchase decision when the best option she revealed while searching exceeds the reservation utility of any unsearched option. In addition, the consumer may be exposed through advertising to information about an option she is not aware of or is not considering (there are more than 1,000 websites in the apparel category in our data), a mechanism documented and commonly used in prior work (Goeree, 2008; Terui et al., 2011; Honka et al., 2017; Tsai and Honka, 2018). Consistent with Renault (2016) and our data patterns, we model consumers' search in response to ads as passive, i.e. consumers choose whether to search an advertised product, but cannot choose which ad to be exposed to. In our main specification (*AP-strong model*), we model consumers as choosing whether to click on the ad they were exposed to, comparing it with the best option observed so far. We also examine other variations of our model, including a version where consumers compare ads not only with the best option observed so far, but also to other unsearched options (*AP-weak model*), and a version where the advertised products are searched actively (using Weitzman optimal search rules), but may have different search costs.

In our empirical application, we demonstrate the better fit of our proposed models of active and passive search over models that treat all ad-initiated searches as active. We model consumers as searching across websites in the four largest apparel subcategories ("shirts, tops, and blouses",

[4]See socialmediatoday.com/news/.

[5]We share the terminology of "active" and "passive" search with Renault (2016) and Ghose and Todri-Adamopoulos (2016), defining active search as the effortful action to seek out product information optimally, and passive search as the reaction to information to which one is exposed. In both cases consumers choose whether to search a product. The difference is that, in contrast to active search, under passive search consumers do not choose which product to search next (i.e. the optimal search order). Similar ideas appear in Honka et al. (2017) and Morozov (2020). More details appear in Section 4.5.

“shoes”, “pants and jeans”, and “underwear”). We find that across all subcategories, our main model of active and passive search (AP-strong) has the best fit, followed by the second variation of the model (AP-weak) and the Weitzman model with different search costs for advertised options. The standard Weitzman model where all searches are treated as active and ads do not affect search costs leads to the worst data fit. This result highlights the different nature of ad-initiated searches and the importance of accounting for the role of advertising when modeling search decisions. Further, we find that treating all searches as active leads to biased estimates – consumer preferences for websites that advertise frequently are overestimated by the Weitzman model by 18% compared to our main model. This bias occurs because most ad-initiated searches happen early on in the search process, leading the Weitzman model to incorrectly assume that these options have high reservation utilities. In contrast, our model predicts that advertised websites will be clicked more early in the search process not because of their high reservation utilities, but because the consumer has not yet searched options with high enough utility. Finally, we show that the improvement in fit of our model comes primarily from lowering the prediction error for advertisers. For example, moving from the Weitzman model to the AP-strong model leads to a 49% (25%) decrease in the root mean squared error (RMSE) for searches (purchases) of advertising companies. In sum, the AP-strong model can more accurately recover data patterns, especially for advertised options.

Our goal in this paper is not to recover causal effects of firm advertisements on consumer decisions. There is a rich literature using experiments to identify such causal effects (e.g. Blake et al., 2015; Fong, 2017; Gordon et al., 2019; Sahni and Zhang, 2020). In fact, our data cannot be used to estimate the causal effect of ads for at least two reasons: (i) we do not have data from an experiment and (ii) we observe over 1,000 websites in our data, each potentially using different advertising strategies, that are unobserved to us. Rather, our goal is to describe and correctly model the consumer search process in the presence of ads. The models we present attempt to do so by testing different hypotheses on the search behavior of consumers when exposed to ads.

Nevertheless, we propose an extension of our main model that tries to account for the possibility of ad targeting. To account for targeting, we extend our model and allow a firm to advertise to specific consumers. More precisely, we model the probability of a consumer seeing an ad from a specific firm as a function of a score, i.e. its value to the firm. Firms vary in their targeting abilities and their knowledge of previous consumer behavior (e.g. previous search and purchase decisions). Thus, we vary the precision of the firm's information in order to account for as many ad targeting scenarios as possible. We find that the fit of this model is dominated by that of our main model specification (AP-strong model). We also find that among the models that account for targeting, those that assume firms' targeting is less precise fit better, further supporting the notion that our model that assumes little or no targeting fits the data better.

In a second extension of our main model, we allow ads to have not only an informative but also a persuasive effect, influencing consumer choices conditional on awareness. Towards this end, we estimate a model where advertising has a persuasive effect (Weitzman model with advertising affecting the purchase utility), as well as a model where advertising can have both an informative and a persuasive effect (AP-strong model with advertising affecting both awareness and the purchase utility). We find that advertising has a positive effect on utility; however, our main model (AP-strong model) fits the data better than a model where ads have a persuasive role. This result is further supported by our empirical patterns: (i) consumers rarely have a prior relationship with websites they search through ads and (ii) advertised options are rarely purchased.

This paper brings together the advertising and consumer search literatures by studying how ad-initiated clicks enter the consumer search process. Our results document the important role of ad-initiated searches – they represent the majority of website visits, occur predominantly early in the search process, but are unlikely to lead to a transaction. These results can help managers account for the different nature of ad-initiated and active clicks. As we have shown, incorrectly assuming that ad-initiated clicks are a result of an active search process may lead to biased parameter estimates, affecting managers' advertising decisions. For example, assuming a consumer has actively searched a product on Nike.com, rather than passively reacted to a Nike ad – even if it is for the same product – implies wrongly assuming the consumer expects Nike's product offerings to dominate those of other brands, inflating consumer brand preferences.

The rest of the paper is organized as follows. Section 2 discusses the related literature. We introduce our data in Section 3. Section 4 provides descriptive results on the nature of ad-initiated searches. Sections 5, 6, 7, and 8 introduce our model, estimation strategy, results, and model extensions, respectively. Section 9 discusses managerial implications and Section 10 concludes.

2 Related Literature

2.1 Consumer Search

Our paper relates and contributes to the literature on consumer search. Both theoretical (e.g., Stigler, 1961; Rothschild, 1974; Weitzman, 1979; Wolinsky, 1986; Anderson and Renault, 1999; Branco et al., 2012, 2016; Ke et al., 2016; Dukes and Liu, 2016; Ke and Villas-Boas, 2019) and empirical (e.g., Hong and Shum, 2006; Moraga-González and Wildenbeest, 2008; Ratchford, 2008; Kim et al., 2010, 2017; De los Santos et al., 2012; Seiler, 2013; Honka, 2014; Koulayev, 2014; Moraga-González et al., 2015; Bronnenberg et al., 2016; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020) branches of this literature consider consumers who actively decide what products to search, in what order to search them, and whether to purchase the best option found. In these models, consumers exert costly effort to seek out product information, and firms affect this search process only indirectly – for instance, through prices, product features, or product recommedations. Our first contribution to this literature consists of documenting that a large fraction of online consumer searches happen through ads – a channel that does not necessarily align with the assumed active nature of consumer search. We additionally contribute to this literature by developing a method to detect ad-initiated searches from clickstream data and by proposing a model where consumers search both actively and passively.

Our data patterns suggest that ad-initiated searches are unlikely to result from an active search process (see Section 4 for more details). To account for these searches, we build on earlier work that describes consumers' passive information acquisition processes, occurring for example through personal sources (e.g. friends, relatives, neighbors), unsponsored sources (customer reports), as well as marketing dominated sources, such as TV, newspaper, or radio ads, similar to our empirical application (Katona and Mueller, 1955; Bennett and Mandell, 1969; Newman and Staelin, 1972; Newman and Lockeman, 1975; Beales et al., 1981; Duncan and Olshavsky, 1982; Furse et al., 1984; Beatty and Smith, 1987; Shim and Drake, 1989). This literature also provides empirical observations on passive search – for instance, Beales et al. (1981) explain that passive search can lead consumers to gather different types of information (e.g. about the existence of a product, rather than about prices or other features) – but does not provide a theoretical formalization of a passive search process. To the best of our knowledge, Renault (2016) is the only paper that proposes a model of passive search, describing consumers who decide whether to click on an ad to obtain additional information about a product or whether to wait for another ad. In the model of Renault (2016), all search is passive; we make a contribution by combining it with the Weitzman (1979) model of active search, developing a model of joint active and passive search decisions. We then estimate our model and show that it outperforms one that treats all searches as active.

Finally, we note that our paper uses the same data as Ursu et al. (2021), but studies a different question. Namely, in Ursu et al. (2021) the focus is on understanding why consumers stop and restart their search across sessions. The authors propose that one mechanism affecting this decision is fatigue. Advertising may also explain why consumers restart their search, but it cannot explain why they frequently stop searching. Nevertheless, to be conservative, Ursu et al. (2021) use searches without clicks on the main advertising types. In contrast, understanding the broad role of online advertising is the focus of our paper.

2.2 Advertising and Consumer Search

Our paper also contributes to prior work on advertising and consumer search. On the theoretical side, papers in this literature consider models where either the only source of information consumers have access to arrives through advertising (Iyer et al., 2005), or where consumers can search in a second stage after receiving ads in a first stage (Butters, 1978; Robert and Stahl, 1993; Anderson and Renault, 2006, 2013; Mayzlin and Shin, 2011; De Corniere, 2016; Burguet and Petrikaite, 2017; Shin and Yu, 2021). Such two-stage models implicitly consider both the passive (first stage) and the active (second stage) nature of search. We contribute to this literature by proposing a model where information acquisition through both passive and active searches can happen throughout the entire decision-making process, not only in the beginning.

In our main model, advertising plays an informative role, consistent with prior work showing or proposing that the primary mechanism through which advertising affects the consumer search and choice process is awareness (e.g., Goeree, 2008; Terui et al., 2011; Honka et al., 2017; Tsai and Honka, 2018). Broadly, our paper fits into the literature documenting the informative effects of advertising (Ackerberg, 2001, 2003; Abhishek et al., 2012; Blake et al., 2015; Sahni and Zhang, 2020). The closest paper to ours is Honka et al. (2017), showing that advertising affects awareness, and that consumers engage in (active) search in the consideration stage (second stage). Unlike Honka et al. (2017), our model allows consumers to search actively and passively throughout the decision-making process.

On the empirical side, our paper relates to the rich literature on advertising effects on search (Yang and Ghose, 2010; Rutz et al., 2011; Goldfarb and Tucker, 2011; Rutz and Bucklin, 2012; Blake et al., 2015; Narayanan and Kalyanam, 2015; Jeziorski and Segal, 2015; Jeziorski and Moorthy, 2017; Golden and Horton, 2020; Fong, 2017; Rao and Simonov, 2018; Simonov et al., 2018; Du et al., 2019; Joo et al., 2013, 2016; Ghose and Todri-Adamopoulos, 2016; Sahni and Zhang, 2020; Simonov and Hill, 2021). Similar to the theoretical literature we discussed above, in these papers, search occurs only in a second stage after consumers were first exposed to advertising. In contrast, we consider the interplay between active and passive searches throughout the search process, and we develop a structural model of consumer search in the presence of advertising. Also, papers using field experiments typically partner with one company (e.g. Blake et al., 2015; Fong, 2017; Sahni and Zhang, 2020), while in our data we observe advertising and search decisions on more than 1,000 websites.

Our paper is also related to prior work on advertising attribution models (Abhishek et al., 2012; Li and Kannan, 2014; Kireyev et al., 2016; Chan et al., 2011). These papers propose methods (frequently Hidden Markov Models) to identify the impact of ads at different stages in the consumer conversion funnel, in order to measure the contribution of ads to the final purchase decision. For example, Abhishek et al. (2012) shows that display ads move consumers from a disengaged state to an awareness state, but not further towards a consideration state. Similarly, we model the effect of advertising on awareness. In contrast to this work, we focus on understanding the interaction of advertising and active search. We are able to do this by employing a rich data set of search across brands, capturing the entire browsing behavior of consumers.

3 Data

3.1 Data Description

In this section, we provide an overview of our data. The data were collected by GfK ("Growth from Knowledge"), the largest German market research company. Our data contain the complete PC browsing histories of an online panel of representative consumers from the Netherlands over the time period February 15, 2018 to May 1, 2018. We observe all search sessions with at least one click on an apparel website,[6] as well as all other browsing activity within the session, including visits to non-apparel websites (e.g. checking email, visiting social networks, using search engines, etc). We define a "spell" as all sessions of a consumer before she makes a transaction, or before our observation period ends.[7] An observation in our data is a URL address of a website clicked by the consumer, together with a time stamp for the visit, and consumer demographics (e.g. age, gender).[8]

The data contain 7,877,551 total clicks and 427,768 apparel clicks. There are 4,612 consumers observed to search across 5,649 spells, purchasing a total of 3,017 apparel products, with 76% of spells containing no purchased product.[9] As summarized in Table 1, on average, in a spell consumers visit 6 websites (of a total of 1,046 websites in the data), make 75 clicks, look at 23 products, and spend 40 minutes searching. There are a total of nine apparel subcategories which were classified as (ordered by total purchases): "shirts, tops and blouses", "shoes", "pants and jeans", "underwear", dresses and skirts", "children's clothes", "jackets and vests", and "accessories". Zalando and H&M are the most popular websites visited across all subcategories.[10]

========================

Insert Table 1 about here

========================

A key feature of our data is that we observe clicks at a very granular level – we know the exact URL address of a website visited by the consumer for each click. This allows us to identify and differentiate clicks that come through the online advertising channel from other clicks that occur organically. We describe our method for detecting searches that are ad-initiated next.

[6]Consistent with the industry standard, GfK groups all clicks that are not interrupted by a time period of inactivity longer than 30 minutes into a "session."

[7]We note that most (62%) spells without a transaction end (have the last session) more than a week before the end of our observation period.

[8]Consistent with prior work on consumer search that utilizes a clickstream data set, we will treat a "click" as a "search" decision (e.g., De los Santos et al., 2012; Koulayev, 2014; Bronnenberg et al., 2016; Chen and Yao, 2017; Ma, 2016; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020). However, we will further differentiate between two types of searches: organic and ad-initiated.

[9]GfK coded the transaction funnel, identifying a website visit, a product view, a basket addition, a checkout, or an order confirmation. This allows us to determine not only search behavior, but also consumers' purchases. The last product searched on a website that has an order confirmation will be marked as purchased. If several products are added to a cart, which is followed by an order confirmation, then all of them will be marked as purchased.

[10]We refer the reader to (Ursu et al., 2021) for further data descriptions. Appendix 11.1 contains details on additional data cleaning steps we performed.

3.2 Detecting Ad-Initiated Searches

We exploit the richness of the data contained in URL addresses to identify ad-initiated searches.[11] There are five main components of a URL, as illustrated in the following example:

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.example.com}}_{\text{website}}\underbrace{\text{/path/index.html}}_{\text{path}}\underbrace{\text{?param=value}}_{\text{parameter}}\underbrace{\text{\#myfragment}}_{\text{fragment}}$$

The uniform resource identifier (URI) scheme gives the http(s) communication protocol. Then, the URL identifies the website visited and a hierarchical path representing different pages and subpages of the website. For example, the path will be empty if the consumer accesses the homepage of the website, but accessing a category page or a product page will populate this component. The parameter component describes the specifics of the last path element identified, and any other information will be stored in the fragment component.

We detect ad-initiated clicks using the parameter component of the URLs. Ad-initiated clicks can be identified in this way because companies that are advertising need to modify URLs to track their advertising campaigns, as well as pay websites for the traffic they refer through ads.[12] The most common URL parameters that identify advertisers contain a series of "UTM" (Urchin Tracking Module) keywords, which contain information about the advertiser, the medium of advertising (e.g. email, display, social media), the ad agency (if any), and other specifics of the ad campaign.[13] The following is an example URL with such UTM parameters highlighted in blue:

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.missetam.nl}}_{\text{website}}\underbrace{\text{/sale/page/906/}}_{\text{path}}$$
$$\underbrace{\text{?utm\_source=copernica\&utm\_medium=email\&utm\_campaign=20180315\_week11}}_{\text{parameter}}$$

In addition, we can identify clicks generated by online advertisements based on unique tracking parameters that ad platforms automatically add to URLs to track ad traffic. For example, "gclid",

[11]We parsed the URLs using the R package called urltools (cran.r-project.org/web/packages/urltools/urltools.pdf). For more information, see docs.python.org/3/library/urllib.parse.

[12]In Appendix 11.2 we document in more detail how companies modify their URLs to track ads online. We note that there are at least two reasons for which our method may undercount the number of ads consumers search. First, consumers may be exposed to ads offline (e.g. on TV) and later search for those products online, a phenomenon that is well-documented (e.g., Joo et al., 2013, 2016). Such searches will be classified as active by our method. Second, a consumer may be exposed to an online ad, not click on it, and later return to search it on her own. Similarly, such searches will be classified as active.

[13]UTMs were developed by Urchin Software Corporation and are used by many ad platforms, such as Google Analytics, Facebook and Microsoft Advertising. For more information on UTMs, see Appendix 11.2, as well as wikipedia.org/wiki/UTM_parameters and ga-dev-tools.appspot.com/campaign-url-builder.

“gclsrc” and “dclid’ are Google ad click identifiers, “msclkid” is a Microsoft ad click identifier, while “fbclic” is a Facebook ad click identifier. An example of such a URL is given below:

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.hunkemoller.nl}}_{\text{website}}\underbrace{\text{/nl\_nl/}}_{\text{path}}\underbrace{\text{?gclid=eaiaiqobchmisky9\&dclid=cpfiv9rfh9ocfuc}}_{\text{parameter}}$$

The rest of the ad-initiated clicks include custom tracking parameters like “track_id” and “affid”, or URLs that include affiliate advertisers’ names or a specific advertising medium.

If a consumer navigates to a website organically the parameter component in the landing URL will not contain advertiser-related information. Instead, it can be empty, or contain information related to the consumer’s search for product information (e.g. her query or her sorting and filtering options used). Below are a few examples of such URLs:

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www2.hm.com}}_{\text{website}}\underbrace{\text{/dames/jassen.html}}_{\text{path}}$$

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.plutosport.nl}}_{\text{website}}\underbrace{\text{/dames/zwemmen}}_{\text{path}}\underbrace{\text{?filter\_size=46\&manufacturer=speedo}}_{\text{parameter}}$$

4 The Role of Online Advertising in the Consumer Search Process

4.1 How Extensive are Ad-Initiated Clicks?

Ad-initiated searches correspond to a substantial fraction of overall website visits in the apparel category. Among all the search-related website visits, 15% are clicks on ads (see Table 2).[14] Moreover, ads initiate the majority of website visits – 53% of first arrivals to a website are through clicks on ads (Table 2).[15] This higher percentage is due to the vast majority of within-website clicks occurring organically – not surprisingly, once on a website, consumers tend to navigate from page to page through links that are not sponsored. Also, consumers rarely have a prior relationship with brands

[14] The percentage we report should be distinguished from ad click through rates which are much smaller (e.g. on the order of 3% for Google ads: smartinsights.com/internet-advertising/internet-advertising-analytics) and which are not conditional on a click as in our case.

[15] Here and throughout the paper, website visits are unique website-session combinations. If instead we considered unique website-spell combinations (thereby ignoring revisits), we would similarly find that 51% of website arrivals occurred through ads.

they visit through ads: of all consumer-website pairs with at least one ad click, 90% of first visits are ad-initiated.[16] Even after we exclude users with visits in the first week of our data (to control for potential left truncation bias) or after considering only websites with at least two clicks, we find that the majority of first visits are ad-initiated (92% and 76%, respectively).

As described in the previous section, URLs frequently also include information on the specific ad campaign ran. Using this information, we were able to classify ad clicks into two types of campaigns: (1) containing season/calendar-related keywords and (2) containing category-related keywords.[17] Across the top 10 websites in our data, an average of 75 (15) consumers clicked on the same type 1 (2) campaign ad. Also, across the top 10 websites in our data, the average length of a type 1 (2) campaign was 3 (15) days. In other words, many consumers over several days saw and clicked on the same ads. These patterns suggest that at least such ads where not targeted to an individual consumer, but rather where part of campaigns that ran in a certain time period or for a certain category, so many consumers saw the same ad.

4.2 When do Consumers Click on Ads?

Consumers are much more likely to click on ads early on in their search process. To illustrate this relationship, in Figure 1a we divide search spells into deciles and compute the average percentage of ad-initiated clicks in each decile.[18] We find that the share of ad-initiated clicks declines as search advances, such that for the first decile the share of ad-initiated searches is 22%, while for the last decile it is only 7%. This relationship is in large part due to the shorter within-website searches early on in the search process; the share of ad-initiated website visits is more stable throughout the spell, though still trending downwards towards the end of the search process, as shown in Figure 1b.

========================

Insert Figure 1 about here

========================

[16] Note that the 53% of first website arrivals that occur through ads include websites that never advertise.

[17] Details on how we classified such ads can be found in Appendix 11.4.

[18] To split searches into deciles, we follow the method used in (Bronnenberg et al., 2016). More precisely, if t denotes the search under consideration and N_i denotes the number of total searches performed in a spell i, then deciles are defined as $d(t, N_i) = ceil\left(\frac{10(t-r(0,1))}{N_i-1}\right)$, where $r(0,1)$ is a draw from a uniform distribution on the interval (0,1). Our results are robust to dividing searches into fewer or more than 10 groups, and to conditioning on spells with at least 3, 5, and 10 searches.

4.3 How Do Ad-Initiated Searches Compare to Organic Searches?

Comparing the behavior of consumers on websites after arriving either through the ad or the organic channels, we find that the quality of ad-initiated website searches is generally lower. Table 3 summarizes differences between the two types of visits; ad-initiated website visits have on average fewer clicks (4.1 vs. 8.37 clicks), are shorter (2.28 vs. 4.32 minutes), and involve fewer products searched (0.58 vs. 1.28 products).[19] As a result, ad-initiated website visits are responsible for only 29% of transactions, while comprising 53% of website visits. Furthermore, in ad-initiated searches, consumers are exposed to more expensive products.[20]

========================

Insert Table 3 about here

========================

There is substantial heterogeneity across consumers and spells in the percent of their searches that are ad-initiated. Figure 2 describes search behavior for quantiles of consumers grouped by the share of searches they perform through the advertising channel. In the median (average) spell in the data, 13% (28%) of clicks are on ads, with 0% ads in the lowest quantile and 85% in the highest quantile (see Figure 2, sixth panel). Consistent with the lower quality of ad-initiated website visits, consumers who heavily rely on searches through the advertising channel tend to search fewer products and make fewer transactions. Consumers in the fifth quantile search predominantly through ads, but only make an average of 24 clicks per spell, with 1.7% of them having a transaction. In contrast, consumers in the second quantile, who click on a small number of ads (4%), have the longest search spells (153 clicks) and 43% of them have a transaction. Search intensity monotonically decreases from the second to the fifth consumer quantile, but it is low in the first quantile – these consumers almost never click on ads and seem to have a clear idea of the products they are looking for given their low search but high purchase frequencies.

========================

Insert Figure 2 about here

========================

[19] All these differences are statistically significant at least at the 5% level, with Table 3 reporting t-statistics.

[20] Since prices vary across apparel subcategories, we present in Table 3 the standardized price normalized by subtracting the average price and dividing by the standard deviation of the price in each subcategory.

4.4 Shopping-Unrelated Activities and Ad Engagement

Using complete URL addresses, we can further classify the types of online ads that bring consumers to a website. Our results appear in Figure 3. We identify eight types of advertisements, ordered by their click frequency: affiliate (45%) (third party ads found on newspapers, blogs, etc), email (24%), search engine (16%), newsletter (6%), other (5%), social (2%), display (2%), retargeting (< 1%).[21,22] Classifying ad types reveals that consumers were engaged in shopping-unrelated activities when they were exposed to ads – they were reading the news, checking email, or visiting social media websites.

We note that clicks induced by search engine advertisements may be classified as active searches, especially if consumers use the search engine to navigate to a specific website. Consistent with this idea we show in the second panel of Figure 3 that search engine ads account for the majority of ad purchases. We thus account for this possibility when estimating our model.[23]

==========================

Insert Figure 3 about here

==========================

In sum, we have shown that product searches arriving through the advertising channel play an important role in the consumer search process – ad-initiated searches represent the majority of website visits, and thus need to be accounted for. In addition, these searches are systematically different from those occurring organically – they occur predominantly early in the search process, have a lower intensity of search after landing on a website, a higher propensity to click on more expensive products, and a higher propensity to leave without purchasing.

4.5 Relating Online Advertising with Passive Searches

The descriptive evidence presented above suggests that ad-initiated searches are unlikely to result from the active search process assumed by standard consumer search models (e.g., Rothschild, 1974; Weitzman, 1979) and used frequently in empirical applications (Kim et al., 2010, 2017; Koulayev, 2014; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020). Active searches are defined as effortful actions to

[21] Appendix 11.3 provides more details on how we classified ads by medium.

[22] Retargeting ads may represent such a small fraction of all ad clicks because of their lower effectiveness (see Lambrecht and Tucker, 2013).

[23] More details can be found in Section 6 and Appendix 11.10.

seek out information about products relevant for a purchase decision. Such active search decisions involve determining which options to search, in what order to search them, and when to stop searching to make a purchase decision. We have shown that searches through online advertisements happen early on in consumers' search processes, but represent visits to lower quality websites, signaled by a lower intensity of searches on these websites, higher prices checked, and a lower likelihood to purchase. These patterns contradict the optimal search rules in Weitzman (1979), according to which consumers should sample the highest reservation utility options first. Also, consumers frequently click on ads while engaged in shopping-unrelated activities, such as checking email or visiting social media websites. This means consumers are not actively seeking out product information in such settings. Thus, the fact that consumers click on ads when engaged in such shopping-unrelated activities is inconsistent with the notion of active search.

Instead, these patterns suggest that consumers may be searching in a more passive manner when exposed to ads. Passive search describes consumer behavior that does not involve optimally choosing what information to see or in what order to see it, (Ghose and Todri-Adamopoulos, 2016; Renault, 2016; Honka et al., 2017; Morozov, 2020). Rather, under passive search, consumers merely react to information to which they are exposed, i.e. they choose whether to search the advertised product, but not which ad to be exposed to.[24] A model of passive search can explain why advertised websites will be clicked earlier in the search process, despite their lower relative quality: consumers have not yet searched options with high enough utility, since advertised products are not necessarily searched in an optimal order. Also, ads may expose consumers to products they were previously unaware of or have not considered – and thus have not computed a reservation utility – consistent with the evidence showing that consumers rarely have a prior relationship with brands they search through ads.

In the next section, we use these insights to formalize the relation between advertising and passive searches.

5 Model

In the canonical sequential search model of Weitzman (1979), each search occasion the consumer decides whether to continue searching, in which case she chooses a product to search, or whether to

[24] We note that the phenomenon we wish to capture is different than one in (Ursu, 2018) where consumers search after seeing an ordered ranking of products. The difference is that in those settings consumers still choose which products to search, i.e. search is active.

stop searching, in which case she decides which product to purchase, if any. We refer to this type of search action as "active", since the consumer determines which product to search if search continues. In contrast, in a model of "passive" search, such as that of Renault (2016), consumers are assumed to search in response to firms' advertising, and thus to not be able to choose what product to search next (i.e. cannot choose their search order optimally). Instead, consumers observe an ad for a product and decide whether to obtain more information about it by searching. In what follows, we develop a model of sequential search where consumers make joint active and passive search decisions.

5.1 Setup: The Joint Active and Passive Search Model

Consider a consumer who is in the market for at most one unit of a product in a given product category. This consumer is aware of and is considering towards her next purchase options $j \in J$. The consumer is uncertain about the options available to her, but may resolve that uncertainty by searching. Searching is costly, $c_j > 0$, but reveals a potential payoff u_j drawn from a distribution function $F_j(\cdot)$ with support $[-\infty, \infty]$. At each decision point, the consumer has searched a set of products S, and a set $\bar{S}$ is available to search, where $S \cup \bar{S} = J$. Let the maximum reward observed among the searched options be given by $y = \max_{j \in S \cup \{0\}} u_j$, where $j = 0$ denotes the outside option of not purchasing. At the end of the search process, the consumer may choose to purchase one of the options searched, or may choose the outside option. This consumer solves the following problem (due to Weitzman (1979))[25]

$$V(\bar{S}, y) = \max_{stop,continue} \{y, \max_{j \in \bar{S}} -c_j + W_j(\bar{S}, y)\}, \tag{1}$$

where $V(\emptyset, y) = y$ and the continuation value $W_j(\cdot)$ for $j \in \bar{S}$ is given by

$$W_j(\bar{S}, y) = V(\bar{S} \setminus j, y) F_j(y) + \int_y^\infty V(\bar{S} \setminus j, u) dF_j(u). \tag{2}$$

In words, at a given moment in the search process, the state space describing the problem of the consumer is given by the set of options she is aware of and is considering, but has not yet searched, $\bar{S}$, and by the best option revealed so far, y. At that moment, the consumer may decide to stop searching and choose the best option revealed so far, y. Alternatively, the consumer may choose to continue searching, in which case she searches one of the options in $\bar{S}$.

In addition to the J options a consumer is aware of and is considering, advertisers may inform the

[25]We assume no time discounting, consistent with prior empirical work, (Kim et al., 2010, 2017; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018; Ursu et al., 2020).

consumer about options $a \in A$. Options in set A are either products the consumer is unaware of or they may be options the consumer is aware of, but does not freely recall/consider in a given search setting, for example, because the product category is large (as in our empirical application, where there are more than 1,000 websites available).[26] Otherwise, the consumer may search these options actively. Our modeling choice is consistent with the empirical evidence presented in Section 4.1, as well as with prior work documenting and proposing that the primary mechanism through which advertising affects the consumer search process is awareness (Goeree, 2008; Terui et al., 2011; Honka et al., 2017; Tsai and Honka, 2018), and with the literature showing the informative effects of advertising more broadly (Ackerberg, 2001, 2003; Abhishek et al., 2012; Blake et al., 2015; Sahni and Zhang, 2020). When a consumer is exposed to an ad from a, her choice is

$$V(\bar{S} \cup a, y) = \max_{stop,continue} \{y, \max_{k \in \bar{S} \cup a} -c_k + W_k(\bar{S} \cup a, y)\}, \tag{3}$$

mirroring equation 1 with a now in the awareness set. We follow Renault (2016) and assume the consumer is passive in her reaction to ads. That is, the consumer does not have control over the probability of observing an ad, the identity of the advertiser, or the timing of the ad (in contrast to active search where she is assumed to choose what product to search next). Thus, the consumer does not change her search process in anticipation of the arrival of ads. However, when exposed to the ad, the consumer decides how to react to it. This is one variation of our model of active and passive search, to which we will refer as the **AP-weak model**.

While the AP-weak model captures the lack of control of consumers over the arrival of advertised product options, it still assumes that consumers compare a to the rest of the products not yet searched, $\bar{S}$. However, the consumer might be engaged in other online activities when exposed to ads (e.g. checking email or visiting social networking websites as we showed in Section 4.4) and may not consider other options at that moment (Renault, 2016); or the ad might focus the consumer's attention and limit her ability to process information about other options due to cognitive constraints (Gossner et al., 2020). To account for these possibilities, we propose a stronger version of the active and passive search model, in which the consumer does not compare the advertised option to other unsearched options when exposed to the ad, and instead compares the ad only with the best option searched so

[26] A longer discussion on the distinction between recall and recognition and their role in human judgments can be found in Lynch and Srull (1982).

far, y, as per

$$V(a, y) = \max_{\text{not } a,\, a} \{y, -c_a + W_a(a, y)\}. \tag{4}$$

Since ads are not anticipated, we assume the consumer can always continue the process (solve equation 1) after deciding whether to search ad a or not. We refer to this model as the **AP-strong model**.[27]

Our model makes a number of assumptions, notably that advertising works by affecting consumers' awareness and that consumers do not have control over the types of ads they are exposed to. We relax these assumptions in later sections. More precisely, in Section 8 we also test for a persuasive effect of ads and we extend our model to allow for ad targeting, where advertising decisions may be (partially) revealing of consumer preferences. Both extensions do not affect the main conclusion of our paper, i.e. that the active and passive search models presented in this section fit the data better than all alternatives considered.

Related to the above, we wish to highlight an important aspect of our model. Our goal in this paper is not to recover causal effects of firm advertisements on consumer decisions, question which has been studied extensively (e.g. Blake et al., 2015; Fong, 2017; Gordon et al., 2019; Sahni and Zhang, 2020). In fact, our data cannot be used to estimate the causal effect of ads for at least two reasons: (i) we do not have data from an experiment and (ii) we observe over 1,000 websites in our data, each potentially using different advertising strategies, that are unobserved to us. Rather, our goal is to correctly model the consumer search process in the presence of ads. The models we presented in this section attempt to do so by testing different hypotheses on the search behavior of consumers when exposed to ads. In the extensions we present in Section 8 below, we propose a method that attempts to additionally account for ad targeting.

5.2 Search Rules

Having laid out the primitives of the joint active and passive search model, we now describe the optimal search rules.

In the absence of ads, the AP-weak and the AP-strong models coincide with the Weitzman model. For this problem, the optimal search strategy is given by the following search rules:

1. **Selection rule:** If a search is to be made, then the option $j^* \in \bar{S}$ with the highest reservation utility

[27]Note that this model is designed to capture consumers' suboptimal decisions in response to ads, for the reasons explained above. Thus, the value function $V(a, y)$ solely captures the main tradeoff consumers make, not their entire continuation value for each decision.

z_{j^*} should be searched next, where

$$c_{j^*} = \int_{z_{j^*}}^{\infty} (u - z_{j^*}) dF_{j^*}(u). \tag{5}$$

2. **Stopping rule:** Search should terminate when the maximum utility observed so far exceeds the reservation utility z_{j^*} of any unsearched option.

3. **Choice rule:** Once search has terminated, the option with the highest revealed utility among those searched (including the outside option) should be chosen.

In words, if the consumer is not exposed to an ad, then she will search using Weitzman's search rules. That is, among the options available to search $\bar{S} \subset J$, the consumer will rank products by their reservation utilities and continue searching if there exists an option j^* with reservation utility larger than the highest utility observed so far, i.e. if $z_{j^*} \geq y$ and $z_{j^*} \geq z_j, \forall j \in \bar{S}$.[28] When the highest utility observed through search exceeds the reservation utility of any option not yet searched, the consumer stops searching and makes a purchase decision.

If instead the consumer is exposed to an ad a in the AP-weak model, then she solves a problem very similar to the one in Weitzman, except that another option has been added (exogenously) to the set of available options to search. Thus, the consumer will search ad a if $z_a \geq z_j$ and $z_a \geq y, , \forall j \in \bar{S}$. In contrast, in the AP-strong model the consumer will search ad a if $z_a \geq y$.

We will estimate four different models on our data.

1. The **AP-weak model**: The joint active and passive search model proposed above where $A \neq \emptyset$, and the consumer compares z_a with the highest utility among options searched so far, y, and with the reservation utility of all options not yet searched in $\bar{S}$.

2. The **AP-strong model**: A stricter version of the joint active and passive search model proposed above where $A \neq \emptyset$, and the consumer compares z_a only with the highest utility among options searched so far, y.

3. The **Weitzman model**: the case where $A = \emptyset$.

[28]Without loss of generality, we assume throughout the paper that when the consumer is indifferent between searching and stopping search, she will continue searching, and that when the consumer is indifferent between buying and choosing the outside option, she will choose to buy.

4. The **Weitzman model with advertising costs**: the case where $A = \emptyset$, and ads affect the search cost of the consumer.

The estimation of the Weitzman model on our data serves several purposes we wish to highlight. First, it allows us to compare our proposed model against one of the most common benchmarks in the literature. Second, it allows us to test several model assumptions. For example, we can test whether consumers are strategic in searching certain websites (e.g. visiting social networking websites) in order to receive ads from specific brands. In this case, it might be more reasonable to treat these searches as active searches, despite their labeling as passive, by estimating the Weitzman model, which we do. In addition, by estimating the Weitzman model with advertising affecting search costs, we can also test whether firms' actions (advertising in our case) merely affect search costs, or have a more structural effect, like the one proposed by our AP models. Finally, our model assumes the ads that consumers observe are not informative of their preferences. However, in the presence of targeting, such ads may be (partially) revealing of consumer preferences. If this targeting is very precise, firms would perfectly predict consumers' optimal search decisions, making Weitzman the right model; we test for this by estimating the Weitzman model, and further extend our model to explicitly account for ad targeting. More details can be found in Section 8.1 below.

5.3 Example Illustrating Consumer Search Rules Across Models

To better illustrate differences between the search rules in the four models above, consider the example provided in Table 4. Suppose there are five products and an outside option available, that the consumer has searched three of them, and that options 2 and 4 exposed the consumer to ads. Also, suppose that only the ad for option 2 was searched. We ignore the choice rule in this example since it is the same across all models.

========================

Insert Table 4 about here

========================

Columns (i), (ii), and (iii) in Table 4 describe the restrictions on the parameters of interest imposed by the Weitzman, AP-strong, and AP-weak models, respectively. Compared to the Weitzman model, the AP-strong model assumes the consumer is not aware of options 2 and 4. Therefore, their reservation utilities do not affect the search order. Rather, when the consumer is exposed to an ad for these two options, she decides whether to search them solely by comparing their reservation utility with the utility of options searched so far. The AP-weak model maintains this assumption, but also models the consumer as comparing the reservation utility of the advertised options to the reservation utility of unsearched options. The fourth variation of our model, the Weitzman model with advertising costs, is not illustrated in this example, but would impose the same restrictions on the reservation and revealed utilities as the Weitzman model. In this variation, ads affect consumers' search costs, rather than affecting their awareness, as in the AP models.

5.4 Related Problems

Our proposed model of joint active and passive search is related to three problems found in the literature. First, the literature on "arm-acquiring bandits", pioneered by Whittle (1981), considers an extension of the traditional multi-armed bandit problem where arms appear continually while the decision maker evaluates them. Whittle (1981) shows that by assigning a state to every arm, the Gittins index solution that applies in the multi-armed bandit problem continues to hold for i.i.d. arrivals of the arms. The decision rule dictates that the decision maker operate the arm in the state with the largest index as long as it is higher than the best observed reward so far; otherwise the decision maker should stop the process and exploit the best arm. Our model is related to this problem if we think about ads as such arms that are added to the problem the consumer is solving. The difference is that in our case, ads do not appear continually and the consumer does not take their arrival into account.

A second problem related to our model is that of endogenous awareness sets, studied in two recent papers (Greminger, 2020; Fershtman and Pavan, 2020). In these papers, the consumer has the choice to search among options she is aware of or to discover n new options that are then added to her awareness set for potential future search. The authors then describe conditions under which an index policy exists. Our model is similar in the sense that we assume consumers are not aware of or are not considering the options to which ads expose them. However, in our model consumers do not choose to expand their awareness set; rather, ads arrive, expanding this set. We make this modeling choice because it is unlikely that consumers check their email, visit a social networking site, or read the news with the specific purpose of discovering new products, as required by a model with endogenous awareness sets that still views search as active. Instead, it is more likely that consumers value these shopping-unrelated activities *per se*, and that while they engage in these activities, they are exposed to retailers' advertising. Another important distinction is that in the Greminger (2020) and Fershtman and Pavan (2020) models, the consumer knows that exactly n options will be revealed if she chooses to discover more options, and that these options are revealed in order of an index. In contrast, in our empirical setting, it is unlikely that consumers who visit a social networking website or check their email will know if and how many ads they will be exposed to. An empirical setting where consumers scroll through a list of products on a website that has the option to "show more" products will fit a model with endogenous awareness sets better. For these reasons, we expect that our approach to dealing with advertised options captures our empirical setting better.

Finally, our model is also related to the rich literature on random search (e.g., Wolinsky, 1986), where consumers do not choose the order in which they search, but merely choose when to stop searching. Our model relaxes the assumption that consumers choose the order of search for the ads they observe, in the spirit of random search models.

6 Empirical Application and Estimation

6.1 Empirical Model

In our empirical application, we model consumers as searching across websites (e.g., adidas.com, nike.com) in one of the four largest apparel subcategories in our data: (1) shirts, tops, and blouses, (2) shoes, (3) pants and jeans, and (4) underwear.[29] Appendix 11.6 provides details on how the estimation samples were constructed. In the model, consumer $i = 1, \ldots, N$ seeks to purchase from website $j = 1, \ldots, J$ or to choose the outside option of not purchasing, $j = 0$. Consumer i's utility of purchasing from website j is given by

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij} \\ &= w_j + \gamma X_{ij} + \eta_{ij} + \epsilon_{ij} \end{aligned} \tag{6}$$

where v_{ij} denotes the information the consumer has about a website before searching it, and ϵ_{ij} denotes the information revealed through search. The information on v_{ij} includes website intercepts, w_j,

[29] We choose to model search across websites (rather than products within websites) for several reasons: (i) consumers are more likely to search websites directly rather than their individual product subpages, since they rarely know which products are available before navigating from the homepage to various list pages that display such products; (ii) ads vary in the types of pages (e.g. homepage vs sales page) to which they direct consumers, so this modeling assumption allows us to keep our analysis consistent across ads; (iii) developing a model of search across as well as within websites is beyond the scope of our paper.

observed website and consumer characteristics, X_{ij} – such as measures of website loyalty (the number of times the consumer has previously searched a website in other subcategories or in previous spells) and price sensitivity (whether the consumer visited the sales page of a website) – and characteristics unobserved by the researcher but observed by the consumer before search, η_{ij}. Since in the apparel industry prices do not vary over a short time period or across consumers, they do not affect consumers' utility after controlling for website fixed effects.[30] We assume that both η_{ij} and ϵ_{ij} are distributed as standard normal distributions (consistent with prior work (Kim et al., 2010; De los Santos et al., 2012; Honka, 2014; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018; Ursu et al., 2020)).[31] The outside option does not require searching and has a utility equal to $u_{i0} = q_0 + \eta_{i0}$, where q_0 is an intercept denoting the value of not purchasing.

Searching to resolve uncertainty about ϵ_{ij} is costly for consumers. Search costs are given by $c_{ij} = \exp(\kappa)$, modeled as exponential functions to ensure that they are positive and consistent with prior work (e.g., Honka, 2014; Chen and Yao, 2017; Ursu, 2018). In the Weitzman model with advertising costs we allow for the possibility that ads have different search costs, $c_{ia} = \exp(\kappa + \delta Ad_{ia})$.

In our data, we only observe ads that consumers have clicked. This is the case because our data are conditional on a click, as is the case in all prior empirical work on consumer search that uses clickstream data (Kim et al., 2010, 2017; Koulayev, 2014; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020). We use our data to impute the probability of ad exposure based on the search history of the consumer, as described in Appendix 11.6. This approach allows us to more accurately capture the magnitude of the effect of passive search and to be able to consider the Weitzman model with advertising affecting search costs. However, this assumption does not drive our results – our estimates are robust to using only ad clicks and not ad exposures (see Table A-5 in Appendix 11.10).

Finally, clicks induced by search engine advertisements are most similar to searches coming from the organic channel, and might be counted as such, especially if consumers use the search engine to navigate to a specific website. We account for this possibility when estimating our model by running robustness checks and classifying clicks coming from search engines as active searches. Our results

[30] Prices vary mostly across seasons (e.g. when companies run sales promotions) and are generally not personalized to individual consumers. For more facts about pricing in the apparel industry, we refer the reader to Ursu et al. (2021).

[31] Deviating from the standard normal assumption on ϵ_{ij} requires an exogenous search cost shifter (Yavorsky et al., 2021), which we do not have in our online application. Also, the standard deviation of η_{ij} needs to be fixed to 1 for identification purposes. More details can be found in Appendix 11.8.

continue to hold (see Appendix 11.10).

6.2 Estimation

The model variations we will estimate with our data are based on the Weitzman (1979) model. Therefore, we will first describe the estimation procedure of the Weitzman (1979) model that is commonly used in the literature (Kim et al., 2010; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018), and then describe how the other variations differ.

In the Weitzman (1979) model, consumers search options in order of their reservation utilities and stop searching when the best observed utility so far exceeds the reservation utility of any unsearched option. The search rules described in Section 5.2 above translate into the following restrictions on preferences and search cost parameters.

Suppose consumer i searched a number s of websites and that she chose j after stopping her search (including the outside option). With a slight abuse of notation, we order websites in J by their reservation utilities and let n denote the website with the nth largest reservation utility. Since consumers searched websites by reservation utilities, according to the *selection rule*, it must be that[32]

$$z_{in} \geq \max_{k=n+1}^{J} z_{ik}, \qquad \forall n \in \{1, \ldots, J-1\}. \tag{7}$$

In addition, the *stopping rule* imposes the following two restrictions. For the set of websites searched, it must be that

$$z_{in} \geq \max_{k=0}^{n-1} u_{ik}, \qquad \forall n \in \{1, \ldots, s\}. \tag{8}$$

In contrast, for the websites that were not searched, it must be that

$$z_{im} \leq \max_{k=0}^{s} u_{ik}, \qquad \forall m \in \{s+1, \ldots, J\}. \tag{9}$$

Finally, consistent with the *choice rule*, if the consumer chooses j (including the outside option), then her utility from this choice is larger than that of any other searched website, i.e.,

$$u_{ij} \geq \max_{k=0}^{s} \; u_{ik}, \qquad \forall j \in \{0, 1, \ldots, s\}. \tag{10}$$

In what follows, we describe how each model we estimate varies from the Weitzman setup.

1. The **AP-weak model**: for all websites, this model uses the same equations (7-10), except that the set of options in equation (7) does not include any of the advertised websites unknown when

[32]For details on how to compute reservation utilities in our setting, we refer the reader to Appendix 11.8.

searching n.

2. The **AP-strong model**: (i) for websites that did advertise, this model does not impose equation 7; and (ii) for all other websites, this model uses equations (7-10), except that the set of options in equation (7) does not include any of the advertised websites unknown when searching n.

3. The **Weitzman model**: no variation.

4. The **Weitzman model with advertising costs**: this model uses the same equations (7-10) with search costs as a function of advertising.

Differences in the selection and stopping rules across these models are illustrated in the example in Section 5.3.

In addition to the restrictions imposed in equations (7-10) and their variations, we assume that the first search performed by a consumer is free.[33] This assumption is common in prior work (e.g., Honka, 2014; Honka and Chintagunta, 2017) and is necessary since all consumers in our data search at least once.

If consumers search using the rules described above (equations (7-10) and their variations), then they make search and purchase decisions jointly. Thus, the probability of observing a certain outcome in the data for consumer i is characterized by the joint probability of equations (7-10) holding. This probability is given by

$$L_i = Pr\Big(\text{Selection rule}_i, \text{ Stopping rule}_i, \text{ Choice rule}_i\Big). \tag{11}$$

Because consumers make these decisions jointly, the likelihood function does not have a closed-form solution. We use a simulated maximum likelihood approach to estimate the parameters of the model. In choosing the simulation method, we use the logit-smoothed AR simulator following the previous literature (McFadden, 1989; Honka, 2014; Honka and Chintagunta, 2017; Ursu, 2018; Ursu et al., 2021). Implementation details are discussed in Appendix 11.7.

6.3 Identification

Parameter identification in the four models we estimate follows from the identification argument used by standard consumer search models based on the Weitzman model (Kim et al., 2010; Chen and Yao,

[33] We allow for the possibility of no search in our Monte Carlo simulation in Section 6.5.

2017; Honka and Chintagunta, 2017; Ursu, 2018). More precisely, utility parameters are identified from search and purchase frequencies observed in the data. For example, websites that are searched and purchased more frequently will have a larger estimated value. Also, variation in the frequencies with which consumers have previously visited websites and whether they visit price discount pages identify γ. In addition, variation in the frequencies with which websites are searched first, second, etc will further pin down website intercepts. These same data patterns together with the selection, stopping, and choice rules described in Section 6.2 help recover preference estimates of advertising websites in all the models we consider.

Similarly, as in prior work, search costs do not affect purchase decisions (i.e. do not enter the choice rule) and are identified from the number of websites that consumers search. More precisely, the search rules impose an upper and a lower bound on the search cost parameter κ that must have made it optimal for the consumer to perform a certain number of searches. These search rules, however, only recover a range of search costs. The level of search costs is pinned down by the functional form and the distribution of the utility function that dictate the expression of the reservation utility.[34] Finally, for the Weitzman model with advertising costs, the parameter δ shifting search costs due to ads is identified from variation in which and how many ads consumers search and buy.

6.4 Biases in Parameter Estimates

The model variations we consider describe consumer search decisions differently, leading to different parameter estimates. In this section, we describe how these parameter estimates compare with those from the Weitzman model. As we will show below, the two active and passive search models we consider, as well as the Weitzman model with advertising affecting search costs, outperform the standard Weitzman model. Therefore, we will interpret these differences in parameter estimates as biases due to misspecifications of the Weitzman model.

Consider first the AP-weak model. This model does not require the reservation utilities of ads to be lower than the reservation utilities of options searched before them. Therefore, it allows for the possibility that ads have higher reservation utilities (higher expected utilities and lower search costs) than those in the Weitzman model. More precisely, the AP-weak model allows for the possibility that advertising websites were not searched because consumers were unaware of them, rather than

[34] See also Appendix 11.8 for details on the separate identification of search costs and the benefit from searching under our standard assumption of normally distributed match values.

because consumers actively chose not to search them. If this model coincided with the true data generation process, then compared to it, the Weitzman model would underestimate the expected utility of frequently advertised websites. Given these underestimated expected utilities, search cost estimates in Weitzman may remain unchanged or may be slightly higher than in the true AP-weak model. Finally, the outside option estimate would be underestimated by the Weitzman model to rationalize consumers' decisions not to purchase when advertised websites have lower expected utility estimates.

Next, consider the AP-strong model. In addition to not requiring reservation utilities of ads to be lower than the reservation utilities of the options searched before them (as in the AP-weak model) this model also does not require that reservation utilities of ads be higher than the reservation utilities of options not yet searched. With both a lower bound and an upper bound on reservation utilities removed, this model may lead to either higher or lower estimates of consumers' valuation for advertised websites compared to the Weitzman model. The direction of the bias depends on the timing of ad-initiated searches. If the advertised websites are searched predominantly early in the search process – as we generally observe in our data – the lower bound on reservation utilities would be relatively more important in affecting estimates, since many options are yet to be searched. This should lead to an upward bias in the expected utility estimates of advertised websites in the Weitzman model, which incorrectly imposes that websites searched early have higher reservation utilities than those searched later or those not searched. In contrast, the Weitzman model will underestimate the expected utilities of advertised websites if they are searched predominantly later in the search process. Biases in search costs and the outside option parameters resemble those in the AP-weak model.

Finally, consider the Weitzman model with advertising affecting search costs. Consistent with the data patterns we presented in Section 4, advertised websites are rarely purchased, but frequently searched. Thus, the Weitzman model with constant search costs will likely overestimate advertised websites' utilities and underestimate mean search costs.

6.5 Monte Carlo Simulation

We now show that the simulated maximum likelihood method using the logit-smoothed AR simulator can recover the parameters of our model. We do so using Monte Carlo simulations. We generate a data set of 1,000 consumers making choices among five options – four websites and an outside option.

We simplified the model estimated to include only website intercepts, an outside option intercept, and a mean search cost parameter. The true values of these parameters are similar to those from a preliminary estimation of our model. Website 4 will serve as the reference option.

To determine how the presence of ads affects estimates, we choose website 2 as the advertiser. With 25% probability, consumers are aware of website 2 and search it according to the Weitzman optimal search rules. All other consumers are not aware of this website but are exposed to its ad. This assumption allows us to better mimic our data where the same website may be searched organically by some consumers, and through an ad by others. However, the results we present below would continue to hold under different scenarios, including in the case where website 2 advertises to all consumers or in the case where some of the unaware consumers are not exposed to an ad.[35]

We varied the timing of ad exposure as follows: consumers searched websites other than website 2 in decreasing order of their reservation utilities; website 2 had a temporary value for its reservation utility equal to the average reservation utility in the data. Therefore, the ad sometimes appeared before the consumer searched any other options, other times it appeared after the last searched option, but most often it appeared somewhere in the middle. When exposed to the ad, consumers chose whether to search it or not.

To estimate our model, we follow the steps described in Section 6.2 and Appendix 11.7 and use 500 draws from the distribution of utility error terms for each consumer-website combination to construct the likelihood function. We repeat the estimation on 50 different data sets generated using the same true parameters, but different seeds for the utility errors terms.

Our Monte Carlo simulation results are displayed in Table 5. In column (i), we present the true parameters; in column (ii), we show results when data were generated according to the AP-strong model; in column (iii) we show results when data were generated according to the AP-weak model. For each set of data we generated, we estimate two models: the corresponding AP model and the Weitzman model. The coefficients reported represent averages across 50 estimations of our model. In parentheses, we also report the standard deviation of these estimated coefficients.

Two findings are worth emphasizing. First, we find that each version of the AP model, when used to estimate parameters on data that it generated, can recover those parameters well. Second, the Weitzman model, when used to estimate parameters on data generated by either version of the AP

[35]Results are available from the authors upon request.

model, recovers a biased estimate of the advertiser's value, with less or no bias for other parameters. More precisely, the Weitzman model underestimates the value of the advertiser, confirming the predictions from Section 6.4. In this simulation, we focused on a simple model with one advertiser. These effects would be inflated when most or all websites advertised. Also, by exposing consumers to the advertising website predominantly early on in the search process (rather than randomly in the current setup), the Weitzman model would overestimate the value of the advertiser in the AP-strong model, but not in the AP-weak model.

==========================

Insert Table 5 about here

==========================

7 Results

7.1 Estimation Results

We estimate our models on four different apparel subcategories, "shirts, tops, and blouses", "shoes", "pants and jeans", and "underwear". Table 6 below presents our results from the first two subcategories, while Table 7 presents results from the other two subcategories. In bold, we identify the three largest advertisers in each subcategory.

To start, we describe the overall takeaways from our estimation results, consistent across all models and subcategories. As expected, we find that Zalando and H&M, the two largest apparel retailers in the Netherlands, are among the top favorite websites for consumers across several categories. All else equal, consumers prefer websites they visited before – in other subcategories or in previous spells – and visiting a price discount page corresponds to a higher indirect utility, potentially signaling consumers' price sensitivity. The search cost estimates are positive and the coefficients are significant, indicating that consumers get disutility from search. The magnitude of the search costs estimates implies that a 10% increase in search costs per website would decrease total searches by approximately 2%.

==========================

Insert Table 6 about here

==========================
==========================

Insert Table 7 about here

==========================

We now turn to comparing the model estimates. In all subcategories, the main differences in the estimates come from the advertised websites' utilities, highlighted in bold. The estimates of utilities of other websites are statistically similar across the models.

First, consider the estimates of advertised websites' utilities in the AP-strong (column (i)) and the Weitzman (column (iii)) models. In subcategories 1, 3, and 4, the AP-strong model has on average 18% lower estimates of advertised websites' utilities compared to the Weitzman model. Since in all of these subcategories the vast majority of advertised websites are visited early on in the search process (as shown in Table A-1 in Appendix 11.6), our result is consistent with the expected upward bias of the Weitzman model described in Section 6.4. Search cost estimates are 15-30% higher in subcategories 1, 3 and 4 in the AP-strong model than in the Weitzman model. In subcategory 2, the bias of the Weitzman model goes in the opposite direction, with the advertised websites' utilities having on average a 10% higher estimate in the AP-strong model. This result similarly aligns with our expectations, since advertising websites in subcategory 2 are more often searched later in the search process (see Table A-1 in Appendix 11.6).

Second, consider the AP-weak model (column (ii)). The estimates of advertised websites' utilities in this model are higher than in the Weitzman model across all subcategories, as predicted in Section 6.4. For example, in subcategory 1, About You, C&A, and Debijenkorf have 4-10% higher estimated website intercepts in the AP-weak model than in the Weitzman model.

Finally, in the Weitzman model with advertising costs (column (iv)), we find that the estimates of advertised websites' utilities are on average 25% lower compared to the Weitzman model, but it is less costly for consumers to search these options, since δ estimates are negative.

7.2 Model Fit and Data Patterns

To determine which model fits our data better, we first look at the log-likelihood measures reported in the bottom panel of Tables 6 and 7. Here we see that the AP-strong model is the most appropriate for our data across all subcategories. For example, the AP-strong model has 2,572 lower log-likelihood in the first subcategory compared to the Weitzman model. The AP-strong model is followed by the AP-weak model and the Weitzman model with advertising costs in terms of fit. Indeed, the Weitzman model has the worse fit on our data based on log-likelihood.

Since the likelihood functions in each of our models are different, we also employ two other measures of fit to compare models: the mean absolute error (MAE) and the root mean squared error (RMSE). Both measures are frequently used to compare values predicted by a model with observed values.[36] Formally, the mean absolute error (MAE) is defined by

$$MAE = \frac{\sum_j^N |(\hat{y}_j - y_j)|}{N}, \tag{12}$$

while the root mean squared error (RMSE) is defined by

$$RMSE = \sqrt{\frac{\sum_j^N (\hat{y}_j - y_j)^2}{N}}, \tag{13}$$

where $\hat{y}_j$ is the predicted value of of a choice for option j, y_j is the equivalent observed value for j in the data, while N gives the total number of values considered. In our case, j will denote websites and the $\hat{y}$ and y will be the total (predict versus observed) number of searches or purchases made on each website. We obtain predicted values as follows: in each subcategory, we use each model's parameter estimates and simulate consumer choices - searches and purchases - (within sample), averaging out the effect of the utility error terms across 50 simulations. The AP models cannot predict searches of advertising websites since ad decisions are not modeled. Therefore, to compare the model fit in terms of consumer searches, we report the predicted searches only for websites searched actively under each model.

Both measures of fit are positive. Also, the lower the measure, the better the fit of the model. The two measures are similar, although not identical. MAE is generally easier to interpret since it is defined as the average absolute difference between predicted and observed values. In contrast, the RMSE squares errors before averaging, and later takes the square root. Thus, a few large errors may have a larger effect on the RMSE measure, than on MAE. For these reasons, we report both measures in the paper. Our results can be see in the the bottom panel of Tables 6 and Table 7.

Across all subcategories, we find once again that the AP-strong model predicts purchases better than all other models (has lower MAE and RMSE values), followed by the AP-weak model, and the Weitzman model with advertising costs. Next, considering search decisions, we again find that the AP-strong model outperforms all other models, generally followed by the AP-weak model and the Weitzman model with advertising costs.

[36]See for example Kim et al. (2017) in Marketing. Also, the famous Netflix Prize was judged on the basis of RMSE (see wikipedia.org/wiki/Netflix_Prize) and most competitions on Kaggle use the same metric (see kaggle.com/general/215997).

Furthermore, we can show that the improvement in fit of the AP-strong model comes primarily from lowering the prediction error for advertisers. More precisely, for searches of frequently advertising websites, the RMSE decreases on average by 49% across subcategories when moving from the Weitzman model to the AP-strong model. Also, when looking at purchases of frequently advertising websites, the RMSE is on average 25% lower across categories when moving from the Weitzman model to the AP-strong model. Similarly, the MAE decreases by 29% for searches and by 13% for purchases when moving from the Weitzman model to the AP-strong model. In sum, the AP-strong model can more accurately recover data patterns, especially for advertised options, making it all the more relevant for advertisers.

Finally, in addition to looking at model fit measures, we also wish to highlight that the AP-strong model can explain the data patterns we presented in Section 4.5 better than all other models considered. For example, the AP-strong model can explain why ads are more likely to be clicked early rather than late in the search process (Figure 1). This is the case because early in the search process, the best option searched so far has a relatively low value compared to its value later in the process. Therefore, even low quality ads are more likely to be clicked. In contrast, neither the AP-weak nor the Weitzman models can rationalize this pattern, since they impose the constraint that advertised websites need to have higher reservation utilities than all options searched after them. Similarly, the lower advertised websites' utilities recovered by the AP-strong model are consistent with the lower quality of the ad-initiated searches described in Section 4.5 – these searches involve fewer and more expensive products, are shorter, and are less likely to lead to a purchase. All these facts support the idea that the AP-strong model is a useful model through which to understand the role of online advertising as passive search. Our results also provide evidence of consumers' limited ability to compare advertised options to other unsearched products, an important behavioral limitation that models of active search ignore.

8 Model Extensions

In this section, we extend our model in two directions, accommodating different assumptions on consumer and firm behavior.

8.1 Advertising Targeting

Our model assumes that consumers do not anticipate ad exposure and cannot affect the probability or the timing of observing a specific ad, since this would contradict the notion of passive search (Renault, 2016; Ghose and Todri-Adamopoulos, 2016; Honka et al., 2017; Morozov, 2020)). The data patterns presented in Section 4, showing that ad campaigns are season and category specific, rather than consumer specific, support this assumption. However, even though this assumption may hold, advertisers may target their ads to consumers, based on, for example, their search and purchase histories. Since in such cases ads are determined with some knowledge of a consumer's type, an endogeneity problem arises: ads may not only affect consumer choices, but variation in ad exposure across consumers may also be (partially) informative of their preferences. Thus, not accounting for firms' ad targeting may bias our estimates (Manchanda et al., 2004).

We did not account for ad targeting in our main model for at least two reasons. First, our data are observational, similar to all prior empirical work on consumer search that uses clickstream data (Kim et al., 2010, 2017; Koulayev, 2014; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020). In other words, we don't have data from an experiment where ads were randomly shown to consumers. Such experiments have been analyzed in the literature, but the data are specific to one firm (e.g. Blake et al., 2015; Fong, 2017; Sahni and Zhang, 2020), thereby not capturing the entire search process of the consumers, which is our goal. Second, our data contain ad decision from more than 1,000 websites. Each website uses potentially different targeting rules (if any). Prior work, such as Manchanda et al. (2004) and Nair et al. (2017), has exploited knowledge of industry practice or specific targeting rules used by the company studied to directly account for targeting. However, given the large number of websites we observe, such an approach is not feasible in our case. Nevertheless, we will attempt to at least partially account for ad targeting in this section using the following approach.

To understand how ad targeting might enter our model, recall that our main departure from a model of active search involves assuming consumers cannot choose the order in which they search the options they are exposed to through ads. Thus, we propose that when ads are targeted, consumers search them in an order influenced by the firm. The amount of information the firm possesses and the level of targeting it chooses will affect the degree to which the firm can influence the search order. At one extreme, ads may be targeted very precisely, such that they are shown to consumers when they are most likely to click on them (similar to scenarios considered in Blake et al. (2015) and Simonov et al. (2018)). If consumers search in order of their reservation utilities, they will be most likely to click on an ad when its reservation utility is highest among unsearched options. Thus, with such a precise level of targeting, the model would be indistinguishable empirically from the Weitzman (1979) model, since ads will be searched in order of their reservation utilities. We tested this possibility by estimating the Weitzman (1979) model on our data. However, we did not find support for this form of precise targeting in our data. At another extreme, with little or no targeting, the set and the timing of the ads consumers are exposed to will not be informative of their preferences (as in a random search model, e.g. Wolinsky (1986)). Both of our AP models correctly account for this possibility. Having identified the boundaries of our model, we also wish to account for the intermediate case where ads are targeted, but not very precisely. For this, we proceed as follows.

Many forms of online ads (e.g. search engine, social media, display, retargeting) are served through an auction (Narayanan and Kalyanam, 2015). A platform (e.g Google, Facebook) collects information about consumers (e.g. keywords searched, content browsed, products purchased) and allows advertisers to bid for the chance to show ads based on this information. Advertisers are then ranked on a score (e.g. called AdRank at Google), which is a function of the advertisers' bids and a quality measure given by the platform (Varian, 2007). Using this knowledge and in the absence of company-specific information of their targeting decisions, we take a reduced-form approach to approximating the probability of a consumer seeing an ad.[37] More precisely, we model the probability of a consumer i seeing an ad from a website j as follows

$$Pr(\text{ad}_{ij}) = Prob(\text{score}_{ij} > \max\{\text{score}_{ij'}\}), \forall j' \in \hat{J}, \tag{14}$$

capturing the idea that advertisers with a higher score/value for consumer i will be more likely to show an ad to that consumer. We specify a firm j's estimate of the score as

$$\text{score}_{ij} \quad = \quad w_j + \gamma X_{ij} + \mu_{ij}, \tag{15}$$

where we assume μ_{ij} is normally distributed with mean zero and standard deviation σ_μ.[38] w_j denotes brand intercepts and X_{ij} are consumer-brand characteristics, similar to the consumer utility specification in equation 6.

[37]Other types of online ads (e.g. email, affiliate) similarly involve the ability to reach a favorable audience based on information collected about it, with companies that find it more beneficial to reach a certain consumer being more willing to pay for that privilege. Therefore, we believe our simple model should also capture decisions for such ads.

[38]Note that we do not include consumer specific fixed effects in the specification of the score, since these would not affect the probability of observing an ad as specified in equation (14).

We vary the score specification and the precision of the firms' information in order to account for as many ad targeting scenarios as possible. First, some firms may have access to data on the previous search history of a consumer, while others may not. To account for this possibility we will vary the set of options that affect $Pr(\text{ad}_{ij})$, i.e. we will vary the size and composition of $\hat{J}$. For firms without knowledge of consumer browsing histories, $\hat{J}$ will include the entire set of available options. In contrast, firms with access to more information can condition $\hat{J}$ on consumer browsing histories. We will consider two options: either $\hat{J}$ = options not yet searched or $\hat{J}$ = options already searched. Second, firms may have more or less uncertainty in their estimate of the score_{ij}. For example, some firms may have access to more data describing the consumer (e.g. demographics), while other may not. Thus, we will let σ_μ take on two extreme values, 3 and 0.1 (extreme values relative to our standard normal assumption on μ_{ij}). To estimate the model accounting for targeting, we include $Pr(\text{ad}_{ij})$ in the likelihood function of the AP-strong model for searches that involve ad clicks (i.e. it enters as a marginal probability multiplying the likelihood in the AP-strong model for such clicks, following Manchanda et al. (2004) and Nair et al. (2017)). Our approach will be to check the fit of the model under different assumptions on firms' informational precision and targeting abilities and to compare it with that of our main model specification, the AP-strong model. Once again, our results should be interpreted as suggestive, since we do not have data from an experiment varying ad exposure and we do not know what targeting rules each of the 1,046 websites in our data use (if any).

Our results for each apparel subcategory can be found in Tables A-6 to A-9 in Appendix 11.10. Across all specifications, we find that the fit of the AP-strong model is better than that of all models that account for targeting (the relevant corresponding results can be found in Tables 6 and 7, column 1). In addition, we find that among the models that account for targeting, those that assume firms' targeting is less precise (larger σ_μ) fit better, further supporting the notion that our model of active and passive search better describes consumer behavior in the data. This evidence is also consistent with our descriptive patterns showing that ad-initiated searches are generally of lower quality and are unlikely to result in a purchase.

8.2 Awareness

In our main model specification, we assumed that advertising affects the consumer's choice process at the awareness stage (informative effect). This assumption is supported by our empirical evidence –

consumers rarely have a prior relationship with a brand they also visit through ads and advertised options are rarely purchased (see Section 4). However, ads may also have a persuasive effect, affecting consumer choices conditional on awareness.[39] Towards this end, we estimate the Weitzman model on our data under the assumption that consumers are aware of all available options (i.e. ads do not have an informative effect), but advertising may affect their purchase utility. We address the potential endogeneity of websites' advertising using a control function approach based on the rationale behind ad targeting presented in the previous section (for details, see Appendix 11.9). Our results for each of the four apparel subcategories can be found in the first column of Tables A-3 and A-4 in Appendix 11.9. As expected, we find a positive effect of advertising. However, our models of active and passive search fit the data better than the model where ads have a persuasive effect (relevant comparison results can be found in Tables 6 and 7, columns 1 and 2).

Finally, we also test the possibility that ads have both an informative and a persuasive effect. Towards this end, we start from our AP-strong model where ads affect consumer awareness and in addition test whether ads also affect the utility of the consumer conditional on awareness. Our results can be found in the third column for each subcategory in Tables A-3 and A-4, Appendix 11.9. Here we see that advertising has a smaller effect on utility than in a model where ads only have a persuasive effect (since part of the effect is informational). In terms of model fit, we see that accounting for an informative effect in addition to the persuasive effect leads to a large improvement in model fit (comparing the first and third columns of Tables A-3 and A-4 for each subcategory). When comparing the AP-strong model with both informative and persuasive ad effects to the one with only an informative effect (column 1 in Tables 6 and 7), our results are more mixed. We find that a model that accounts for both persuasive and informative effects fits worse in some subcategories, but better in others. For example, in subcategories 1 and 4, the AP-strong model with both informative and persuasive ad effects explains purchases worse, but searches better, while in the subcategory 3 the reserve is true. However, in all cases the differences are very small (most are less than 3 on the MAD or the RMSE scale), leading us to conclude that accounting for informative effects of advertising with our AP-strong model captures the primary effect of advertising.

[39] In addition to an informative and a persuasive effect, advertising may also serve to merely remind consumers of a brand. Testing for this effect is beyond the scope of our paper, since our model, which is based on the Weitzman model, cannot accommodate revisit behavior (required in order to measure the effect of reminding consumers after an initial ad exposure). Such revisit behavior can be accommodated by a search with learning model, such as Gardete and Hunter (2020); Ursu et al. (2020). As we report in 11.6, approximately 30% of clicks are revisits in our data. Additionally, we find that retargeting ads (most likely to have a reminder effect) constitute fewer than 1% of clicks.

9 Managerial Implications

Our results describe the role of ad-initiated clicks in the consumer search process – ad-initiated searches represent the majority of website visits, occur predominantly early in the search process, and are unlikely to lead to a transaction. Not taking this into account will distort estimates of consumer preferences.

For advertisers, understanding the extent to which consumers seek out their products actively rather than only react to their product messages, can help inform advertising decisions. For example, assuming a consumer has actively searched a product on Nike.com, rather than passively reacted to a Nike ad – even if it is for the same product – implies wrongly assuming that a consumer expects Nike's product offerings to dominate those of other brands, inflating consumer brand preferences. Our results show that a model of passive search allows companies to better predict search and purchase decisions of advertised products than the Weitzman model of active search.

More broadly, our paper questions the assumption that every click performed by a consumer online is an outcome of an active search process. Beyond the case we focus on in this paper, where advertising affects search decisions, there are several other cases in which we expect a click to not reflect an active search decision. For example, in many settings, consumers may merely be curious about a product or may be browsing rather than searching for information with the goal of making a purchase decision (e.g. a consumer typing in "Ferrari" into Google out of curiosity, rather than because she is interested in gathering information towards her next purchase). In such cases, companies should account for passive searches when running their (re-)targeting advertising campaigns – if a consumer stumbled upon a product webpage while browsing, it might be a weak signal of the consumer's interest in buying the product and it may thus be wasteful to (re-)target this consumer with online advertising. A broader understanding of passive search settings (perhaps using similar methods as Moe (2003)), as well as a formal treatment of decision making in such cases would be theoretically and managerially relevant.

10 Conclusion

In this paper, we model model the nature of ad-initiated searches. Using a detailed clickstream data set capturing website visits at the exact URL level, we develop and apply a method that classifies clicks into ad-initiated and organic searches. We then show that ad-initiated searches are extensive – driving more than half of all website arrivals in our category– happen early on in the search process, and lead to less in-depth searches and fewer purchases. These patterns do not align with standard models of active information search (Weitzman, 1979), and instead are consistent with models of passive search, such as Renault (2016). To account for such passive search, we develop a simple model that accommodates both active and passive search decisions by consumers and estimate this model on our data. The results show that a model of active and passive search fits the data the best, while treating all searches as active leads to substantial biases in the estimates. Finally, we show that our model can more accurately recover data patterns, especially for advertising brands, and we explore two extensions of it.

References

Abhishek, V., Fader, P., and Hosanagar, K. (2012). Media exposure through the funnel: A model of multi-stage attribution. Working paper.

Ackerberg, D. A. (2001). Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, pages 316–333.

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: An empirical examination. *International Economic Review*, 44(3):1007–1040.

Amano, T., Rhodes, A., and Seiler, S. (2022). Flexible demand estimation with search data. Working Paper.

Anderson, S. P. and Renault, R. (1999). Pricing, product diversity, and search costs: A Bertrand-Chamberlin-Diamond model. *The RAND Journal of Economics*, 30(4):719–735.

Anderson, S. P. and Renault, R. (2006). Advertising content. *American Economic Review*, 96(1):93–113.

Anderson, S. P. and Renault, R. (2013). The advertising mix for a search good. *Management Science*, 59(1):69–83.

Beales, H., Mazis, M. B., Salop, S. C., and Staelin, R. (1981). Consumer search and public policy. *Journal of Consumer Research*, 8(1):11–22.

Beatty, S. E. and Smith, S. M. (1987). External search effort: An investigation across several product categories. *Journal of Consumer Research*, 14(1):83–95.

Bennett, P. D. and Mandell, R. M. (1969). Prepurchase information seeking behavior of new car purchasers—The learning hypothesis. *Journal of Marketing Research*, 6(4):430–433.

Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1):155–174.

Branco, F., Sun, M., and Villas-Boas, J. M. (2012). Optimal search for product information. *Management Science*, 58(11):2037–2056.

Branco, F., Sun, M., and Villas-Boas, J. M. (2016). Too much information? Information provision and search costs. *Marketing Science*, 35(4):605–618.

Breen, R., Karlson, K., and Holm, A. (2018). Interpreting and understanding logits, probits, and other nonlinear probability models. *Annual Review of Sociology*, 44:39–54.

Bronnenberg, B. J., Kim, J. B., and Mela, C. F. (2016). Zooming in on choice: How do consumers search for cameras online? *Marketing Science*, 35(5):693–712.

Burguet, R. and Petrikaite, V. (2017). Targeted advertising and costly consumer search. Working paper.

Butters, G. R. (1978). Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, 44(3):465–419.

Cameron, S. and Heckman, J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106:262–333.

Chan, T. Y., Wu, C., and Xie, Y. (2011). Measuring the lifetime value of customers acquired from google search advertising. *Marketing Science*, 30(5):837–850.

Chen, Y. and Yao, S. (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science*, 63(12):4345–4365.

De Corniere, A. (2016). Search advertising. *American Economic Journal: Microeconomics*, 8(3):156–88.

De los Santos, B., Hortaçsu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80.

De los Santos, B. and Koulayev, S. (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science*, 36(4):542–564.

Du, R. Y., Xu, L., and Wilbur, K. C. (2019). Immediate responses of online brand search and price search to tv ads. *Journal of Marketing*, 83(4):81–100.

Dukes, A. and Liu, L. (2016). Online shopping intermediaries: The strategic design of search environments. *Management Science*, 62(4):1064–1077.

Duncan, C. P. and Olshavsky, R. W. (1982). External search: The role of consumer beliefs. *Journal of Marketing Research*, 19(1):32–43.

Fershtman, D. and Pavan, A. (2020). Sequential learning with endogenous consideration sets. Working Paper.

Fong, N. M. (2017). How targeting affects customer search: A field experiment. *Management Science*, 63(7):2353–2364.

Furse, D. H., Punj, G. N., and Stewart, D. W. (1984). A typology of individual search strategies among purchasers of new automobiles. *Journal of Consumer Research*, 10(4):417–431.

Gardete, P. and Hunter, M. (2020). Guiding consumers through lemons and peaches: An analysis of the effects of search design activities. Working Paper.

Ghose, A. and Todri-Adamopoulos, V. (2016). Toward a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly*, 40(4):889–910.

Goeree, M. S. (2008). Limited information and advertising in the US personal computer industry. *Econometrica*, 76(5):1017–1074.

Golden, J. M. and Horton, J. J. (2020). The effects of search advertising on competitors: An experiment before a merger. *Management Science*, 67(1):342–362.

Goldfarb, A. and Tucker, C. (2011). Search engine advertising: Channel substitution when pricing ads to context. *Management Science*, 57(3):458–470.

Gordon, B. R., Zettelmeyer, F., Bhargava, N., and Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Marketing Science*. Publisher: INFORMS.

Gossner, O., Steiner, J., and Stewart, C. (2020). Attention please! Forthcoming, *Econometrica*.

Greminger, R. (2020). Optimal search and discovery. Working Paper.

Hong, H. and Shum, M. (2006). Using price distributions to estimate search costs. *The RAND Journal of Economics*, 37(2):257–275.

Honka, E. (2014). Quantifying search and switching costs in the US auto insurance industry. *The RAND Journal of Economics*, 45(4):847–884.

Honka, E. and Chintagunta, P. (2017). Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Science*, 36(1):21–42.

Honka, E., Hortaçsu, A., and Vitorino, M. A. (2017). Advertising, consumer awareness and choice: Evidence from the US banking industry. *The RAND Journal of Economics*, 48(3):611–646.

Iyer, G., Soberman, D., and Villas-Boas, J. M. (2005). The targeting of advertising. *Marketing Science*, 24(3):461–476.

Jeziorski, P. and Moorthy, S. (2017). Advertiser prominence effects in search advertising. *Management Science*, 64(3):1365–1383.

Jeziorski, P. and Segal, I. (2015). What makes them click: Empirical analysis of consumer demand for search advertising. *American Economic Journal: Microeconomics*, 7(3):24–53.

Joo, M., Wilbur, K. C., Cowgill, B., and Zhu, Y. (2013). Television advertising and online search. *Management Science*, 60(1):56–73.

Joo, M., Wilbur, K. C., and Zhu, Y. (2016). Effects of TV advertising on keyword search. *International Journal of Research in Marketing*, 33(3):508–523.

Katona, G. and Mueller, E. (1955). A study of purchase decisions. In *Consumer Behavior: The Dynamics of Consumer Reaction*, volume 1, pages 30–87. New York: New York University Press, 1955.

Ke, T. T., Shen, Z.-J. M., and Villas-Boas, J. M. (2016). Search for information on multiple products. *Management Science*, 62(12):3576–3603.

Ke, T. T. and Villas-Boas, J. M. (2019). Optimal learning before choice. *Journal of Economic Theory*, 180:383–437.

Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2010). Online demand under limited consumer search. *Marketing Science*, 29(6):1001–1023.

Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2017). The probit choice model under sequential search with an application to online retailing. *Management Science*, 63(11):3911–3929.

Kireyev, P., Pauwels, K., and Gupta, S. (2016). Do display ads influence search? Attribution and dynamics in online advertising. *International Journal of Research in Marketing*, 33(3):475–490.

Koulayev, S. (2014). Search for differentiated products: Identification and estimation. *The RAND Journal of Economics*, 45(3):553–575.

Lambrecht, A. and Tucker, C. (2013). When does retargeting work? information specificity in online advertising. *Journal of Marketing Research*, 50(5):561–576.

Li, H. and Kannan, P. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51(1):40–56.

Lynch, J. and Srull, T. (1982). Memory and attentional factors in consumer choice: Concepts and research methods. *Journal of Consumer Research*, 9(1):18–37.

Ma, L. (2016). Only the interested learn: A model of proactive learning of product reviews. Working Paper.

Maddala, G. (1983). In *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge Univ. Press.

Manchanda, P., P, R., and P, C. (2004). Resoonse modeling with nonrandom marketing-mix variables. *Journal of Marketing Research*, 41(4):467–478.

Mayzlin, D. and Shin, J. (2011). Uninformative advertising as an invitation to search. *Marketing Science*, 30(4):666–685.

McFadden, D. (1989). A method of simulating moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5).

Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology*, 13(1-2):29–39.

Moraga-González, J. L., Sándor, Z., and Wildenbeest, M. R. (2015). Consumer search and prices in the automobile market. Working Paper.

Moraga-González, J. L. and Wildenbeest, M. R. (2008). Maximum likelihood estimation of search costs. *European Economic Review*, 52(5):820–848.

Morozov, I. (2020). Measuring benefits from new products in markets with information frictions. Working paper.

Nair, H., Misra, S., Hornbuckle, W. I., Mishra, R., and Acharya (2017). Big data and marketing analytics in gaming: Combining empirical models and field experimentation. *Marketing Science*, 36(5):699–725.

Narayanan, S. and Kalyanam, K. (2015). Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science*, 34(4):612–627.

Newman, J. W. and Lockeman, B. D. (1975). Measuring prepurchase information seeking. *Journal of Consumer Research*, 2(3):216–222.

Newman, J. W. and Staelin, R. (1972). Prepurchase information seeking for new cars and major household appliances. *Journal of Marketing Research*, 9(3):249–257.

Petrin A, T. K. (2010). Optimizing click-through in online rankings with endogenous search refinement. *Journal of Marketing Research*, 47(1):3–13.

Rao, J. M. and Simonov, A. (2018). Firms' Reactions to Public Information on Business Practices: Case of Search Advertising. *Quantitative Marketing and Economics*, Forthcoming.

Ratchford, B. T. (2008). Consumer search behavior and its effect on markets. *Foundations and Trends® in Marketing*, 3(1):1–74.

Renault, R. (2016). Passive internet search. Working paper.

Robert, J. and Stahl, D. O. (1993). Informative price advertising in a sequential search model. *Econometrica: Journal of the Econometric Society*, pages 657–686.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202.

Rutz, O. J. and Bucklin, R. E. (2012). Does banner advertising affect browsing for brands? clickstream choice model says yes, for some. *Quantitative Marketing and Economics*, 10(2):231–257.

Rutz, O. J., Trusov, M., and Bucklin, R. E. (2011). Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science*, 30(4):646–665.

Sahni, N. and Zhang, C. (2020). Are consumers averse to sponsored messages? The role of search advertising in information discovery. Working Paper.

Seiler, S. (2013). The impact of search costs on consumer behavior: A dynamic approach. *Quantitative Marketing and Economics*, 11(2):155–203.

Shim, S. and Drake, M. F. (1989). Information search in the purchase of employment apparel: A synthesis of two theories. *Clothing and Textiles Research Journal*, 7(3):40–46.

Shin, J. and Yu, J. (2021). Targeted advertising and consumer inference. *Marketing Science*, 40(5):900–922.

Simonov, A. and Hill, S. (2021). Competitive advertising on brand search: Traffic stealing and click quality. *Marketing Science*, Forthcoming.

Simonov, A., Nosko, C., and Rao, J. M. (2018). Competition and crowd-out for brand keywords in sponsored search. *Marketing Science*, 37(2):200–215.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.

Terui, N., Ban, M., and Allenby, G. M. (2011). The effect of media advertising on brand consideration and choice. *Marketing Science*, 30(1):74–91.

Tsai, Y.-L. and Honka, E. (2018). Non-informational advertising informing consumers: How advertising affects consumers' decision-making in the US auto insurance industry. Working paper.

Ursu, R., Wang, Q., and Chintagunta, P. K. (2020). Search duration. *Marketing Science*, 39(5):849–871.

Ursu, R., Zhang, Q., and Honka, E. (2021). Search gaps and consumer fatigue. Working Paper.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552.

Varian, H. R. (2007). Position auctions. *international Journal of industrial Organization*, 25(6):1163–1178. Publisher: Elsevier.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, pages 641–654.

Whittle, P. (1981). Arm-acquiring bandits. *The Annals of Probability*, 9(2):284–292.

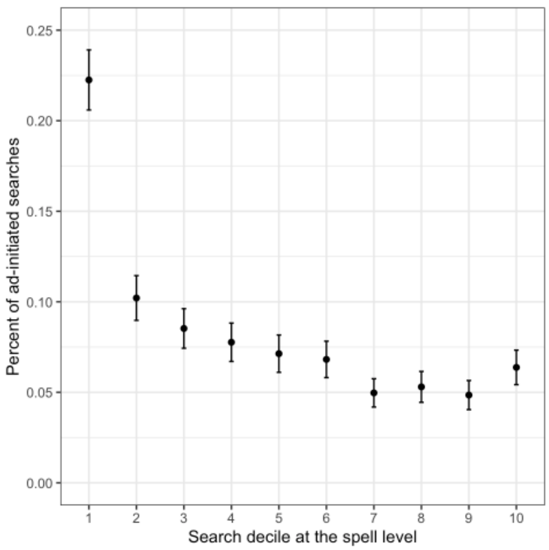
Wolinsky, A. (1986). True monopolistic competition as a result of imperfect information. *The Quarterly Journal of Economics*, 101(3):493–511.

Yang, S. and Ghose, A. (2010). Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, 29(4):602–623.

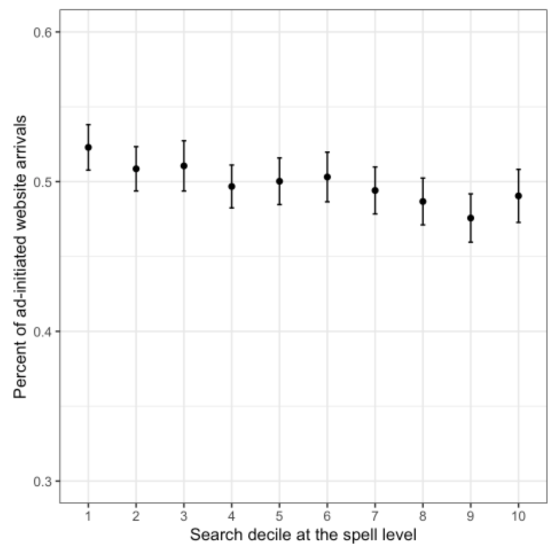
Yavorsky, D., Honka, E., and Chen, K. (2021). Consumer search in the u.s. auto industry: The role of dealership visits. *Quantitative Marketing and Economics*, 19:1–52.

Figures and Tables

Figure 1: Percent of Ads by Progress in the Spell



(a) Ad-Initiated Searches



(b) Ad-Initiated Website Arrivals

Figure 2: Average Statistics by Quantiles of Ad-Initiated Searches per Spell

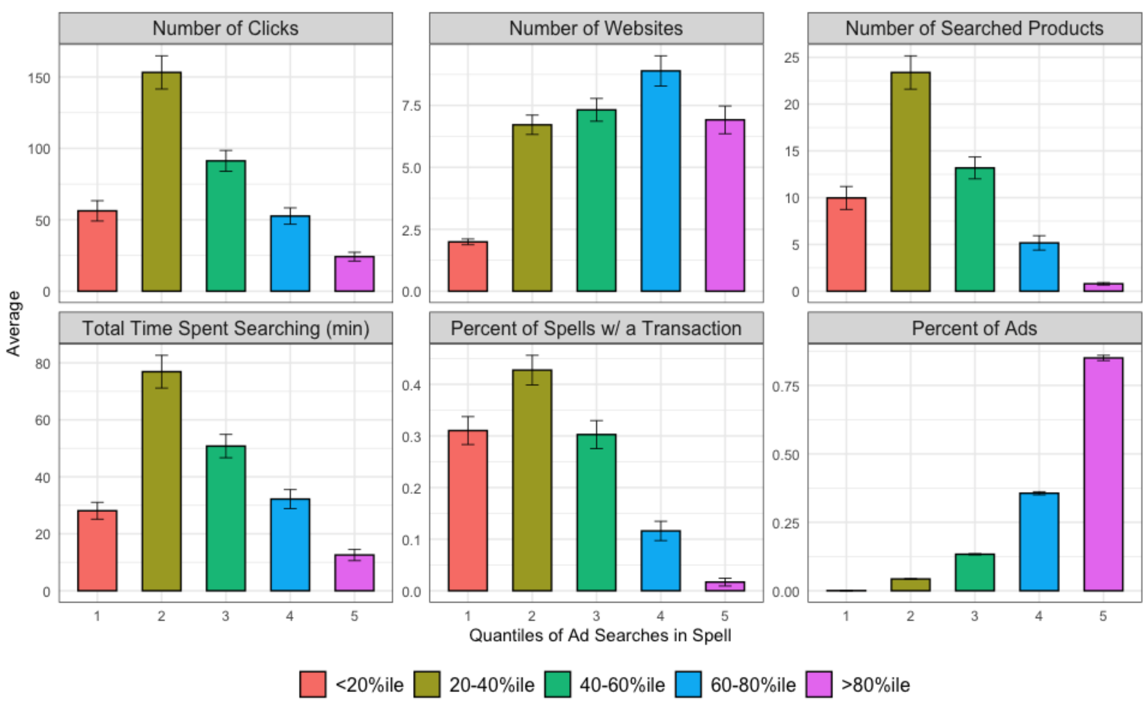
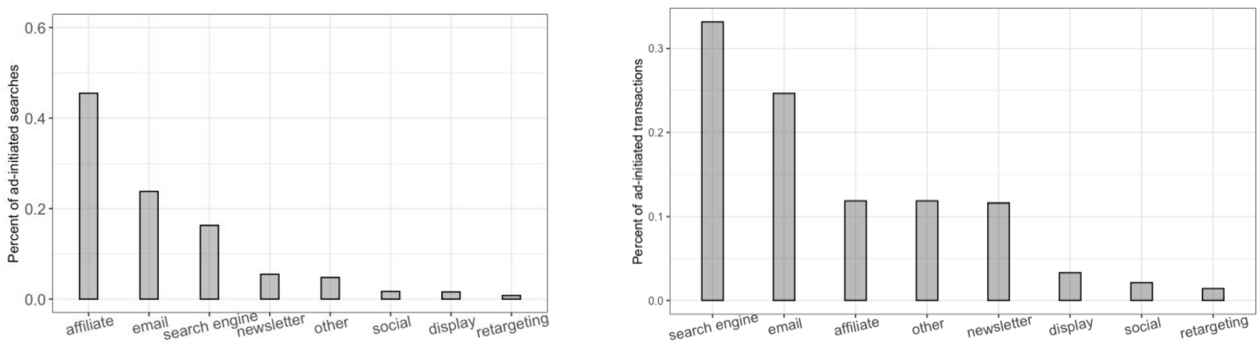


Figure 3: Ad types



(a) Percent of searches by ad type

(b) Percent of transactions by ad type

Table 1: Characteristics of Apparel Searches in a Spell

	Mean	Median	St. Dev.
Number of Sessions	7.19	3.00	164.00
Number of Clicks	75.72	32.00	2447.00
Number of Websites Searched	6.36	3.00	137.00
Number of Products Searched	22.88	9.00	596.00
Total Duration (min)	40.20	17.63	1162.05

Table 2: Frequency of Organic and Ad-Initiated Searches

search type	observations	percent
Among all clicks		
organic	328,508	85.5%
ad-initiated	55,838	14.5%
Among all website arrivals		
organic	33,194	47.3%
ad-initiated	36,980	52.7%

Table 3: Website-level Summary Statistics Based on the Website Arrival Type

	Organic	Ad-Initiated	*T-stat*
Percent of All Transactions	0.71	0.29	–
Number of Clicks	8.37	4.10	30.74
Time Spent on Website (min)	4.32	2.28	31.75
Number of Searched Products	1.28	0.58	24.60
Standardized Price of Clicked Products	-0.14	0.16	-1.95

Notes: The last column reports the t-statistic for the difference in means.

Table 4: Example Illustrating Differences Across Models

option	searched	ad	(i) *Weitzman model*	(ii) *AP-strong model*	(iii) *AP-weak model*
1	1	0	$z1 \geq z2, z1 \geq z3, z1 \geq z4, z1 \geq z5$	$\cancel{z1 \geq z2}, z1 \geq z3, \cancel{z1 \geq z4}, z1 \geq z5$	$\cancel{z1 \geq z2}, z1 \geq z3, \cancel{z1 \geq z4}, z1 \geq z5$
2	1	1	$z2 \geq u1, z2 \geq z3, z2 \geq z4, z2 \geq z5$	$z2 \geq u1, \cancel{z2 \geq z3}, \cancel{z2 \geq z4}, \cancel{z2 \geq z5}$	$z2 \geq u1, z2 \geq z3, \cancel{z2 \geq z4}, z2 \geq z5$
3	1	0	$z3 \geq max\{u1, u2\}, z3 \geq z4, z3 \geq z5$	$z3 \geq max\{u1, u2\}, \cancel{z3 \geq z4}, z3 \geq z5$	$z3 \geq max\{u1, u2\}, \cancel{z3 \geq z4}, z3 \geq z5$
4	0	1	$z4 < max\{u1, u2, u3\}$	$z4 < max\{u1, u2, u3\}$	$z4 < max\{u1, u2, u3\}$
5	0	0	$z5 < max\{u1, u2, u3\}$	$z5 < max\{u1, u2, u3\}$	$z5 \leq max\{u1, u2, u3\}$

Table 5: Monte Carlo Simulation Results

	(i)	(ii)		(iii)	
Data Generating Model:		*AP-strong*		*AP-weak*	
Estimation Model:		*AP-strong*	*Weitzman*	*AP-weak*	*Weitzman*
	True values	Estimates (SD)		Estimates (SD)	
Utility					
Outside option	0.5	0.48	0.43	0.49	0.45
		(0.08)	(0.08)	(0.08)	(0.07)
Website 1	-1	-0.88	-0.87	-0.90	-0.88
		(0.06)	(0.08)	(0.07)	(0.08)
Website 2 (advertiser)	-0.5	**-0.49**	**-0.65**	**-0.43**	**-0.63**
		(0.07)	(0.07)	(0.07)	(0.07)
Website 3	-0.3	-0.27	-0.27	-0.27	-0.26
		(0.06)	(0.07)	(0.06)	(0.07)
Search cost (exp)					
Constant	-3	-2.97	-3.03	-2.86	-2.92
		(0.10)	(0.10)	(0.11)	(0.10)
Log-likelihood		-3,745	-4,071	-3,726	-3,995
Number of Observations		5,000	5,000	5,000	5,000

Table 6: Estimation Results

	Subcat. 1: "Shirts, tops, & blouses"			
	(i)	(ii)	(iii)	(iv)
	AP-strong	*AP-weak*	*Weitzman*	*Weitzman*
Utility				
aboutyou.com	-1.28***	-1.10***	-1.18***	-1.44***
	(0.04)	(0.03)	(0.03)	(0.03)
c-and-a.com	-0.78***	-0.61***	-0.68***	-0.90***
	(0.03)	(0.03)	(0.03)	(0.03)
debijenkorf.nl	-1.62***	-1.51***	-1.57***	-1.76***
	(0.04)	(0.03)	(0.04)	(0.04)
esprit.nl	-1.66***	-1.71***	-1.72***	-1.62***
	(0.05)	(0.04)	(0.04)	(0.04)
hm.com	-1.19***	-1.21***	-1.23***	-1.17***
	(0.03)	(0.03)	(0.03)	(0.03)
jbfo.nl	-2.47***	-2.53***	-2.54***	-2.46***
	(0.12)	(0.12)	(0.13)	(0.11)
msmode.nl	-1.69***	-1.72***	-1.72***	-1.62***
	(0.04)	(0.04)	(0.04)	(0.05)
peterhahn.nl	-1.74***	-1.83***	-1.85***	-1.76***
	(0.06)	(0.05)	(0.07)	(0.06)
your-look-for-less.nl	-1.39***	-1.42***	-1.45***	-1.37***
	(0.04)	(0.04)	(0.04)	(0.04)
zalando.nl	-1.09***	-1.09***	-1.10***	-1.04***
	(0.04)	(0.04)	(0.03)	(0.03)
Number of previous website visits	0.20***	0.19***	0.16***	0.16***
	(0.01)	(0.01)	(0.01)	(0.01)
Visit to a price discount page	1.94***	2.00***	1.76***	1.62***
	(0.03)	(0.03)	(0.04)	(0.04)
Outside option	2.07***	2.10***	1.98***	1.95***
	(0.03)	(0.04)	(0.03)	(0.03)
Search cost (exp)				
Constant	-3.97***	-4.18***	-4.15***	-3.71***
	(0.04)	(0.06)	(0.07)	(0.05)
Advertising				-2.16***
				(0.08)
Observations	32422	32422	32422	32422
LL	-8974	-10381	-11546	-11095
MAE (purchase)	34.59	36.77	41.13	36.55
MAE (search)	59.15	65.73	93.86	74.84
RMSE (purchase)	59.80	61.76	70.73	67.77
RMSE (search)	65.00	75.02	101.86	86.15

	Subcat. 2: "Shoes"			
	(i)	(ii)	(iii)	(iv)
	AP-strong	*AP-weak*	*Weitzman*	*Weitzman*
Utility				
adidas.com	-0.95***	-1.09***	-1.04***	-0.88***
	(0.03)	(0.03)	(0.02)	(0.04)
debijenkorf.nl	-1.64***	-1.81***	-1.75***	-1.52***
	(0.05)	(0.04)	(0.03)	(0.05)
nelson.nl	-1.33***	-1.42***	-1.33***	-1.18***
	(0.03)	(0.03)	(0.03)	(0.03)
nike.com	-1.07***	-1.31***	-1.21***	-0.93***
	(0.03)	(0.03)	(0.03)	(0.03)
omoda.nl	-1.40***	-1.53***	-1.45***	-1.23***
	(0.04)	(0.03)	(0.03)	(0.04)
schuurman-shoenen.nl	-0.64***	-0.64***	-0.78***	-1.05***
	(0.03)	(0.02)	(0.02)	(0.02)
spartoo.nl	-1.09***	-1.05***	-1.14***	-1.52***
	(0.03)	(0.03)	(0.03)	(0.03)
vanharen.nl	-0.79***	-0.79***	-0.87***	-1.13***
	(0.03)	(0.02)	(0.02)	(0.02)
zalando.nl	-0.57***	-0.74***	-0.73***	-0.56***
	(0.03)	(0.02)	(0.02)	(0.03)
ziengs.nl	-1.58***	-1.62***	-1.54***	-1.41***
	(0.05)	(0.04)	(0.03)	(0.04)
Number of previous website visits	0.21***	0.21***	0.21***	0.20***
	(0.01)	(0.01)	(0.01)	(0.02)
Visit to a price discount page	1.47***	1.48***	1.18***	0.96***
	(0.05)	(0.04)	(0.03)	(0.04)
Outside option	2.35***	2.25***	2.09***	2.11***
	(0.03)	(0.03)	(0.02)	(0.04)
Search cost (exp)				
Constant	-5.13***	-5.28***	-4.97***	-4.11***
	(0.05)	(0.05)	(0.05)	(0.08)
Advertising				-2.78***
				(0.07)
Observations	34812	34812	34812	34812
LL	-13129	-15907	-18585	-17312
MAD (purchase)	36.57	40.52	48.54	44.65
MAD (search)	109.27	92.52	180.38	376.04
RMSE (purchase)	57.60	65.96	76.24	69.18
RMSE (search)	118.10	102.96	197.40	435.22

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers in each subcategory.

Table 7: Estimation Results (continued)

	Subcat. 3: "Pants & Jeans"			
	(i)	(ii)	(iii)	(iv)
	AP-strong	*AP-weak*	*Weitzman*	*Weitzman*
Utility				
c-and-a.com	-0.77***	-0.83***	-0.82***	-0.81***
	(0.04)	(0.03)	(0.03)	(0.04)
debijenkorf.nl	-1.35***	-1.42***	-1.41***	-1.40***
	(0.05)	(0.05)	(0.04)	(0.05)
esprit.nl	-1.33***	-1.39***	-1.36***	-1.32***
	(0.05)	(0.05)	(0.05)	(0.05)
g-star.com	-1.81***	-1.86***	-1.85***	-1.84***
	(0.08)	(0.07)	(0.07)	(0.07)
hm.com	-0.86***	-0.92***	-0.91***	-0.89***
	(0.04)	(0.04)	(0.03)	(0.04)
jeanscentre.nl	-1.48***	-1.41***	-1.44***	-1.67***
	(0.07)	(0.05)	(0.05)	(0.05)
missetam.nl	-0.91***	-0.43***	-0.43***	-0.66***
	(0.04)	(0.03)	(0.03)	(0.03)
tommy.com	-1.98***	-1.95***	-1.93***	-1.96***
	(0.11)	(0.08)	(0.08)	(0.08)
your-look-for-less.nl	-1.21***	-1.28***	-1.27***	-1.24***
	(0.05)	(0.04)	(0.04)	(0.05)
zalando.nl	-0.61***	-0.61***	-0.66***	-0.79***
	(0.03)	(0.03)	(0.03)	(0.03)
Number of previous website visits	0.17***	0.17***	0.16***	0.16***
	(0.01)	(0.01)	(0.01)	(0.01)
Visit to a price discount page	1.78***	1.68***	1.57***	1.51***
	(0.05)	(0.05)	(0.05)	(0.05)
Outside option	2.26***	2.23***	2.20***	2.16***
	(0.03)	(0.03)	(0.03)	(0.03)
Search cost (exp)				
Constant	-3.70***	-3.80***	-3.85***	-3.58***
	(0.06)	(0.05)	(0.04)	(0.06)
Advertising				-2.04***
				(0.10)
Observations	27552	27552	27552	27552
LL	-7665	-8917	-9357	-9083
MAE (purchase)	29.07	31.96	34.26	32.92
MAE (search)	49.17	56.62	73.49	68.93
RMSE (purchase)	59.25	61.71	64.55	64.91
RMSE (search)	52.31	80.95	93.85	78.36

	Subcat. 4: "Underwear"			
	(i)	(ii)	(iii)	(iv)
	AP-strong	*AP-weak*	*Weitzman*	*Weitzman*
Utility				
asos.nl	-1.78***	-1.91***	-1.87***	-1.77***
	(0.07)	(0.06)	(0.07)	(0.06)
debijenkorf.nl	-1.46***	-1.57***	-1.56***	-1.47***
	(0.06)	(0.05)	(0.05)	(0.05)
happysocks.nl	-1.74***	-1.76***	-1.74***	-1.69***
	(0.07)	(0.06)	(0.06)	(0.06)
hm.com	-1.12***	-1.18***	-1.17***	-1.11***
	(0.05)	(0.04)	(0.05)	(0.04)
hunkemoller.nl	-0.65***	-0.42***	-0.49***	-0.65***
	(0.05)	(0.04)	(0.04)	(0.03)
livera.nl	-1.41***	-1.34***	-1.42***	-1.55***
	(0.06)	(0.05)	(0.05)	(0.05)
mona-mode.nl	-2.04***	-2.19***	-2.13***	-2.02***
	(0.12)	(0.11)	(0.10)	(0.09)
ullapopken.nl	-1.45***	-1.55***	-1.52***	-1.42***
	(0.06)	(0.05)	(0.05)	(0.05)
wibra.eu	-1.57***	-1.67***	-1.66***	-1.57***
	(0.06)	(0.05)	(0.06)	(0.06)
zalando.nl	-1.06***	-1.05***	-1.09***	-1.18***
	(0.05)	(0.05)	(0.04)	(0.04)
Number of previous website visits	0.12***	0.12***	0.10***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)
Visit to a price discount page	1.83***	1.77***	1.79***	1.77***
	(0.09)	(0.07)	(0.08)	(0.14)
Outside option	2.21***	2.20***	2.13***	2.11***
	(0.05)	(0.04)	(0.03)	(0.04)
Search cost (exp)				
Constant	-3.84***	-4.07***	-4.10***	-3.69***
	(0.06)	(0.07)	(0.06)	(0.08)
Advertising				-1.96***
				(0.09)
Observations	17988	17988	17988	17988
LL	-4466	-5294	-5681	-5497
MAD (purchase)	14.32	15.32	18.45	17.74
MAD (search)	24.05	25.71	49.99	44.63
RMSE (purchase)	25.92	27.05	32.57	32.39
RMSE (search)	26.14	36.14	57.28	49.13

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers in each subcategory.

11 Appendix

11.1 Data Cleaning

Our initial data source is the data used in Ursu et al. (2021). We further cleaned the data on apparel searches, which initially consisted of 428,651 clicks, 40,735 sessions and 5,665 spells, as follows. We dropped sessions if they consisted of all non-search clicks (59 sessions), given that in this paper our focus is on active versus passive searches. We defined clicks as "non-searches" if their URLs included information on: transaction-related activities (e.g., add to basket, checkout, and order confirmation), login/out decisions, actions related to managing/viewing users' accounts or subscriptions, finding/creating a password, locating a store, tracking a shipment, or reaching customer service.[40] Below are two examples of non-search clicks.

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www2.hm.com}}_{\text{website}}\underbrace{\text{/my-account/overview}}_{\text{path}}$$

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.missetam.nl}}_{\text{website}}\text{/}\underbrace{\text{orderdetails/page/880/}}_{\text{path}}\underbrace{\text{?orderid=3173951}}_{\text{parameter}}$$

These non-search clicks account for 10.20% of all consumer apparel clicks. In this paper, we describe the nature of search clicks.

Also, we dropped spells if the first session in a spell was dropped (i.e. if it contained only non-search clicks), since in such cases we may not observe the user's previous search activities (16 spells). Our final data sample consists of 427,768 apparel clicks with 40,625 sessions and 5,649 spells.

11.2 How Traffic is Tracked and Monetized Online

Consumers can arrive at a website in two distinct ways: either by visiting the website organically or by clicking on a paid ad link that refers them to the website. Web analytics services help companies track traffic to and from their website in order to allow referral websites to get paid for the ad traffic they generated, as well as to enable websites receiving such traffic to analyze it. For example, the New York Times website may refer consumers to the Nike website through a banner ad placed on its website.

[40] Non-search clicks that are transaction related identify the products purchased. For more details, see (Ursu et al., 2021).

For this referral, the New York Times expects payment. At the same time, Nike wants to understand whether its advertising campaigns ran on the New York Times website are effective, therefore it wants to be able to identify and track its campaigns across many sources (e.g. news websites, social media, email, etc). Web analytics companies can help both such companies.

As of 2019, Google Analytics was the most widely used web analytics service, with 86.2% market share, followed by Facebook.[41] More than 28.8 million websites used Google Analytics in October 2021, and the Netherlands, the country we obtained data from, is ranked 7th among countries worldwide for its frequent use of Google Analytics.[42]

Google Analytics, as well as other web analytics companies, works by modifying the URL address of a website in order to indicate the source of the traffic referred to it by an ad. It is necessary for ad companies to modify their URLs in order to pay referral websites for their traffic, as well as collect data on the effectiveness of their advertising campaigns. Google offers two ways of modifying URLs to identify ad clicks: automatic and manual.[43] For automatic ad tagging, Google appends a Google Click ID (gclid) to the end of a URL, before any fragments (i.e. in the parameter component of the URL).[44] An example is "http://www.example.com/?parameter=1&gclid=TeSter-123". Approximately 13% of ad-initiated clicks contain such automatic tags in our data.

For manual ad tagging, companies can customize their URLs by choosing up to 5 UTM parameters (and at least the first 3) to add to their URLs: utm_source (identifying the source sending traffic to the website, e.g., advertiser, site, publication), utm_medium (indicating the advertising or marketing medium, e.g., banner, email, newsletter), utm_campaign (describing the individual campaign name, slogan, promo code), utm_term (identifying paid search keywords), and utm_content (differentiating similar content or links of the same ad). An example of such a URL with manual tagging is: "https://www.example.com/?utm_source=summer-mailer&utm_medium=email&utm_campaign=summer-sale".[45] Among ad-initiated clicks, approximately 68% include UTM parameters in our data.

Other web analytics companies work the same way. For example, Microsoft Advertising also

[41] See the reports at w3techs.com/technologies/overview/traffic_analysis.

[42] Additional details can be found at ahrefs.com.

[43] For the use of Google Click Ids see support.google.com/google-ads/answer/9744275. For additional information on manual versus automatic ad click tagging see support.google.com/analytics/answer/1733663.

[44] See support.google.com/analytics/answer/2938246.

[45] Additional examples are described at support.google.com/analytics/answer/1033863.

allows advertisers either to auto tag with the Microsoft id (msclkid) or with the same 5 UTM tags.[46]. Similarly, in the case of Facebook, ads are tracked by Facebook identifiers and UTMs.[47].

11.3 Classifying Ads by Type

We classified ad-initiated searches into eight types using the parameter component of a URL:

1. display – if the parameter component included keywords such as: display, banner, image.
2. email – if the parameter component included keywords such as: email, e-mail, mail, gmail, outlook, live.com.
3. newsletter – if the parameter component included keywords such as: nwl, newsletter, nieuwsbrief.
4. retargeting – if the parameter component included keywords such as: retarget, remarket.
5. search engine – if the parameter component included keywords such as: (a) gclid, gclsrc, dclid, or msclkid, (b) search engine names like google, bing, and yahoo, or (c) cpc, seo, ppc, sem, engine.
6. social – if the parameter component included keywords such as: social, instagram, facebook, fb, twitter.
7. affiliate – if the parameter component did not include (1)-(6) related variables but did include affiliate advertisers' names or affiliate ids such as: affiliate, refid, affid, partnerid, zanox, awin, daisycon, tradetracker, criteo, adtraxx, affilinet, copernica, and zanpid. Affiliate ads that contained information on the type of third-party website the ad was placed on (e.g. email, search engine, social, etc), were reclassified to match the groups identified above. Finally, based on the click performed before an affiliate ad click, we classified affiliate ads placed on cashback websites (31%) as active searches. This is a conservative approach, since consumers may chose to search these cashback websites looking for apparel deals.
8. other – otherwise. For example, some URLs only included campaign ids, so we cannot identify the ad type.

[46]For information, see https://help.ads.microsoft.com.
[47]For reference, see facebook.com/business.

11.4 Classifying Ad Campaigns by Type

11.5 Types of ad campaign ids

We classified ad-initiated searches into several types:

1. Type 1: Season/calendar-related keywords (14.8% of ad-initiated searches):

    - e.g.: week, date, FW, SS, Easter

    > https://www.vanharen.nl/nl/nl/shop/home-dames/home-dames-schoenen.cat
    >
    > ecmid=545/2109& $\underbrace{\text{newsletter=mailing\_545/2018\_kw10/20180309/header\_collectie}}_{\text{campaign id}}$
    >
    > &ecmuid=217722&showall=true&filter-collection=fv_collection_springandsummer

2. Type 2: Category-related keywords (6.3% of ad-initiated searches):

    - women, men, kids, boys, girls
    - sales (e.g., sales, spring discount, top deals)
    - product category (e.g., dresses, bathing suit)
    - brand names (e.g., NIKE)
    - welcome/new registration, birthday, gift, celebration
    - low/mid/upper funnel
    - retargeting

    > https://www.zalando.nl/kinderen/gender=1& $\underbrace{\text{wmc=osm\_gc-aw16\_boys\_nl}}_{\text{campaign id}}$

3. Other (78.9% of ad-initiated searches):

    - Other ad-initiated searches cannot be easily classified by campaign id, since this id contains a string of undistinguishable numbers and/or letters, such as in the example below:

    > https://www.vanarendonk.nl/utm_source=daisycon&utm_medium=affiliate
    >
    > & $\underbrace{\text{utm\_campaign=daisycon\_182499}}_{\text{campaign id}}$ &utm_content=algemeen

11.6 Estimation Samples

We constructed four estimation samples corresponding to the four most commonly purchased product subcategories in our data: (1) "shirts, tops, and blouses", (2) "shoes", (3) "pants and jeans", and (4) "underwear". For each subcategory, we determined the top 10 most searched websites (accounting for approximately 65% of clicks in each subcategory), for which we estimated website intercepts. All other websites were grouped together into a composite website which we call "Other."

Since neither our AP models, nor the Weitzman model can accommodate revisits, we removed search revisits (to the a website) from the data, accounting for approximately 30% of observations. Also, we removed spells that had a search session within the last two days of our observation period, but no transaction, in order to avoid concerns about right truncation.[48] A small fraction (less than 1%) of spells contained more than one product purchased (after the changes we made to the samples), which we removed from the sample. In estimation, an ad-initiated search is a website where the arrival to the website (first click) was passive. The resulting estimation samples can be summarized as follows (Table A-1):[49]

Table A-1: Summary Statistics by Estimation Sample

	Subcat. 1	Subcat. 2	Subcat. 3	Subcat. 4
Observations	32,422	34,812	27,552	17,988
Spells	2,702	2,901	2,296	1,499
Converting Spells	359	316	271	152
Ads Searched	1,526	3,068	872	706
Ads Searched First/Ads Searched	0.63	0.50	0.80	0.78

In our data we only observe ads consumers clicked on. However, to more accurately capture the magnitude of the effect of passive search and to be able to consider an effect of advertising on search costs (the Weitzman model with advertising costs), we need understand the extent to which consumers might have been exposed to ads. We assume a consumer i was exposed to an ad from website j that she did not click if all of the following criteria are met: (i) consumer i clicked on an ad from website j in a different subcategory in the past, increasing her likelihood of receiving ads from the same website; (ii) consumer i had an open account with website j, increasing her likelihood of email and newsletter ads; (iii) website j advertises extensively in a given subcategory (more than

[48] Spells that end within the first week of our observation period (before February 23rd, 2018) were dropped from the original data sample, alleviating concerns about left truncation.

[49] Note that the reported number of observations includes an outside option for each spell.

the 90th percentile of the ad distribution in a subcategory), increasing the consumer's probability of being exposed to ads from this website; and (iv) consumer *i* clicked on at least one ad in the current spell, suggesting the consumer may be more likely to be exposed to ads (for example because she does not use ad blocker software; also, this allows us to be more conservative in our approach to infer ad exposure). For robustness, we also estimate our proposed model on the raw data, without any ads on websites that were not searched, and show that our results continue to hold. These results can be found in Table A-5 in Appendix 11.10.

In estimation, any ads that were not searched will be assumed to have occurred after the last searched website. To show that this assumption is in most cases innocuous, we use two approaches. First, we note that our robustness check, estimating the model on the raw data without any ads on websites that were not searched, also provides a robustness check for this assumption. Second, we demonstrate analytically in what narrow set of cases this assumption fails. If a consumer was exposed to ads she did not click on earlier in the search process than after the last searched website, then it means she searched other options after ad exposure. Let's denote by *ad* the ad the consumer did not click on, and by *next* any such options she searched after the ad she did not click on. In both AP models, if the consumer does not search an ad, then it must be that

$$z_{ad} < y, \tag{A1}$$

where y denotes the best option searched up to that moment in the search process. In contrast, because the consumer has searched an option after the ad she did not click on, then it must be that $z_{next} \geq y$. Thus, we conclude that

$$z_{next} \geq z_{ad}. \tag{A2}$$

Using this same logic for every website searched after the ad, we conclude that the ad the consumer did not click on has a lower reservation utility than all searched websites. This means that although the ad may have been shown earlier in the search, assuming it was presented to the consumer after all other searched websites will not produce a bias in the order of reservation utilities.

A bias may arise only because the reservation utility of the ad not clicked is compared against the utility of additional options when we assume it was presented to the consumer after all searched websites, rather than earlier. However, those additional options have reservation utilities z_{next} that are higher than z_{ad}, making it likely that their utilities are also higher than z_{ad} (since $z_j = u_j - \epsilon_j + fcn(c) -$

see Kim et al. (2010) for more details on the functional form of reservation utilities), thus not affecting the set of inequalities that identify our parameters. Only in the unlikely event that $fcn(c) - \epsilon_{next}$ is very large (e.g. very low search costs or very low utility shock draw), then our assumption would lead to a higher upper bound on reservation utilities z_{ad} than if we had observed ad exposure timing (which *may* lead to a higher reservation utility estimate, but does not need to). Given that we do not observe ad exposure timing, assuming consumers were exposed to ads they did not click on after the last searched website produces minimal (if any) bias in parameter estimates. Also, we note that our assumption is preferred over other alternatives, such as random or early exposure timing, because it does not disrupt the true order of reservation utilities, as demonstrated.

11.7 Estimation Procedure

The estimation procedure using the logit-smoothed AR simulator is standard in the literature (e.g. Honka and Chintagunta, 2017; Ursu, 2018) and involves the following steps for the Weitzman (1979) model:

1. Make $d = \{1, \ldots, D\}$ draws of η_{ij} and ϵ_{ij} for each consumer-website combination and calculate utility u_{ij}^d.
2. Compute z_j^d using the method proposed by Kim et al. (2010).
3. Calculate the following expressions for each draw d:
   (a) $\nu_1^d = z_{in}^d - \max_{k=n+1}^{J} z_{ik}^d \qquad \forall n \in \{1, \ldots, J-1\}$
   (b) $\nu_2^d = z_{in}^d - \max_{k=0}^{n-1} u_{ik}^d \qquad \forall n \in \{1, \ldots, s\}$
   (c) $\nu_3^d = \max_{k=0}^{s} u_{ik}^d - z_{im}^d \qquad \forall m \in \{s+1, \ldots, J\}$
   (d) $\nu_4^d = u_{ij}^d - \max_{k=0}^{s} u_{ik}^d \qquad \forall j \in \{0, 1, \ldots, s\}$
4. Compute $V^d = \frac{1}{1+M^d}$ for each draw d, where

$$M^d = \sum_{k=1}^{4} e^{-\nu_k^d/\rho} \tag{A3}$$

   and where ρ is a scaling parameter, chosen using Monte Carlo simulations. In our application, the scaling parameter equals $\rho = -15$.

5. The average of V^d over the D draws and over consumers and websites gives the simulated likelihood function.

It is straightforward to modify the above expressions for the AP-weak and the AP-strong models using the discussion in Section 6.2.

11.8 Computing Reservation Utilities and Additional Identification Issues

11.8.1 Computing Reservation Utilities

Recall from Section 6 that consumer i's utility of purchasing from website j is defined as (equation 6)

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij} \\ &= w_j + \gamma X_{ij} + \eta_{ij} + \epsilon_{ij} \end{aligned} \tag{A4}$$

The value v_{ij} denotes the part of utility that the consumer observes without search, while ϵ_{ij} denotes the match value revealed through search. Let η_{ij} be normally distributed with mean zero and standard deviation ω_{ij}, and let ϵ_{ij} be normally distributed with mean zero and standard deviation σ_{ij}.[50]

Given search costs c_{ij}, we know that reservation utilities are the solution to (equation 5)

$$c_{ij} = \int_{z_{ij}}^{\infty} (u - z_{ij}) dF_{ij}(u). \tag{A5}$$

In words, the reservation utility z_{ij} is the value that equates the marginal cost of searching with the expected marginal benefit. Kim et al. (2010) show how to compute these values under similar distributional assumptions to those we have made in Section 6.2. More precisely, they show that we can compute reservation utilities under the assumption of normally distributed utilities from

$$z_{ij} = v_{ij} + m_{ij}\sigma_{ij}, \tag{A6}$$

where the value of m_{ij} is obtained by solving

$$\frac{c_{ij}}{\sigma_{ij}} = \phi(m_{ij}) + m_{ij}\Phi(m_{ij}) - m_{ij}. \tag{A7}$$

A unique solution to this equation exists (see Weitzman (1979)), so to compute reservation utilities we can invert this relation, solve for m_{ij}, and the compute the reservation utility from $z_{ij} = v_{ij} + m_{ij}\sigma_{ij}$.

11.8.2 Identifying the Match Value Standard Deviation σ_{ij}

From equation A7 above we can see that consumer behavior (i.e. the number of searches – see Section 6.3) only identifies the ratio $\frac{c_{ij}}{\sigma_{ij}}$. Thus, without an exogenous search cost shifter, the search cost level

[50]Note that this model is more general than the one presented in Section 6, where both ω_{ij} and σ_{ij} are normalized to 1, for reasons we explain below.

is not identified (Yavorsky et al., 2021). For this reason, all prior work on consumer search, except Yavorsky et al. (2021), has fixed the standard deviation of the match value revealed through search, σ_{ij}, to 1 (see (e.g., Kim et al., 2010; De los Santos et al., 2012; Honka, 2014; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018; Ursu et al., 2020, 2021)). Yavorsky et al. (2021) use data on offline search (consumers searching for cars offline) and have access to an exogenous search cost shifter (distance from a consumer's home to a car dealership). However, online such exogenous search cost shifters are not readily available, so in our paper we will also normalize σ_{ij} to 1, consistent with all prior work that uses online data.[51]

11.8.3 Identifying the Observed Utility Standard Deviation ω_{ij}

The parameter ω_{ij} affects the part of utility that the consumer observes before search, v_{ij}. All prior work on consumer search, including Yavorsky et al. (2021), fix this parameter to 1. This is because v_{ij} is only identified up to scale, as is the case in standard logit or probit models (see Cameron and Heckman, 1998; Maddala, 1983; Breen et al., 2018). It is easy to see this in our model.

Consider the selection rule for two arbitrary options, j and k, and suppose the consumer searches j before k. In this case, we want to write the probability that z_{ij} is greater than z_{ik}. To simplify notation, we will consider writing the probability that the difference between these two terms is positive as $Pr(\Delta z > 0)$. Then

$$\begin{aligned} Pr(\Delta z > 0) &= Pr(\beta\Delta\tilde{X} + \omega\Delta\eta > 0) \\ &= Pr(\frac{\beta}{\omega}\Delta\tilde{X} + \Delta\eta > 0) \end{aligned} \tag{A8}$$

where η is a standard normally distributed variable and where, without loss of generality, we have collected terms $w_j + \gamma X_{ij} + m_{ij}\sigma_{ij}$ into a composite term $\beta\tilde{X}_{ij}$ to emphasize the relation of interest.[52] We can immediately see that only the ratio between observed product features and the standard deviation ω can be identified. The same holds when considering the search and the choice rules. Thus, we will follow prior work and normalize ω to 1 (Kim et al., 2010; De los Santos et al., 2012; Honka, 2014; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018; Ursu et al., 2020, 2021; Yavorsky et al., 2021).

[51] As the authors Yavorsky et al. (2021) also emphasize, "[...] it is hard to imagine an exogenous search cost shifter in the online context" (page 3).

[52] Also, as explained above, σ_{ij} needs to be fixed to 1 for identification purposes, and search costs are constant across options in our empirical application. Thus, the term $m_{ij}\sigma_{ij}$ drops from the difference in equation A8.

11.9 Accounting for a Persuasive Ad Effect

In addition to accounting for an informative ad effect, we also run a robustness check to see if ads have a persuasive effect entering directly into the consumer utility function. As explained in Section 8.1, firms may target their ads to specific consumers. If ads have a persuasive effect and enter the consumer utility function, then a commonly used approach to dealing with advertising endogeneity is to use a control function approach (Petrin A, 2010; Honka et al., 2017; De los Santos and Koulayev, 2017). We will employ this approach. More precisely, we model the ad as a linear function of all relevant exogenous variables, denoted by X, instruments Z, that do not enter a consumer's utility directly, but that could affect advertising, and unobserved error ξ_{ij}:

$$ad_{ij} = w_j + \beta X_{ij} + \theta Z_i + \xi_{ij}. \tag{A9}$$

As exogenous variables X, we include all variables that enter the utility function: website fixed effects, number of previous website visits, and visits to a price discount page. As instruments Z_i, we use the total number of websites visited in a session and indicators for visits to social networking sites and email, since time spent online and visits to those websites are likely to affect advertising exposure, but not purchase utility directly. We also control for consumer fixed effects, the size of the product category, and include an indicator for when/if the consumer creates an account on a website.[53] Our results can be found in Table A-2 below. The adjusted R-squared in this regression equals 0.35, and the F-test for joint significance of our instruments equals 57.59, suggesting that the instruments we chose are valid.

We then obtain the predicted residual ξ_{ij} by estimating equation A9 and include it in the utility function, along with the advertising variable, to account for endogeneity. Because we are using an estimate of ξ_{ij}, rather than the true ξ_{ij} in our model, we report standard errors obtained using bootstrapping methods. Our results can be found in Tables A-3 and A-4 (column 2). Here we find a positive effect of advertising on utility and a worse fit than in our main models (relevant comparison results are in Tables 6 and 7, columns 1 and 2). Additionally, we find a larger effect of advertising on utility than when we do not account for endogeneity, with a mostly negative effect of the residual. This result suggests that websites advertise more to consumers that generally are unlikely to respond to their ads. Therefore, once we account for this selection, we observe a larger effect of advertising.

[53] The specification in equation (A9) above mirrors the one in equation (15) in Section 8.1, except that it is able to control for additional variables that would either drop out of equation (A9) (e.g. total number of websites visited in a session and indicators for visits to social networking sites and email) or could not be included in equation (A9) due to computational constraints (e.g. consumer fixed effects).

Table A-2: Predicting Advertising Decisions

	Advertising	
Instruments		
(OLS) Number of website visits in session	-0.0015***	(0.0002)
Visit to social network	0.0095***	(0.0015)
Visit to email	0.0164***	(0.0018)
Utility		
Visit to a price discount page	-0.0012	(0.0017)
Number of previous website visits	0.0002	(0.0001)
Controls		
Website account	-0.0721***	(0.0016)
Size of category	0.0000***	(0.0000)
Constant	0.1232***	(0.0023)
Consumer FE	Yes	
Website FE	Yes	
R-squared	0.35	
F-test (instruments)	57.59	
Observations	367,285	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

11.10 Additional Results

Table A-3: Estimation Results with a Persuasive Ad Effect

	Subcat. 1: "Shirts, tops, & blouses"				Subcat. 2: "Shoes"		
	(i)	(ii)	(iii)		(i)	(ii)	(iii)
	persuasive	persuasive (CF)	persuasive &informative		persuasive	persuasive (CF)	persuasive &informative
	Weitzman	*Weitzman*	*AP-strong*		*Weitzman*	*Weitzman*	*AP-strong*
Utility				*Utility*			
aboutyou.com	-1.40***	-1.44***	-1.38***	adidas.com	-0.89***	-0.82***	-0.88***
	(0.03)	(0.07)	(0.04)		(0.04)	(0.08)	(0.03)
c-and-a.com	-0.88***	-0.92***	-0.87***	debijenkorf.nl	-1.51***	-1.44***	-1.46***
	(0.03)	(0.07)	(0.03)		(0.05)	(0.11)	(0.04)
debijenkorf.nl	-1.73***	-1.77***	-1.69***	nelson.nl	-1.17***	-1.11***	-1.2***
	(0.04)	(0.08)	(0.04)		(0.05)	(0.11)	(0.04)
esprit.nl	-1.63***	-1.62***	-1.62***	nike.com	-0.93***	-0.87***	-0.92***
	(0.05)	(0.09)	(0.04)		(0.04)	(0.09)	(0.03)
hm.com	-1.18***	-1.16***	-1.18***	omoda.nl	-1.23***	-1.16***	-1.26***
	(0.04)	(0.08)	(0.03)		(0.05)	(0.10)	(0.04)
jbfo.nl	-2.47***	-2.44***	-2.43***	**schuurman-shoenen.nl**	-1.02***	-1.17***	-0.86***
	(0.12)	(0.18)	(0.11)		(0.03)	(0.09)	(0.03)
msmode.nl	-1.64***	-1.62***	-1.66***	**spartoo.nl**	-1.49***	-1.66***	-1.39***
	(0.05)	(0.09)	(0.05)		(0.03)	(0.10)	(0.03)
peterhahn.nl	-1.76***	-1.74***	-1.71***	**vanharen.nl**	-1.10***	-1.26***	-1.00***
	(0.06)	(0.14)	(0.06)		(0.03)	(0.09)	(0.03)
your-look-for-less.nl	-1.37***	-1.35***	-1.36***	zalando.nl	-0.53***	-0.47***	-0.44***
	(0.04)	(0.09)	(0.04)		(0.04)	(0.09)	(0.03)
zalando.nl	-1.05***	-1.03***	-1.08***	ziengs.nl	-1.41***	-1.34***	-1.48***
	(0.04)	(0.08)	(0.04)		(0.05)	(0.10)	(0.05)
Number of previous	0.16***	0.16***	0.20***	Number of previous	0.20***	0.19***	0.20***
website visits	(0.01)	(0.02)	(0.01)	website visits	(0.02)	(0.02)	(0.01)
Visit to a	1.62***	1.60***	1.88***	Visit to a	0.98***	0.87***	1.36***
price discount page	(0.04)	(0.08)	(0.04)	price discount page	(0.04)	(0.13)	(0.04)
Advertising	0.67***	0.79***	0.39***	Advertising	0.93***	1.24***	0.81***
	(0.03)	(0.06)	(0.03)		(0.02)	(0.07)	(0.02)
Residual		-0.16*		Residual		-0.41***	
		(0.08)				(0.08)	
Outside option	2.13***	2.16***	2.15***	Outside option	2.55***	2.64***	2.74***
	(0.04)	(0.1)	(0.03)		(0.04)	(0.12)	(0.02)
Search cost (exp)				*Search cost (exp)*			
Constant	-4.23***	-4.24***	-4.01***	Constant	-5.28***	-5.27***	-5.53***
	(0.07)	(0.19)	(0.05)		(0.08)	(0.24)	(0.04)
Observations	32422	32422	32422	Observation	34812	34812	34812
LL	-11236	-11230	-8921	LL	-17373	-17291	-12592
MAE (purchase)	39.51	39.11	35.36	MAD (purchase)	44.19	43.97	34.61
MAE (search)	67.22	67.38	54.03	MAD (search)	128.74	132.90	83.51
RMSE (purchase)	67.98	67.23	60.79	RMSE (purchase)	71.67	70.17	57.40
RMSE (search)	77.06	77.94	60.58	RMSE (search)	140.84	144.91	94.39

Standard errors in parentheses.

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

Note: Websites in bold identify the three largest advertisers in each subcategory. Standard errors in the second column for each set of results are computed using bootstrapping with 100 estimation samples.

Table A-4: Estimation Results with a Persuasive Ad Effect (continued)

	Subcat. 3: "Pants & Jeans"				Subcat. 4: "Underwear"		
	(i)	(ii)	(iii)		(i)	(ii)	(iii)
	persuasive	persuasive (CF)	persuasive &informative		persuasive	persuasive (CF)	persuasive &informative
	Weitzman	*Weitzman*	*AP-strong*		*Weitzman*	*Weitzman*	*AP-strong*
Utility				*Utility*			
c-and-a.com	-0.78***	-0.80***	-0.76***	asos.nl	-1.77***	-1.77***	-1.79***
	(0.03)	(0.11)	(0.03)		(0.06)	(0.13)	(0.07)
debijenkorf.nl	-1.36***	-1.37***	-1.35***	debijenkorf.nl	-1.47***	-1.46***	-1.46***
	(0.04)	(0.13)	(0.05)		(0.04)	(0.12)	(0.06)
esprit.nl	-1.30***	-1.32***	-1.32***	happysocks.nl	-1.67***	-1.67***	-1.73***
	(0.05)	(0.14)	(0.05)		(0.06)	(0.14)	(0.07)
g-star.com	-1.80***	-1.83***	-1.80***	hm.com	-1.09***	-1.09***	-1.12***
	(0.07)	(0.17)	(0.07)		(0.03)	(0.11)	(0.05)
hm.com	-0.86***	-0.88***	-0.85***	**hunkemoller.nl**	-0.61***	-0.64***	-0.67***
	(0.03)	(0.11)	(0.04)		(0.03)	(0.10)	(0.04)
jeanscentre.nl	-1.59***	-1.59***	-1.51***	**livera.nl**	-1.51***	-1.56***	-1.42***
	(0.05)	(0.14)	(0.06)		(0.05)	(0.13)	(0.06)
missetam.nl	-0.61***	-0.62***	-0.92***	mona-mode.nl	-2.02***	-2.01***	-2.03***
	(0.03)	(0.1)	(0.04)		(0.09)	(0.17)	(0.10)
tommy.com	-1.90***	-1.94***	-1.97***	ullapopken.nl	-1.41***	-1.40***	-1.44***
	(0.08)	(0.18)	(0.11)		(0.05)	(0.12)	(0.06)
your-look-for-less.nl	-1.20***	-1.22***	-1.20***	wibra.eu	-1.57***	-1.56***	-1.57***
	(0.04)	(0.12)	(0.05)		(0.05)	(0.13)	(0.07)
zalando.nl	-0.74***	-0.74***	-0.61***	**zalando.nl**	-1.13***	-1.16***	-1.07***
	(0.03)	(0.11)	(0.03)		(0.03)	(0.10)	(0.05)
Number of previous	0.16***	0.16***	0.17***	Number of previous	0.09***	0.09***	0.12***
website visits	(0.01)	(0.02)	(0.01)	website visits	(0.01)	(0.02)	(0.01)
Visit to a	1.53***	1.51***	1.76***	Visit to a	1.80***	1.81***	1.83***
price discount page	(0.05)	(0.12)	(0.06)	price discount page	(0.07)	(0.17)	(0.09)
Advertising	0.64***	0.62***	0.05	Advertising	0.62***	0.69***	0.11*
	(0.04)	(0.08)	(0.04)		(0.03)	(0.09)	(0.05)
Residual		0.06		Residual		-0.08	
		(0.09)				(0.10)	
Outside option	2.32***	2.30***	2.27***	Outside option	2.32***	2.32***	2.23***
	(0.04)	(0.14)	(0.03)		(0.03)	(0.14)	(0.04)
Search cost (exp)				*Search cost (exp)*			
Constant	-3.93***	-3.91***	-3.72***	Constant	-4.20***	-4.20***	-3.84***
	(0.08)	(0.19)	(0.05)		(0.06)	(0.30)	(0.06)
Observations	27552	27552	27552	Observation	17988	17988	17988
LL	-9157	-9151	-7664	LL	-5543	-5540	-4462
MAE (purchase)	33.23	33.81	28.75	MAD (purchase)	17.35	17.47	14.38
MAE (search)	66.42	64.41	49.46	MAD (search)	39.15	39.80	23.49
RMSE (purchase)	64.10	65.48	58.64	RMSE (purchase)	30.38	30.66	26.08
RMSE (search)	76.50	74.86	52.89	RMSE (search)	41.72	42.53	25.31

Standard errors in parentheses.

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

Note: Websites in bold identify the three largest advertisers in each subcategory. Standard errors in the second column for each set of results are computed using bootstrapping with 100 estimation samples.

Table A-5: Robustness Checks

	(i)	(ii)		(iii)		(iv)		(v)
	AP-strong: search engine-active	*AP-strong: raw data*						
	Subcat. 1	Subcat. 1	Subcat. 2		Subcat. 3		Subcat. 4	
Utility			*Utility*		*Utility*		*Utility*	
aboutyou.com	-1.28***	-1.30	adidas.com	-0.96***	c-and-a.com	-0.77***	asos.nl	-1.78***
	(0.09)	(0.04)		(0.04)		(0.04)		(0.07)
c-and-a.com	-0.77***	-0.79***	debijenkorf.nl	-1.64***	debijenkorf.nl	-1.35***	debijenkorf.nl	-1.47***
	(0.07)	(0.03)		(0.04)		(0.06)		(0.05)
debijenkorf.nl	-1.64***	-1.64***	nelson.nl	-1.33***	esprit.nl	-1.33***	happysocks.nl	-1.73***
	(0.06)	(0.04)		(0.04)		(0.06)		(0.06)
esprit.nl	-1.67***	-1.65***	nike.com	-1.08***	g-star.com	-1.81***	hm.com	-1.12
	(0.09)	(0.05)		(0.03)		(0.08)		(0.04)
hm.com	-1.22***	-1.19***	omoda.nl	-1.40***	hm.com	-0.85***	**hunkemoller.nl**	-0.67***
	(0.06)	(0.03)		(0.04)		(0.05)		(0.04)
jbfo.nl	-2.45***	-2.44***	**schuurman-shoenen.nl**	-0.69***	**jeanscentre.nl**	-1.50***	**livera.nl**	-1.43***
	(0.18)	(0.13)		(0.03)		(0.07)		(0.06)
msmode.nl	-1.71***	-1.68***	**spartoo.nl**	-1.14***	**missetam.nl**	-0.92***	mona-mode.nl	-2.02***
	(0.14)	(0.04)		(0.04)		(0.04)		(0.10)
peterhahn.nl	-1.76***	-1.72***	**vanharen.nl**	-0.84***	tommy.com	-1.98***	ullapopken.nl	-1.46***
	(0.10)	(0.06)		(0.03)		(0.12)		(0.05)
your-look-for-less.nl	-1.41***	-1.39***	zalando.nl	-0.58***	your-look-for-less.nl	-1.21***	wibra.eu	-1.57***
	(0.05)	(0.04)		(0.03)		(0.05)		(0.06)
zalando.nl	-1.07***	-1.08***	ziengs.nl	-1.59***	**zalando.nl**	-0.63***	**zalando.nl**	-1.07***
	(0.07)	(0.03)		(0.04)		(0.04)		(0.05)
Number of previous website visits	0.20***	0.20***	Number of previous website visits	0.21***	Number of previous website visits	0.17***	Number of previous website visits	0.12***
	(0.01)	(0.01)		(0.02)		(0.01)		(0.01)
Visit to a price discount page	1.90***	1.94***	Visit to a price discount page	1.45***	Visit to a price discount page	1.76***	Visit to a price discount page	1.84***
	(0.07)	(0.04)		(0.05)		(0.06)		(0.08)
Outside option	2.06***	2.07***	Outside option	2.34***	Outside option	2.26***	Outside option	2.20***
	(0.04)	(0.04)		(0.04)		(0.04)		(0.05)
Search cost (exp)			*Search cost (exp)*		*Search cost (exp)*		*Search cost (exp)*	
Constant	-4.00***	-3.99***	Constant	-5.16***	Constant	-3.73***	Constant	-3.84***
	(0.11)	(0.06)		(0.07)		(0.07)		(0.07)
Observations	32422	32422	Observations	34812	Observations	27552	Observations	17988
LL	-9327	-9023	LL	-13242	LL	-7689	LL	-4490

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers in each subcategory.

Table A-6: Estimation Results Accounting for Ad Targeting

	Subcat. 1: "Shirts, tops, & blouses"					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
	$\hat{J}$ = *all*, $\sigma_\mu = 3$	$\hat{J}$ = *all*, $\sigma_\mu = 0.1$	$\hat{J}$ = *notsearched*, $\sigma_\mu = 3$	$\hat{J}$ = *notsearched*, $\sigma_\mu = 0.1$	$\hat{J}$ = *searched*, $\sigma_\mu = 3$	$\hat{J}$ = *searched*, $\sigma_\mu =$
Utility						
aboutyou.com	-0.37***	-0.18***	-0.89***	-0.65***	-0.58***	-0.13***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)
c-and-a.com	-0.20***	-0.10***	-0.53***	-0.40***	-0.37***	-0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
debijenkorf.nl	-0.72***	-0.36***	-1.18***	-0.89***	-0.80***	-0.30***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
esprit.nl	-1.95***	-1.41***	-2.45***	-1.84***	-1.51***	-1.27***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)
hm.com	-1.30***	-0.98***	-1.50***	-1.26***	-1.07***	-0.86***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
jbfo.nl	-4.46***	-2.34***	-10.20***	-2.81***	-2.34***	-2.22***
	(1.28)	(0.12)	(2.15)	(0.11)	(0.15)	(0.13)
msmode.nl	-2.17***	-1.45***	-2.48***	-1.83***	-1.54***	-1.35***
	(0.06)	(0.04)	(0.08)	(0.04)	(0.06)	(0.04)
peterhahn.nl	-2.38***	-1.72***	-3.40***	-2.01***	-1.68***	-1.66***
	(0.09)	(0.06)	(0.28)	(0.05)	(0.06)	(0.05)
your-look-for-less.nl	-1.66***	-1.30***	-1.91***	-1.56***	-1.32***	-1.19***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
zalando.nl	-1.19***	-0.88***	-1.36***	-1.14***	-0.95***	-0.74***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Number of previous	0.21***	0.12***	0.24***	0.18***	0.17***	0.12***
website visits	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Visit to a	1.58***	0.82***	2.24***	1.64***	1.25***	0.66***
price discount page	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Outside option	1.40***	0.19***	2.22***	2.09***	1.14***	-0.20***
	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.02)
Search cost (exp)						
Constant	-2.50***	-1.32***	-4.28***	-4.01***	-2.28***	-0.85***
	(0.05)	(0.03)	(0.05)	(0.07)	(0.04)	(0.03)
Observations	32422	32422	32422	32422	32422	32422
LL	-38907	-23845	-22161	-14911	-19377	-20171
MAE (purchase)	89.44	156.39	45.26	47.53	94.94	171.43
MAE (search)	109.19	113.92	67.19	81.17	81.26	112.58
RMSE (purchase)	138.90	231.36	67.27	71.08	140.26	246.46
RMSE (search)	150.02	175.94	94.06	118.93	117.72	188.11

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers.

Table A-7: Estimation Results Accounting for Ad Targeting (continued)

	Subcat. 2: "Shoes"					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
	$\hat{J}$ = *all*, $\sigma_\mu = 3$	$\hat{J}$ = *all*, $\sigma_\mu = 0.1$	$\hat{J}$ = *notsearched*, $\sigma_\mu = 3$	$\hat{J}$ = *notsearched*, $\sigma_\mu = 0.1$	$\hat{J}$ = *searched*, $\sigma_\mu = 3$	$\hat{J}$ = *searched*, σ_μ
Utility						
adidas.com	-1.32***	-0.83***	-1.48***	-1.16***	-0.69***	-0.67***
	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
debijenkorf.nl	-2.09***	-1.32***	-2.93***	-1.88***	-1.29***	-1.08***
	(0.06)	(0.04)	(0.07)	(0.05)	(0.05)	(0.03)
nelson.nl	-1.92***	-1.05***	-2.26***	-1.48***	-1.02***	-0.91***
	(0.06)	(0.03)	(0.06)	(0.04)	(0.04)	(0.03)
nike.com	-1.49***	-0.98***	-1.81***	-1.42***	-0.81***	-0.75***
	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
omoda.nl	-2.03***	-1.18***	-2.51***	-1.61***	-1.11***	-0.98***
	(0.08)	(0.04)	(0.07)	(0.05)	(0.05)	(0.04)
schuurman-shoenen.nl	-0.01	0.05*	-0.50***	-0.4***	0.02	0.20***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
spartoo.nl	-0.15***	0.01	-0.73***	-0.52***	-0.14***	0.18***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
vanharen.nl	-0.07*	0.03	-0.61***	-0.44***	-0.09	0.18***
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
zalando.nl	-0.79***	-0.56***	-1.08***	-0.81***	-0.40***	-0.32***
	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
ziengs.nl	-2.39***	-1.26***	-3.02***	-1.70***	-1.32***	-1.13***
	(0.07)	(0.04)	(0.08)	(0.05)	(0.06)	(0.04)
Number of previous	0.27***	0.14***	0.34***	0.23***	0.17***	0.11***
website visits	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Visit to a	1.49***	0.82***	2.24***	1.42***	0.97***	0.61***
price discount page	(0.05)	(0.03)	(0.05)	(0.04)	(0.04)	(0.03)
Outside option	1.31***	0.23***	2.43***	2.23***	1.06***	-0.14***
	(0.04)	(0.02)	(0.03)	(0.05)	(0.03)	(0.02)
Search cost (exp)						
Constant	-2.41***	-1.50***	-5.42***	-4.91***	-2.12***	-0.97***
	(0.05)	(0.02)	(0.06)	(0.08)	(0.03)	(0.02)
Observations	34812	34812	34812	34812	34812	34812
LL	-61826	-35425	-40562	-23926	-32990	-28975
MAE (purchase)	112.11	184.18	43.25	51.28	135.39	200.94
MAE (search)	115.1	102.47	91.23	100.67	90.31	103.40
RMSE (purchase)	179.78	256.37	71.17	76.88	181.49	273.82
RMSE (search)	136.11	140.27	108.65	120.44	116.03	160.90

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers.

Table A-8: Estimation Results Accounting for Ad Targeting (continued)

	Subcat. 3: "Pants & Jeans"					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
	$\hat{J} = all, \sigma_\mu = 3$	$\hat{J} = all, \sigma_\mu = 0.1$	$\hat{J} = notsearched, \sigma_\mu = 3$	$\hat{J} = notsearched, \sigma_\mu = 0.1$	$\hat{J} = searched, \sigma_\mu = 3$	$\hat{J} = searched, \sigma_\mu =$
Utility						
c-and-a.com	-0.87***	-0.68***	-1.01***	-0.87***	-0.71***	-0.57***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
debijenkorf.nl	-1.52***	-1.10***	-1.84***	-1.45***	-1.25***	-0.99***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
esprit.nl	-1.55***	-1.17***	-1.86***	-1.46***	-1.22***	-1.08***
	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
g-star.com	-2.28***	-1.68***	-2.90***	-1.97***	-1.78***	-1.62***
	(0.1)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
hm.com	-1.00***	-0.74***	-1.18***	-0.98***	-0.78***	-0.62***
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
jeanscentre.nl	-0.48***	-0.18***	-0.95***	-0.67***	-0.74***	-0.15***
	(0.04)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)
missetam.nl	-0.19***	-0.05*	-0.39***	-0.23***	-0.57***	-0.13***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)
tommy.com	-2.71***	-1.78***	-3.74***	-2.05***	-1.95***	-1.70***
	(0.12)	(0.08)	(0.05)	(0.08)	(0.1)	(0.09)
your-look-for-less.nl	-1.40***	-1.14***	-1.65***	-1.36***	-1.17***	-1.04***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
zalando.nl	-0.11***	0.00	-0.49***	-0.36***	-0.29***	0.04
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Number of previous	0.14***	0.10***	0.18***	0.15***	0.14***	0.10***
website visits	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Visit to a	1.55***	0.96***	1.88***	1.63***	1.15***	0.78***
price discount page	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Outside option	1.67***	0.44***	2.39***	2.31***	1.39***	0.01
	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Search cost (exp)						
Constant	-2.51***	-1.31***	-3.88***	-3.78***	-2.27***	-0.80***
	(0.06)	(0.03)	(0.05)	(0.07)	(0.04)	(0.03)
Observations	27552	27552	27552	27552	27552	27552
LL	-27448	-18199	-16410	-11862	-13740	-15660
MAE (purchase)	66.33	130.86	33.38	35.75	73.19	145.11
MAE (search)	89.18	95.97	61.92	70.17	58.49	90.27
RMSE (purchase)	106.83	185.58	58.66	59.82	107.70	199.48
RMSE (search)	125.12	151.50	91.87	112.74	84.72	155.15

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers.

Table A-9: Estimation Results Accounting for Ad Targeting (continued)

	Subcat. 4: "Underwear"					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
	$\hat{J} = all, \sigma_\mu = 3$	$\hat{J} = all, \sigma_\mu = 0.1$	$\hat{J} = notsearched, \sigma_\mu = 3$	$\hat{J} = notsearched, \sigma_\mu = 0.1$	$\hat{J} = searched, \sigma_\mu = 3$	$\hat{J} = searched, \sigma_\mu =$
Utility						
asos.nl	-2.22***	-1.51***	-2.74***	-2.00***	-1.64***	-1.34***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
debijenkorf.nl	-1.63***	-1.13***	-2.03***	-1.63***	-1.34***	-0.98***
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
happysocks.nl	-2.10***	-1.43***	-2.43***	-1.80***	-1.65***	-1.34***
	(0.08)	(0.05)	(0.07)	(0.05)	(0.07)	(0.06)
hm.com	-1.21***	-0.87***	-1.46***	-1.23***	-1.00***	-0.78***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
hunkemoller.nl	-0.04	-0.01	-0.35***	-0.25***	-0.40***	-0.06
	(0.02)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)
livera.nl	-0.37***	-0.17***	-0.99***	-0.74***	-0.72***	-0.14***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
mona-mode.nl	-3.66***	-1.87***	-4.07***	-2.29***	-1.91***	-1.72***
	(0.11)	(0.08)	(0.07)	(0.08)	(0.1)	(0.08)
ullapopken.nl	-1.72***	-1.26***	-2.01***	-1.64***	-1.39***	-1.13***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
wibra.eu	-1.79***	-1.36***	-2.21***	-1.76***	-1.50***	-1.20***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
zalando.nl	-0.34***	-0.15***	-0.91***	-0.70***	-0.59***	-0.11***
	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)
Number of previous	0.11***	0.07***	0.14***	0.12***	0.07***	0.07***
website visits	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Visit to a	1.81***	1.20***	2.12***	1.76***	1.37***	1.07***
price discount page	(0.07)	(0.06)	(0.06)	(0.08)	(0.06)	(0.06)
Outside option	1.42***	0.23***	2.23***	2.27***	1.14***	-0.23***
	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)	(0.02)
Search cost (exp)						
Constant	-2.13***	-1.09***	-3.83***	-4.00***	-2.05***	-0.59***
	(0.05)	(0.04)	(0.05)	(0.07)	(0.05)	(0.03)
Observations	17988	17988	17988	17988	17988	17988
LL	-20223	-12380	-11318	-7569	-9090	-10030
MAE (purchase)	47.22	88.74	17.57	17.87	51.46	98.25
MAE (search)	58.01	49.34	34.64	34.97	30.47	46.61
RMSE (purchase)	74.61	127.03	29.73	28.25	74.62	136.56
RMSE (search)	76.48	83.29	47.15	56.29	43.78	87.85

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers.