Search Gaps and Consumer Fatigue*

Current version: December 2021

Raluca M. Ursu† Qianyun Zhang‡ Elisabeth Honka§

Abstract

In the canonical sequential search model, consumers inspect options consecutively until they decide to stop searching, a decision which occurs only once before consumers determine whether and what to purchase. However, using data on consumers' online browsing histories, we document that consumers frequently take breaks during their search ("search gaps"), that is, they obtain information on a number of options, pause, and later resume their search. Further, we provide model-free evidence that consumers take breaks from searching due to fatigue. To describe search processes that include gaps due to fatigue, we extend the Weitzman (1979) framework and develop a sequential search model that rationalizes search gaps by allowing consumers to additionally decide when to search an option: now or after a break. Fatigue enters the model through increasing search costs: the more a consumer searches, the higher her search costs per option; taking a break reduces these costs to a baseline and enables the consumer to resume her search at a later time. We estimate the proposed model using our data and quantify the effect of fatigue on consumer search and purchase decisions. We find the effect of fatigue to be larger than that of baseline search costs. Lastly, we illustrate the impact of search gaps and consumer fatigue on market outcomes via counterfactuals.

Keywords: Sequential Search, Search Fatigue, Search Delay, Online Browsing, Apparel Industry

JEL Classification: D83, L81, M31

*We are thankful for comments from Tomomichi Amano, Simon Anderson, On Amir, Heski Bar-Issac, Alix Barasch, Bryan Bollinger, David Brown, Daria Dzyabura, Eric Greenleaf, Robert Kent, Sarah Komisarow, Matthew McGranaghan, Kanishka Misra, Dee Muir, Eitan Muller, Zach Nolan, Brennan Platt, Anita Rao, Andrew Rhodes, Robbie Sanders, Daniel Savelle, Sandro Shelegia, Jackie Silverman, Andrey Simonov, James Smith, Yi-Lin Tsai, Marton Varga, and Ken Wilbur; seminar participants at the Rotman School of Management (Toronto), the Rady School of Management (UCSD), and the University of Delaware; attendees of the 2019 Marketing Science/INFORMS conference, the 2020 Consumer Search Digital Seminar Series, and the 2021 Virtual Quant Marketing Seminar (VQMS). We have greatly benefited from the careful and meticulous help offered by Ishita Verma, Jian Zhang, and Xinyu Wei in augmenting and processing our data. We would also like to thank Siham El Kihal and GfK for helping us obtain the data. We appreciate the funding received from the Center for Global Economy and Business, NYU Stern. The usual disclaimer applies.

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

‡Stern School of Business, New York University, qianyunzhang@stern.nyu.edu.

§Anderson School of Management, University of California Los Angeles, elisabeth.honka@anderson.ucla.edu.

1 Introduction

Consumers can typically purchase products such as TVs or shoes from tens, if not hundreds, of different online retailers. Since evaluating each retailer and its merchandise is costly, consumers usually only investigate a small number of websites. Models of consumer search describe how individuals decide which options to become informed about and which ones to ignore (Stigler 1961, Weitzman 1979). These models characterize consumer search behavior as an *uninterrupted* process, with one stopping decision occurring after the consumer has evaluated all options that are optimal to be inspected.

However, using data on consumers' entire online browsing histories, we document that search processes frequently involve breaks ("search gaps"). That is, consumers often obtain information on a number of options during a session, then take a break from searching, and later resume their search in a different session (e.g., a day later). Our data come from GfK and capture all web traffic (8 million clicks) of a panel of 4,600 Dutch consumers during ten weeks in 2018. The data include all clicks in our focal category – apparel – as well as all other browsing activities that consumers performed during the same session (e.g., checking emails, visiting Facebook, or using search engines). Importantly, consumers' online activities can be linked over time, i.e., across sessions, revealing when consumers search products versus when they take a break from searching. With these data, we show that search gaps are prevalent: on average, 43% of consumers take at least one break while searching. Further, conditional on pausing the search at least once, the average search process contains three gaps.

Such search gaps are ignored by previous consumer search literature for several reasons. First, prior empirical work often does not observe search gaps because it employs data containing information on purchases only, survey data with information on purchases and only searched sets, or browsing data at the session level that cannot be linked across sessions or to an individual consumer (e.g., Seiler 2013; Honka 2014; Honka and Chintagunta 2017; Ursu 2018). And second, even when data on search gaps are available, prior literature assumes that sessions are either independent or that they can be grouped together as part of a (gap-free) larger search (e.g., Chen and Yao 2017, De los Santos and Koulayev 2017; Ursu, Wang, and Chintagunta 2020).[1] We provide empirical evidence that search sessions are not independent in our empirical context. Rather, later sessions are a continuation of a search started in earlier sessions.

[1] Another way of interpreting the decision to ignore search gaps is to say that prior work has assumed that only the last stopping decision is relevant.

Assuming away search gaps is innocuous if the reason for such gaps is unrelated to consumer search decisions, e.g., work emails or planned offline activities. However, there are several reasons to expect that search gaps are – to a large extent – conscious consumer decisions. For example, search gaps may occur when consumers expect prices or product features to change over time and thus think that they may benefit from delaying their search. A second potential reason is shopping fatigue: the more options a consumer searches, the higher her search costs per option due to fatigue; taking a break reduces these costs and enables the consumer to resume her search at a later time.[2]

Using model-free evidence, we show that search gaps are largely conscious consumer decisions and that fatigue is a main driver of search gaps in our data. To demonstrate the first point, we investigate what consumers do during search gaps. We find that approximately 90% of consumers engage in other online activities after ceasing their apparel search during a session. Such activities are often leisure activities (e.g., visiting social networking websites), suggesting that consumers had more uncommitted time during which they could have continued searching apparel products, but that they chose not to do so. Also, we find that consumers often resume their apparel search within a session when interrupted by an email notification, further supporting the idea that search gaps are a consumer's choice.

Next, we provide three pieces of evidence in support of the notion that search gaps are related to fatigue. First, we proxy for fatigue using consumer demographics and website characteristics. We show that consumers who are older and who visit websites that are slower to load or are harder to read have generally more search gaps. Second, we show that the more websites a consumer has searched and the more time she has spent searching since the last break, the higher her likelihood of a search gap. Finally, we present empirical evidence against several alternative explanations for the occurrence of search gaps: (i) expecting future changes in prices or other product features, (ii) having a limited budget of time, (iii) forgetting previously obtained information, and (iv) indecision.

We then develop a model of sequential search that endogenizes search gaps. More specifically, we extend the Weitzman (1979) sequential search model in two directions. First, we allow consumers to not only decide which products to search and in what order, but also *when* to search them: now or after a break. Second, motivated by our empirical evidence, we allow the decision of when to search a product to be influenced by fatigue. To this end, we model search costs as having two components: a baseline level and a component that depends on the number of options searched after the latest

[2]The following quote illustrates shopping fatigue: "Car shopping is exhausting and confusing. With every search online, I have to drink a sip of wine." (http://business.time.com/consumer-fatigue-shopping-has-never-been-easier-or-as-mentally-exhausting/).

break. This means that the cost of gathering product information after taking a break is equal to the baseline search cost level, while subsequent searches involve paying a higher search cost per option due to the fatigue that accumulated from previous searches within the same session. Note that these two components affect consumer decisions differently. For example, higher fatigue levels increase the number of gaps, while higher baseline search costs (indirectly) decrease the number of gaps by reducing the number of searched options.

The model we develop captures search gaps, but – in contrast to the Weitzman (1979) problem – no longer has an index policy solution. This is the case because, in our model, the optimization problems of different options interact for two reasons: (i) due to the increasing nature of search costs, i.e., searching an alternative increases search costs for *all* so far unsearched alternatives, and (ii) due to the choice of when to search an option, i.e., choosing to take a break resets search costs for *all* unsearched options. This interaction of optimization problems of different options violates the assumption in Weitzman (1979) that searching an option does not affect the payoffs from any other option, leading to a failure of the index policy solution. However, we show that under a set of fairly general conditions that are met in our empirical context (e.g., high fatigue given the observed prevalence of search gaps), as well as in simulations with more than 1,000,000 parameter combinations, the optimal search order in our problem coincides with the one in Weitzman (1979). Using this result, we then describe a consumer's optimal search rules for the entire set of decisions she makes in a model with search gaps: (i) which alternatives to search, (ii) when to search an alternative, and (iii) whether to continue to search or to stop. These optimal search rules are characterized by a set of four reservation utilities rather than one reservation utility as in Weitzman (1979).

We estimate our model and quantify consumer preferences, baseline search costs, and search fatigue parameters in the two largest apparel subcategories, "shirts, tops, & blouses" and "shoes." Our empirical results are consistent across both apparel subcategories. We treat all browsing on a retailer's website, such as hm.com or nike.com, as one search and recover consumers' utilities for the ten most popular websites in each subcategory.[3] We find that consumers are loyal to websites they have frequently visited before. More importantly, we show that fatigue has a large effect on search decisions, equivalent to increasing baseline search costs at least tenfold with every searched option. In

[3]Note that, in contrast to most literature studying online consumer search behavior, we examine consumer search of online retailers (websites) and not individual products in our empirical application. We do so because we expect consumers to be more likely to choose which websites to search, rather than which products to search, since the latter are often unknown to consumers and are discovered only after they navigate through a number of subpages on a website.

contrast, the baseline search cost estimate is relatively small. However, estimating the model using the canonical Weitzman (1979) framework leads to an overestimate of the baseline search cost. The Weitzman (1979) model ignores search gaps and assumes search costs are independent of the number of previously searched options. Thus, when estimated on the same data as a model in which fatigue affects search costs, the Weitzman (1979) model rationalizes the same number of searched options by inflating baseline search costs. We also show that an adapted version of the Weitzman (1979) model that ignores search gaps, but models fatigue (i.e., search costs are an increasing function of the number of previously searched options) also overestimates baseline search costs, albeit less, and underestimates the effect of fatigue. This occurs because ignoring the fact that fatigue causes consumers to take a break from searching (not only increases search costs) leads to the mistaken impression that fatigue has a relatively smaller impact on consumer decisions.

Finally, we measure the impact of consumer fatigue and search gaps on market outcomes via counterfactuals. First, we quantify the effects of fatigue on searches and purchases, and compare them to those of baseline search costs. Decreasing fatigue by 50% increases the number of searched websites by $1-4\%$, increases transactions by $0.5-1.2\%$, and lowers search gaps by $11-22\%$. The effects of a fatigue reduction are larger than the effects of a baseline search cost reduction of the same magnitude. Furthermore, while all websites suffer from high fatigue levels, larger and more popular website are less negatively affected than smaller and less popular websites. And second, we investigate the consequences of consumers not being able to reduce their fatigue levels via search gaps. Such a situation might occur when consumers face challenging times such as the Covid-19 pandemic or when consumers are constantly being stimulated by (tiring) marketing activities. We find that not being able to reset fatigue during a break leads to a significant reduction in the number of products consumers search and purchase: searches decrease by approximately 20% and purchases by more than 6%. Most importantly, larger and more popular websites are hurt less in such a situation. In other words, consumers become more likely to buy from larger and more familiar websites. These findings emphasize the importance of search gaps for competition and brand value.

The contribution of this paper is two-fold. First, we document the presence of search gaps during consumers' search processes and propose the first model of consumer search that accounts for such search gaps. To the best of our knowledge, we are also the first to model consumer search fatigue

before a purchase.[4] And second, search gaps reveal that consumers might be stopping their search due to a high fatigue level rather than a low match value with a brand. Identifying and targeting such consumers might be profitable for companies since these consumers are still active in the market and thus more likely to resume searching and ultimately purchase (Schmittlein, Morrison, and Colombo 1987). The observation that consumers may stop searching because of a high fatigue level also has implications for firms' pricing decisions. Prior work on ordered search has found that firms' optimal pricing decisions depend on the order in which they are searched (e.g., Arbatskaya 2007; Armstrong, Vickers, and Zhou 2009; Petrikaite 2018). However, these results no longer necessarily hold when consumers can take search breaks.[5]

The rest of the paper is organized as follows. In the next section, we discuss relevant prior work. In Section 3, we introduce our data and in the following section, we provide model-free evidence for search gaps being conscious decisions and related to fatigue. We develop our theoretical model in Section 5. In Section 6, we describe our empirical model, estimation procedure, and identification. In the following section, we present our results, while, in Section 8, we provide a description of two counterfactual exercises. We conclude in the last section.

2 Relevant Literature

This paper is primarily related to three strands of the literature: (i) the theoretical consumer search literature, (ii) empirical work using individual-level search data to quantify consumer preferences and search costs, and (iii) prior work on choice deferral. We describe and delineate our paper vis-à-vis extant research.

We contribute to theoretical work on consumer search in two main ways. First, we develop a new model of consumer search, adding to a rich literature that generally follows one of two frameworks: either the sequential search model of Weitzman (1979) or the simultaneous search model by Stigler (1961). In both these frameworks, consumers inspect products consecutively until they decide to stop searching, a decision which occurs once before determining whether to purchase. For example,

[4] Carlin and Ederer (2019) develop a model of search fatigue in which fatigue affects consumers across purchases rather than before a purchase, which is our focus.

[5] For example, Armstrong, Vickers, and Zhou (2009) show that the non-prominent firm can infer that the consumer searching it obtained a low match value at the prominent firm. In this case, the non-prominent firm will face a relatively more inelastic demand for its product, allowing it to charge a higher price than the prominent firm in equilibrium. This inference is weakened when the consumer has the option to visit the non-prominent firm after a search gap, since such a decision may be motivated by the low search cost after the break, rather than a low match with the prominent firm. Thus, observing when the consumer searches an option (before or after a search gap), not only whether she searches, may help companies' pricing strategies.

in Weitzman (1979)'s sequential search model, the consumer searches options as long as the benefit from search exceeds the cost. When this relation no longer holds, search ceases and the consumer determines whether to purchase. Similarly, in Stigler (1961)'s simultaneous search model, there is one stopping decision: after searching the set of options for which the expected benefit exceeds the search cost, the consumer stops and decides whether to buy one of the searched products. Thus – in contrast to our model – neither framework can be used to study search gaps, which involve multiple stopping decisions.

The only exception is a model for homogenous goods developed by Morgan and Manning (1985). The authors demonstrate that, under very general conditions, neither simultaneous nor sequential search is optimal, but rather a combination of the two is, i.e., a process during which the consumer searches sets of options sequentially. Morgan and Manning (1985)'s model can give rise to search gaps since consumers may choose sets of options to search at every occasion and take breaks between sets. However, since their theory was developed for homogenous goods, i.e., all products are ex ante identical, it can explain how consumers choose the number of options to search in every set, but not the identity of those options. Therefore, to the best of our knowledge, no prior theoretical work exists that can account for search gaps when consumers search among heterogenous goods, i.e., also choose which products to search. Our model fills this gap in the literature.

Second, we contribute to the theoretical consumer search literature through our definition of search costs. Most prior work assumes that search cost per product are independent of the number of products searched.[6] To the best of our knowledge, there are only a few exceptions. Stiglitz (1987) studies the effect of convex search costs on competition and the equilibrium number of firms in the market. The author links this effect to the increasing scarcity of time and money that intensifies as the consumer continues searching, but does not consider the effect of resetting these costs on search decisions. Levav et al. (2010) show experimentally that participants who need to customize a product (a suit or a car) are more likely to choose the default option when first presented with options that have many rather than few attributes. The authors argue that this result can be partially explained by convex costs of evaluating attributes, as demonstrated by literature in psychology and economics modeling self-control as a muscle that requires more effort on future rather than identical early stimulation (Ozdenoren, Salant, and Silverman 2012; Vohs et al. 2008).

[6] For a review of theoretical work on consumer search, see Baye, Morgan, and Scholten (2006) and Anderson and Renault (2018).

Carlin and Ederer (2019) develop a model of search fatigue in which fatigue affects search decisions across purchase trips, i.e., the more products the consumer searched before the previous purchase, the higher her search costs are when searching towards the next purchase decision. In contrast, in our paper, we focus on the effect of fatigue on search decisions before a given purchase, i.e., the more the consumer searches before the current purchase, the higher her search costs.[7] Also, in Carlin and Ederer (2019) the goal is to study the effect of search fatigue on firm pricing decisions in equilibrium, while we take our model to data and quantify the effect of fatigue on consumer decisions. Most closely related to our paper, Ursu and Dzyabura (2020) posit that search costs increase linearly in the number of alternatives searched and affect current search and purchase decisions, modeling choices which we also make. However, the presence of increasing search costs is not sufficient for the occurrence of search gaps. More precisely, such search costs may explain why the consumer stops searching, but not why she restarts. For consumers to be willing to resume their search, search costs must also decrease during a gap (if the consumer's utility from the available options remains unchanged). To the best of our knowledge, no prior work on consumer search suggests this possibility. Instead, prior economics work on education finds that taking a break from academic classes to perform physical exercises, helps students to recover from cognitive fatigue and to perform better academically (Bednar and Rouse 2019). We posit that a similar mechanism may drive search fatigue.

Our paper is also related to empirical work quantifying preference and search cost parameters using individual-level data on consumers' search activities (e.g., De los Santos, Hortaçsu, and Wildenbeest 2012; Koulayev 2014; Chen and Yao 2017; Honka and Chintagunta 2017; Honka, Hortaçsu, and Vitorino 2017; De los Santos and Koulayev 2017; Dong et al. 2020; Yavorsky, Honka, and Chen 2021).[8] Most of this work assumes that search costs per product are independent of the number of products searched. The exception is Koulayev (2014) who estimates higher search costs for products searched later, providing empirical support for the assumption of increasing search costs. Furthermore, this stream of the literature rests on the theoretical models of Weitzman (1979) and Stigler (1961) and assumes that consumers search options consecutively and stop searching only once. Although some prior work recognizes the fact that consumers search in sessions (e.g., consumers learn across sessions in Wu et al. 2015), it does not explicitly model consumers' decisions to stop and resume searching

[7] A concept that may seem similar to fatigue is that of obfuscation (e.g., Ellison and Ellison 2009, Ellison and Wolitzky 2012). However, the difference is that by obfuscating the consumer, firms increase their (baseline) search costs and the cost of making successive searches. As such, our model and predicted behavior differ from those observed in a model with search obfuscation.

[8] For a review of empirical work on consumer search, see Honka, Hortascu, and Wildenbeest (2019).

several times, and is thus not accounting for the presence of search gaps.

Finally, our paper relates to the literature on choice deferral. Work in consumer behavior shows that choice difficulty increases the probability of the consumer choosing none of the options and thus delaying her choice (Dhar 1997; Novemsky et al. 2007). In the context of a search model, we view this finding as broadly suggesting that search gaps are more likely as search difficulty increases, a result which is in line with our empirical patterns. More closely related is the work of Greenleaf and Lehmann (1995) that identifies several possible reasons for consumers delaying the decision to purchase a product such as the absence of time to devote to the task or the expectation of future price decreases. Although not described in the context of consumer search, these reasons could also influence search decisions. We contribute to this literature by developing a model of consumer search in which consumers may stop and restart searching, thereby formalizing the idea of delay in the context of search.

3 Data

3.1 Data Sources

Our primary data come from GfK, Germany's largest market research company. GfK recruits and maintains an online panel of representative consumers for whom online browsing data are collected via a browser extension. This browser extension is installed on the panelists' devices (PC, smartphone, tablet) and records all their online activities. GfK groups all clicks which are not interrupted by a time period of inactivity longer than 30 minutes (the industry standard) into "sessions." The data are at the exact URL level clicked by a consumer and also contain the time of each click, the visited website, and consumer demographics (e.g., age, gender). Furthermore, GfK classifies clicks into activities such as email, social networking, apparel, search engine use, banking, or gaming. And finally, GfK codes the transaction funnel identifying website visits, product views, basket additions, checkouts, and order confirmations.

Our data contain the complete PC browsing histories of online panel members from the Netherlands from February 15, 2018, until May 1, 2018 (ten weeks), for sessions during which they made at least one click to an apparel website. In other words, our data are conditional on an apparel click (not conditional on a purchase) occurring during a session, but show all visited websites (including non-apparel

websites) during such sessions. We chose to focus on products in the apparel category for two reasons: first, this category is frequently visited by consumers, allowing us to observe multiple search actions. And second, we were able to scrape product information for the URLs in the GfK data because they are stable over time and are generally not personalized to individual consumers.[9] While choosing a category such as travel would allow us to observe enough search activity, we would not be able to scrape product information since this information changes frequently and is often personalized. On the other hand, a durable goods category only contains searches from a small number of consumers and would restrict our analysis given our relatively short observation window.

We augmented the GfK data in several ways. First, we scraped product information from 44 of the top 50 apparel websites. These 44 websites account for more than 57% of all apparel clicks, a large percentage given the 1,046 unique apparel websites in our data. This data collection stage occurred within one month of the last day of our observation period to prevent changes on the webpages. The product information we obtained includes price (current and any promotions), page title, brand name, product name, product color, reviews, star rating, number of photos, product description, shipping information, speed score of the website, word counts, and page readability.[10] Second, we identified the purchased product as the last product searched before engaging in transaction related clicks (e.g., adding to cart, checking out, confirming an order) on the same website. Using this information, we defined a "spell" as all search sessions conducted by a consumer before a purchase (or before the end of our observation period if no purchase occurred). Next, we use URLs, page titles, and the scraped information (e.g., product description) to identify nine product subcategories (e.g., "shoes" or "accessories") that the consumer searched. Finally, we define a "search gap" as the break a consumer takes between subsequent search sessions. Figure 1 provides an example of a search process, defining the concepts we use in this paper. Detailed information about the data collection, classification, and cleaning steps are provided in Web Appendix A.

==========

Insert Figure 1 about here

==========

[9] Cavallo (2017) finds that 92% of apparel prices are the same online and offline within a chain, suggesting little personalization to individual consumers' visits.

[10] We obtain website speed score information from Google (https://developers.google.com). The website speed score is the page loading speed with values ranging from 0 to 100. Google PageSpeed Insights considers $0-49$ as slow, $50-89$ as medium, and greater than 90 as fast speed. We obtained other website features such as word counts, number of images, and readability from `https://urlprofiler.com/`.

3.2 Data Description

Our data contain 7,877,551 observations with 428,651 apparel clicks. We observe searches made by 4,622 consumers in 5,665 spells and 40,735 sessions across nine distinct apparel subcategories. There are a total of 3,036 products purchased in the apparel category, with 76% of spells containing no purchased product, 11% of spells containing one purchased product, and 13% of spells containing at least two purchased products. 65% of consumers are female; the average (median) age is 48 (49) with a large standard deviation of 16. Click duration is, on average, half a minute and is slightly longer for apparel than non-apparel clicks (0.54 versus 0.50 minutes, respectively).

We summarize session characteristics in Table 1. Activity in each session is extensive: on average, consumers make 190 clicks on 30 websites and spend more than one hour online. In contrast, apparel search in a session is more modest: the average consumer makes 11 apparel clicks, spends about five minutes searching, and visits one apparel subcategory. The most popular activities in our data are email, social networking, and apparel. Together they account for more than 33% of all clicks. The most popular websites are google.com, live.com, and facebook.com.

==========

Insert Table 1 about here

==========

Table 2 lists the top apparel websites in terms of their searches and transactions. Zalando is the most popular apparel website in our data and among online retailers in the Netherlands.[11] More precisely, Zalando has more than 22% of transactions in our data (15% of apparel clicks), followed by H&M with 13% of transactions (10% of apparel clicks). The two most commonly purchased apparel subcategories are "shirts, tops, & blouses" and "shoes." In Table 3, we additionally display the most popular websites searched and purchased in each of these two subcategories. Once again we note the overall popularity of Zalando, as well as C&A and H&M in the "shirts, tops, & blouses" subcategory, and of Schuurman Shoenen and Van Haren in the "shoes" subcategory. Finally, the subcategory "jackets & vests" is the most expensive one with an average transaction price of 60€, while "children's clothes," the cheapest subcategory, has an average transaction price of less than 20€.

==========

Insert Tables 2 and 3 about here

==========

In Table 4, we demonstrate that search gaps are very prevalent in our data. More specifically, across all nine apparel subcategories, on average, 43% of search spells contain at least one search gap.

[11] For details, see https://ecommercenews.eu/top-10-online-stores-in-the-netherlands/.

Conditional on a spell having at least one search gap, the average number of search gaps is three per spell. The median length of a search gap (number of days between search sessions) is less than four days, with 25% of search gaps lasting less than one day.[12] Therefore, we focus on studying why search gaps occur and how they can be understood from the lens of a search model.[13] Spells range from 7 to 17 days, on average, depending on the subcategory.[14] In comparison, the average time between spells (for the approximately 30% of consumers who have more than one spell during our observation period) is longer, typically lasting about two weeks.

==========
Insert Table 4 about here
==========

To summarize, we find that search gaps occur frequently both across consumers and across apparel subcategories. In the next section, we delve deeper and attempt to relate search gaps to fatigue.

4 Search Sessions, Search Gaps, Fatigue, and Alternative Explanations

In this section, we first provide evidence that search sessions are not independent, i.e., that later search sessions are a continuation of a search started in earlier sessions, and that search gaps are conscious consumer decisions, i.e., that consumers *choose* when to take a break. We then show empirically that these decisions are affected by fatigue. And finally, we present evidence against possible alternative explanations for the occurrence of search gaps.

4.1 Empirical Evidence Showing that Search Sessions are not Independent

Here, we show several pieces of evidence in support of later search sessions being a continuation of a search started in earlier sessions, i.e., of sessions not being independent.

First, across all apparel subcategories, 71% of spells do not contain any website revisits, 26% of spells contain revisits without a purchase, and 3% of spells contain revisits with a purchase (for statistics on each subcategory separately, see Table B-1 in Web Appendix B). The empirical pattern that 71% of spells do not contain any revisits supports the idea that sessions are not independent, as consumers continue their search by visiting new websites and obtaining new information in the same

[12] The average length of a search gaps is approximately one week, skewed by a small number of longer gaps (less than 10%).

[13] We also observe when a consumer takes a break *within* a search session. 31% of search sessions contain such gaps which, on average (median), only last 3.73 (0.27) minutes. Because such gaps are less prevalent and are unlikely to lead to a change in fatigue due to their short duration, we analyze search gaps across sessions.

[14] The average spell length may be lower than the average search gap length because only a fraction of spells contain gaps.

apparel subcategory. Further, even when consumers revisit the same website (e.g., Nike.com), they typically look at new products in the same subcategory (Nike Air sneakers rather than Nike Pegasus sneakers) they did not see on the previous visit. In fact, only 5.3% of spells contain revisits to the same product page. And lastly, conditional on ending in a purchase, 68% of spells do not contain a revisit and 32% of spells contain a revisit. Taken together, these data patterns suggest that search across sessions is connected.

Second, in Figure B-1 in Web Appendix B, we plot prices of searched products within a spell relative to the price of the purchased product (for converting spells) for all nine apparel subcategories.[15] We find that, across most apparel subcategories, products searched closer to the end of the spell contain searched prices that are more similar to the price of the purchased product than products searched earlier in the spell. Although consumers likely do not search solely to learn prices in the apparel category, this convergence in searched prices further supports the idea that searches across sessions are not independent. Finally, we study the degree to which products searched in the same subcategory are similar. Within each apparel subcategory, there are natural subdivisions. For example, we divide searches within the "shirts, tops, & blouses" subcategory into searches for "short sleeve shirts" and "long sleeve shirts", and searches in the "shoes" subcategory into separate searches for "sneakers," "sandals," "boots," and "heels."[16] Looking at these subdivisions, we find that, on average, 36% of spells contain at least one search gap, with as many as 48% of search spells for "sneakers" containing at least one search gap. In other words, we find that consumers frequently continue their search for very similar products after a break, suggesting that search sessions are not independent.

4.2 Empirical Evidence Showing That Search Gaps are Conscious Decisions

There are several pieces of evidence in our data pointing to search gaps being conscious decisions, i.e., occurring as a result of a decision by the consumer to delay her search. First, we find that consumers' online activity rarely ends when their search in the apparel category ceases. More specifically, only 13% of sessions end with an apparel click. Furthermore, the two most popular activities after the last apparel click are email and social networking, accounting for more than 23% of all clicks. These two

[15] To show this, we split searches within a spell into deciles, and compute the percentage price difference in the products searched in decile n relative to the price of the purchased product. Deciles are computed following the method in Bronnenberg, Kim, and Mela (2016): $d(n, N_i) = ceil\left(\frac{10(n-r(0,1))}{N_i-1}\right)$, where $r(0,1)$ is a draw from a uniform distribution on the interval $(0,1)$, n denotes the search under consideration, and N_i denotes the total number of searches performed in spell i.

[16] These classifications account for at least two thirds of all searches within each subcategory.

activities remain the most popular ones even when restricting the data to clicks in the evenings (6pm to midnight) or on weekends, increasing the chances of them capturing leisure activities. These data patterns suggest that consumers had more time available to allocate to online activities, but that they chose not to spend more time searching for apparel.

Second, although we do not observe what consumers do during search gaps across sessions, we observe what they do when pausing their search within a session. Here again, we find that email is the most popular activity with 16% of clicks, followed by social networking with 6% of clicks.

And lastly, we observe when a notification announcing that an email was received interrupts the apparel search and how consumers react to this event. Receiving such a notification is arguably exogenous to the consumer search for apparel products. We find that 91% of consumers who get a notification while searching in the apparel category return to searching apparel in the same session. In other words, consumers do not pause their search in response to an email notification, i.e., take a search gap. Taken together, these data patterns suggest that search gaps are a result of conscious decisions made by consumers.

4.3 Empirical Evidence for Fatigue Affecting Search Gaps

Here, we aim to link fatigue and search gaps. Doing so is challenging because we do not directly observe a consumer's fatigue level and thus cannot directly relate it to the decision to take a break versus to continue searching without a break. An ideal experiment would manipulate consumers' fatigue levels (e.g., by making some websites slower to load or by increasing the amount of information available) and directly test whether fatigue affects search gaps. In what follows, we seek to mimic this experiment and test the relation between fatigue and search gaps using observational data.

First, we consider two proxies for fatigue: consumer demographics and website characteristics. We then check whether these fatigue proxies are related to the number of gaps a consumer makes while searching. We define our dependent variable as the logarithm of the number of search gaps in a spell (plus one) and present the results in columns (i) and (ii) in Table 5. Column (i) shows the results for "shirts, tops, and blouses" and column (ii) shows the results for "shoes," the two most commonly purchased apparel subcategories in our data. All regressions control for the number of searches consumers perform since consumers who search longer generally also have more search gaps.

==========
Insert Table 5 about here
==========

We find that older consumers have more search gaps in a spell. There is an abundance of medical research supporting the idea that mental processing abilities are affected by age, with observed declines in conceptual reasoning, memory, processing speed, and attention to stimuli in older individuals (Harada, Love, and Triebel 2013). These changes in mental abilities can affect decision-making processes in marketing-relevant contexts (Peters 2010; Carpenter and Yoon 2011). For example, older consumers have been shown to make better decisions when presented with fewer options (Tanius et al. 2009; Abaluck and Gruber 2013). Also, research shows that older consumers are more likely to use heuristics, to search for a shorter amount of time, and to build smaller consideration sets to reduce cognitive effort (Kim et al. 2005; Lambert-Pandraud, Laurent, and Lapersonne 2005). Motivated by this evidence, we consider age as a possible proxy of a consumer's proneness to fatigue and find age to have a positive effect on search gaps.[17] Furthermore, consumers who visit websites that are slower to load (lower speed score) and harder to read (higher readability/SMOG index) also have more search gaps.[18] Assuming that these variables are suitable proxies for consumer fatigue, our results show that higher fatigue levels lead to more search gaps.

Second, we check whether the number of options searched after the last gap affects the probability of a gap. If more search increases fatigue levels, then the more websites the consumer searches after the latest break, the higher the probability of a search gap. We consider two measures of the number of searches performed: (i) the cumulative number of websites searched since the previous gap and (ii) the total number of minutes spent searching since the previous gap. The results are displayed in columns (iii) to (vi) in Table 5. Consistent with our hypothesized relation between fatigue and search gaps, we find that the more websites a consumer searched after the latest break or the more time she spent searching, the more likely it is that she takes a break.

[17] For consistency with the empirical setup in Section 7, we also estimated the regressions in Table 5 with an indicator for age greater than or equal to 50. The results are very similar to those shown in Table 5: the age coefficient in subcategory 1 (2) is significant and equal to 0.1349 (0.3169) with a standard error of 0.0284 (0.0333). The results are available from the authors upon request.

[18] Page readability is measured in terms of its SMOG index, which computes the number of years of education needed to understand a piece of text. Therefore, a larger SMOG index means a less readable text. More information about the SMOG index can be found at `https://en.wikipedia.org/wiki/SMOG`.

4.4 Empirical Evidence Against Alternative Mechanisms

In what follows, we briefly discuss several potential alternative explanations for the occurrence of search gaps. Our goal is not to rule out *all* alternative explanations. Indeed, multiple reasons may affect the occurrence of search gaps. Rather, our goal is to provide empirical evidence against some of the most common alternatives. Modeling the relation between search gaps and factors other than fatigue is left for future research.

One alternative explanation for search gaps is that consumers may delay their search because they expect prices to decrease or other product features to improve. This reason is particularly pertinent in a category such as travel, in which airfare and hotel prices change very frequently and dynamically in response to changes in demand and the available supply of options. However, in the apparel category, price and product feature changes are less frequent. More precisely, product features change mostly every season and there are only two major seasons annually (Fall/Winter, running from July to December, and Spring/Summer, running from January to June),[19] while prices typically change around holidays or at the end of each season when products go on sale.[20] Our observation period (February 15 to May 1) falls within a single season and does not overlap with any major sales periods.[21] Further, the median search gap in our data is shorter than four days. Therefore, it is unlikely that search gaps occur because consumers expect prices or other product features to change in our empirical setting.

A second alternative explanation is that consumers resume their search to obtain additional information about the same products or because they forgot the information they gathered previously. Note that these behaviors occur even absent gaps, i.e., they may occur even if the consumer does not take any breaks during her search. Both these decisions involve revisits of previously searched options (see, e.g., Ursu, Wang, and Chintagunta 2020; Dang, Ursu, and Chintagunta 2020). However, in our data, most spells do not contain any website revisits (71%) and product page revisits account for only 5.3% of all clicks, with most of these product page revisits (72%) occurring within a session rather than across sessions. Therefore, revisits cannot explain search gap behavior.

Third, search gaps may occur because consumers only have a predetermined budget of time

[19] For more details, see https://www.leaf.tv/articles/when-do-fashion-seasons-start/.

[20] Spring/summer goods usually go on sale in June and July and Fall/Winter goods usually go on sale in January after the winter holidays (see https://money.usnews.com/shopping-holidays-the-best-days-to-shop-this-year, https://www.thebalance.com/comprehensive-guide-to-seasonal-sales, or for the Netherlands https://www.amsterdamsights.com/shopping/sales-period).

[21] The exception is Easter, which occurred on April 1, 2018. However, comparing transactions one week before and one week after Easter shows no significant difference in prices (difference= 0.02, t = 1.32).

available to allocate to searching during the current session.[22] We present three pieces of evidence against this alternative explanation. First, as previously noted, only 13% of sessions end with an apparel click and apparel searches are typically followed by leisure activities. This observation suggests that most consumers had more uncommitted time available, but chose not to devote it to additional apparel searches. Second, we plot the length of a search session in minutes in Figure 2(a). If consumers had a predetermined budget of time for shopping for apparel, we would expect to see spikes at, e.g., 30 or 60 minutes. However, we do not observe such a pattern in Figure 2(a). Third, in Figure 2(b), we plot the second of the last apparel click in a search session (0 seconds represents any full hour during a day, 1,800 seconds represents any half hour during a day, etc.). If consumers had a predetermined budget of time for shopping for apparel until, e.g., the beginning of a meeting, we would expect to see spikes at, e.g., 0 or 1,800 seconds since meetings tend to start at those times. However, we do not observe such a data pattern in Figure 2(b). To summarize, we find no evidence that consumers had a predetermined budget of time to search for apparel products.

==========

Insert Figure 2 about here

==========

And finally, we also consider the possibility that search gaps occur when consumers are undecided about which product(s) to buy or whether to make a purchase at all. Although we do not observe consumers' levels of indecision, website and especially product page revisits represent behavior that is consistent with indecision: consumers may revisit websites and product pages to gather more information about a product and to resolve their indecision. However, as also discussed in Section 4.1, 71% of the spells do not contain any website revisits. Even when consumers revisit the same website, they typically look at new products in the same subcategory they did not see on the previous visit. In fact, only 5.3% of spells contain revisits to the same product page. Such data patterns are inconsistent with indecision. Further, we hypothesize that searching for (i) products with more features and (ii) more expensive products makes purchasing a more challenging decision, and thus may more often lead to indecision. If this hypothesis is correct, we should observe more search gaps in subcategories in which products have more features and are more expensive relative to subcategories in which products have fewer features and are cheaper.[23] Using the statistics from Table 4, we find little difference in the

[22] Note that this reason may also involve changing search costs over time, but it differs from fatigue because of the predetermined nature of a budget of time.

[23] We view "underwear," "children's clothing," and "accessories" as subcategories with simpler products (fewer features) and "shirts, tops, and blouses," "shoes," "pants and jeans," and "dresses and skirts" as subcategories with more complex products (more features). Looking at prices, the three most expensive subcategories are "jackets and vests," "shoes," and "dresses and skirts;" the three least expensive subcategories are "children's clothes," "accessories," and "shirts, tops, and blouses."

percent of spells with at least one search gap or in the number of search gaps in these subcategories when comparing (i) products with fewer and more features and (ii) more expensive and less expensive products. And lastly, we consider another indicator of consumers' indecision: cart additions followed by search gaps. When consumers are undecided about whether to make a purchase, they may add a product to their cart, and then abandon it by taking a break from searching. However, we find that only 3% of search gaps occur immediately after a cart addition and that most cart additions happen in the last session of a spell (53%). Together, these data patterns do not support the notion that indecision is the main factor driving search gap decisions.

In sum, we showed that search gaps are mostly driven by fatigue and not by expectations, a limited budget of time, forgetting or indecision in our empirical context. Next, we develop a model of sequential search that endogenizes search gaps due to fatigue.

5 Model

5.1 Setup

A consumer seeks to purchase from option $j = 1, \ldots, J$ or to choose the outside option of not purchasing (denoted by $j = 0$).[24] The consumer knows the utility distributions $F_j(\cdot)$, but has to search to learn the actual utility of a product u_j, i.e., u_j is an independent draw from the continuous distribution function $F_j(\cdot)$. Search occurs sequentially. Let S denote the set of searched options, while $\bar{S}$ denotes the set of options still available for search. The best option among the searched ones is denoted by y, i.e., $y = \max_{j \in S \cup \{0\}} u_j$. Consumer choices depend on the state variables $\bar{S}$ and y.

We extend the framework developed by Weitzman (1979) to account for search gaps. This involves making two modifications. First, we allow the consumer not only to decide which option to search, but also when to search it: now or after taking a break. To this end, we define a new (additional) state variable, $t \geq 0$, which tracks the number of options searched *after* the last break, implying $|S| \geq t$. If the consumer decides to continue searching, she can either search an option j immediately at t or after taking a break, which resets t to zero.[25]

[24] We omit consumer i subscripts in what follows, but our model should be understood as applying to every single consumer.

[25] Many factors (e.g., advertising) may influence a consumer's decision of when to resume her search. However, modeling a consumer's choice of search gap length is beyond the scope of this paper and we assume that we observe the consumer searching again after t was reset to zero. Consistent with this modeling choice, a preliminary analysis shows that longer gaps do not correlate with the number of searches performed in the next session. This finding suggests that only a short amount of time is needed to reset search costs, making it not paramount to model search gap length. The analysis is available from the authors upon request.

And second, we allow search cost per option to increase with the number of searched options.[26] Following Ursu and Dzyabura (2020), search costs are given by

$$c_j = c_{j0} + \alpha t \tag{1}$$

where c_{j0} are baseline search costs and t captures the number of searches without a break. Recall that $t = 0$ after the consumer takes a break or before she searches any options. The first component c_{j0} captures the cost of searching j regardless of the number of other options searched. It depends on characteristics of that option, e.g., a website's prominence among results in a search engine ranking. The second component depends on t and represents the consumer's fatigue from searching, i.e., is due to the number of previously searched options. For simplicity, we assume that the difference in cost between subsequent searches is constant, i.e., it costs an additional $\alpha > 0$ for the consumer to search option j after having searched t other options. This functional form implies that the cost of the first search and that of searching an option after taking a break equal c_{j0}, while other searches involve paying a higher cost per option.

Given the state variables $(\bar{S}, t, y)$, at each search occasion, the problem solved by the consumer is given by

$$V(\bar{S}, t, y) = \max_{stop,\, continue} \left\{ y,\ \max_{j \in \bar{S}} \Phi_j(\bar{S}, t, y) \right\}, \tag{2}$$

where $V(\emptyset, t, y) = y$. The search value $\Phi_j(\bar{S}, t, y)$ is defined as

$$\Phi_j(\bar{S}, t, y) = \max_{now,\, later} \left\{ -c_{j0} - \alpha t + W_j(\bar{S}, t+1, y),\ \beta \left[-c_{j0} + W_j(\bar{S}, 1, y) \right] \right\}, \tag{3}$$

with $0 < \beta < 1$ being the discount factor if the consumer decides to search later. Furthermore, the continuation value $W_j(\bar{S}, t+1, y)$ is given by

$$W_j(\bar{S}, t+1, y) = V(\bar{S} \setminus j, t+1, y) F_j(y) + \int_y^\infty V(\bar{S} \setminus j, t+1, u)\, dF_j(u). \tag{4}$$

The interpretation of the value function in equation (2) is as follows: given a set of options available for search $\bar{S}$, a number t of options searched after the latest break, and a best option observed so far y, the consumer makes three decisions. First, she decides whether to stop or to continue searching. If she

[26] An alternative model would be to let search costs be a function of the elapsed time since the prior break. We leave such a model to future research. Instead and consistent with most of the literature that considers increasing search costs (e.g., Stiglitz 1987, Carlin and Ederer 2019, Ursu and Dzyabura 2020), we let search costs be a function of the number of searched options and note the similarity between these approaches (i.e., greater elapsed time is a direct consequence of additional searches).

stops searching, she gets a payoff y which represents the option of buying the alternative with the highest utility revealed among those searched or of choosing the outside option of not purchasing. Second, if she decides to continue searching, she has to make two decisions (simultaneously): (i) which option j to search among those not yet searched and (ii) whether to search the chosen option now (at t) or after a break. $\Phi_j(\bar{S}, t, y)$ denotes the value of choosing to search option j from $\bar{S}$. Searching j immediately involves paying a relatively high search cost due to t. In contrast, choosing to search j after a break involves paying a lower search cost, but discounting the continuation value at a rate β. If the consumer decides to search j after taking a break, she receives no utility in the current time period and t resets to zero. In this model, a higher level of fatigue α encourages more search gaps, making the value of searching an option now less desirable. The continuation value $W_j(\bar{S}, t+1, y)$ defined in equation (4) is given by the probability of revealing a utility u lower than y by searching j, and thus continuing the process with the same best option y, and by the probability of revealing a utility u higher than y, and thus continuing the process with u. We assume that $-c_j + W_j(\bar{S}, 1, y) \geq 0, \forall j \in J$. Search for options for which this inequality does not hold is postponed indefinitely. It follows that $\Phi_j(\bar{S}, 0, y) = -c_{j0} + W_j(\bar{S}, 1, y)$.

If t did not affect search decisions, i.e., if $\alpha = 0$ and t were not a state variable, then our problem would coincide with the Weitzman (1979) problem, in which $V(\bar{S}, y) = \left\{y, \max_{j \in \bar{S}} -c_{j0} + W_j(y)\right\}$. A consumer's optimal search strategy in the Weitzman (1979) problem can be described as an index policy: for each option j, compute an index, i.e., a reservation utility, and proceed to search options in a decreasing order of these indices until all options have been searched or any unsearched options have an index lower than the best observed utility among those searched. The index of option j is the unique solution z_j to $c_{j0} = W_j\left(z_j\right) - z_j$, where the continuation value $W_j(y)$ simplifies to $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$ since the one-step ahead policy is optimal (for more details, see Weitzman 1979). In this paper, we extend the Weitzman (1979) framework to account for search gaps and consumer fatigue. This also requires proposing a new solution to the problem we presented above since, as we describe in the next two subsections, it does not have an index policy solution.

5.2 Indexability and Related Problems

To the best of our knowledge, no index policy solution to the general problem presented in equation (2) exists. To understand why this is the case, consider the following examples of simpler versions of

our model: if gaps happen exogenously (i.e., the consumer does not choose whether to search now or later) and if search costs are constant, our problem coincides with the Weitzman (1979) problem and an optimal index policy exists. However, if gaps happen exogenously but search cost increase with the number of searches, the problem above does *not* coincide with the Weitzman (1979) problem. In particular, it is not indexable because the optimization problems of different options interact: although utility draws are independent across options, searching an option increases the search costs for *all* so far unsearched options. This means that the assumption in Weitzman (1979) that searching an option does not affect the payoffs of any other option is violated, leading to a failure of the index policy solution. Therefore, one reason why no index policy solution exists in the general version of our problem is the increasing nature of search costs.

In addition, in our problem, consumers decide not only which options to search, but also when to search them (now or later), adding an extra layer of decision-making not present in the Weitzman (1979) model (or most other search models in the literature). This feature makes our model resemble a bandit superprocess (BSP), a generalization of a multi-armed bandit problem (itself a generalization of the Weitzman 1979 problem). In BSP, a decision maker not only chooses which of a set of independent arms/processes to play in each time period, but also chooses from a number of actions for each arm/process conditional on playing (Gittins, Glazebrook, and Weber 2011).[27] Similarly, in our problem, a consumer chooses not only which products to search, but also when to search them (now or later). BSPs are generally not indexable because the optimizations for different processes interact: choosing a certain action for one process can affect the rewards of a different process (Whittle 1980, Brown and Smith 2013). In our case, choosing to take a break resets search costs for *all* unsearched options, meaning that the payoffs from one option are affected by choices related to other options. Thus, this additional layer of decision-making present in our model also prevents indexability.

5.3 Deriving a Solution

Despite the aforementioned indexability issues, in this subsection, we propose a solution for the entire set of decisions a consumer makes in our model: (i) which alternatives to search, (ii) when to search an alternative, and (iii) whether to continue searching. The purchase decision remains the same (purchase

[27] Our model is not a bandit superprocess because the searchable products are not independent – they are connected through t.

the product with the largest realized utility among those searched), so we omit it in the following.[28] We are able to derive an optimal solution after showing that, under a set of fairly general conditions, the optimal search order in our problem coincides with the one in Weitzman (1979).

5.3.1 Selection rule

The selection rule determines which option the consumer searches next if she decides to continue searching. If t did not affect search decisions, then the optimal selection rule would be that described in Weitzman (1979): search options in decreasing order of their reservation utility. In contrast, when t affects the search process, the selection rule may vary with t, as we show next.

Theorem 1. *The optimal selection rule depends on t.*

Proof: To show this, it suffices to provide an example in which the search order is different for two values of t. Consider any two options, j and k, and compare their search order at $t > 0$ and $t = 0$. The value of searching j or k at t is given by equation (3):

$$\begin{aligned}
\Phi_j(\bar{S},t,y) &= \max_{now,\,later}\left\{-c_{j0}-\alpha t+W_j(\bar{S},t+1,y),\ \beta\left[-c_{j0}+W_j(\bar{S},1,y)\right]\right\} \\
&= \max_{now,\,later}\{a,\ \beta A\} \\
\Phi_k(\bar{S},t,y) &= \max_{now,\,later}\{-c_{k0}-\alpha t+W_k(\bar{S},t+1,y),\ \beta[-c_{k0}+W_k(\bar{S},1,y)]\} \\
&= \max_{now,\,later}\{b,\ \beta B\}
\end{aligned}$$

Suppose that, if $t = 0$, the consumer prefers searching j before k, i.e.,

$$\Phi_j(\bar{S},0,y)-\Phi_k(\bar{S},0,y) = W_j(\bar{S},1,y)-W_k(\bar{S},1,y)-\left(c_{j0}-c_{k0}\right) = A-B > 0. \tag{5}$$

We show that this does not necessarily imply that $\Phi_j(\bar{S},t,y)-\Phi_k(\bar{S},t,y) > 0$ for any $t > 0$, i.e., the search orders at $t = 0$ and $t > 0$ may be different. There are four cases to consider:[29]

1. Suppose $a > \beta A$ and $b > \beta B$. Then

   $$\Phi_j(\bar{S},t,y)-\Phi_k(\bar{S},t,y) \;=\; a-b \;=\; W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)-\left(c_{j0}-c_{k0}\right).$$

   This difference may be positive or negative.

2. Suppose $a > \beta A$ and $b < \beta B$. Then

[28] We note that, as in Weitzman (1979), consumers can choose to purchase any of the options they searched previously – before and after any breaks they might have taken during a search spell.

[29] We only consider cases for which the expressions in equation (5) hold with inequality since other cases are straightforward to solve based on these results.

$$\Phi_j(\bar{S},t,y) - \Phi_k(\bar{S},t,y) \;=\; a - \beta B \;.$$

Since $A > B$, then $a > \beta B$, so the search order at $t > 0$ coincides with that at $t = 0$.

3. Suppose $a < \beta A$ and $b > \beta B$. Then

$$\Phi_j(\bar{S},t,y) - \Phi_k(\bar{S},t,y) \;=\; \beta A - b \;=\; \beta W_j(\bar{S},1,y) - W_k(\bar{S},t+1,y) - \beta c_{j0} + c_{k0} + \alpha t \;.$$

This difference may be positive or negative.

4. Suppose $a < \beta A$ and $b < \beta B$. Then

$$\Phi_j(\bar{S},t,y) - \Phi_k(\bar{S},t,y) \;=\; \beta A - \beta B \;.$$

This difference is positive since $A > B$, so the search order at $t > 0$ coincides with that at $t = 0$.

In sum, in cases 2 and 4, the same search order prevails for any t. However, in the other two cases, such a result is not generally true. □

In Theorem 1 we showed that the selection rule may depend on t. In other words, a consumer may want to search options in one order at t, but may want to search the same options in a different order at $t' \neq t$. The following example illustrates this result. Suppose there are two options, j and k, a consumer could search. Also suppose that, for small values of t, the consumer prefers to search option j before option k. As t increases, the consumer is more fatigued and thus more likely to want to search option j after a break than to want to search option j without a break. However, it is possible that the consumer may still want to search k without a break, implying that she may want to switch her search order and search option k before option j for larger values of t.

Such a switch does not necessarily have to occur. More precisely, for large t, if the consumer wants to search j after a break and also wants to search all other options after a break, then the same search order will prevail for all t. In other words, if the benefit from searching now versus after a break changes monotonically across options, then the optimal search order will be independent of t. In what follows, we describe a sufficient condition for the optimal search order in our problem to be the same for all t.

Condition 1. *The difference in continuation values of two options j and k is* ***monotonic*** *if, for any $t > 0$, $W_j(\bar{S},t+1,y) - W_k(\bar{S},t+1,y) \;\geq\; W_j(\bar{S},1,y) - W_k(\bar{S},1,y)$ whenever searching j before k is optimal for $t = 0$.*

In general, if the consumer wants to search j before k for $t = 0$, then she will want to search these options in the same order after a break as well (since after a break t resets to zero). Condition 1 ensures

that the consumer will also want to search j and k in the same order without a break, i.e., for $t > 0$. Using Condition 1, we can now show the following.

Theorem 2. *Under Condition 1, the optimal search order for $t > 0$ coincides with the optimal search order for $t = 0$.*

Proof: This result follows from the proof of Theorem 1. In case 1, it is straightforward to see that this monotonicity condition is sufficient for the statement to be true. Also, in cases 2 and 4, the same search order prevails for any $t \geq 0$ even absent the monotonicity condition. In case 3, the consumer prefers to search j after the break ($a < \beta A$) and k before a break ($b > \beta B$). Under the monotonicity condition, she will also prefer to search j before k before a break (since $a > b$ and $\beta A - b > 0$). Since this holds for any pair of alternatives and any values of t, our statement follows. □

5.3.2 Evaluating Condition 1

To better understand Condition 1 and to determine the optimal search order in our problem, we need to characterize the relation between the continuation value, $W_j(\bar{S}, t+1, y)$, and its arguments. This is challenging because continuation values are recursive functions of all future decisions a consumer will make (see equation (4)), and thus do not have simple closed form expressions that can be analyzed. However, as is common in finite horizon dynamic programming problems, we can make progress using backward induction.

<u>One Option Left to Search</u>

Suppose there is only one unsearched option j left. Also, suppose that t options have been searched after the latest break and that the best option observed so far is y. Then the consumer solves the following problem:

$$V(j,t,y) = \max\Big\{\, y,\ -c_{j0} - \alpha t + W_j(y),\ \beta\Big[-c_{j0} + W_j(y)\Big] \,\Big\}, \tag{6}$$

where $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$ since the consumer will stop searching after j. The continuation value here coincides with the one in Weitzman (1979). Note that $W_j(\{j\}, t+1, y) = W_j(\{j\}, 1, y) = W_j(y)$, so the continuation value has a relatively simple expression.

Two Options Left to Search

Now suppose there are two options left to search, j and k. In this case, the consumer solves the following problem:

$$V(\{j,k\},t,y) = \Big\{\, y,\ \Phi_j(\{j,k\},t,y),\ \Phi_k(\{j,k\},t,y)\ \Big\}, \tag{7}$$

where the value of searching j equals

$$\Phi_j(\{j,k\},t,y) = \max_{now,\,later} \Big\{ -c_{j0} - \alpha t + W_j(\{j,k\},t+1,y),\ \beta\Big[-c_{j0} + W_j(\{j,k\},1,y)\Big]\ \Big\}. \tag{8}$$

The continuation value $W_j(\{j,k\},t+1,y)$ is now given by

$$\begin{aligned} W_j(\{j,k\},t+1,y) &= V(k,t+1,y)F_j(y) + \int_y^\infty V(k,t+1,u)\,dF_j(u) \qquad (9)\\ &= \max\Big\{ \underbrace{y}_{\text{I}},\ \underbrace{-c_{k0} - \alpha(t+1) + W_k(y)}_{\text{II}},\ \underbrace{\beta[-c_{k0} + W_k(y)]}_{\text{III}} \Big\} F_j(y) \\ &+ \max\left\{ \underbrace{\int_y^\infty u\,dF_j(u)}_{\text{I}'},\ \underbrace{\int_y^\infty [-c_{k0} - \alpha(t+1) + W_k(u)]dF_j(u)}_{\text{II}'},\ \underbrace{\int_y^\infty \beta[-c_{k0} + W_k(u)]\,dF_j(u)}_{\text{III}'} \right\}, \end{aligned}$$

and the consumer will stop searching after j and k, explaining the presence of $W_k(\cdot)$ above.

We inspect the expression for $W_j(\{j,k\},t+1,y)$ using three approaches: first, using simulation studies, we use the function derived in equation 9 to directly check when the difference $\Big[W_j(\{j,k\},t+1,y) - W_k(\{j,k\},t+1,y)\Big] - \Big[W_j(\{j,k\},1,y) - W_k(\{j,k\},1,y)\Big]$ is non-negative for any two options j and k, i.e., when Condition 1 holds. Second, we provide additional analytical results on the generality of Condition 1 by investigating the relation between $W_j(\{j,k\},t+1,y)$ and $W_j(\{j,k\},1,y)$. And third, we compare the functional form of the continuation value in our problem to that in Weitzman (1979) by examining the relation between $W_j(\{j,k\},t+1,y)$ and $W_j(y)$. In the following, we present our main findings. A more detailed discussion with additional analyses and results in available in Web Appendix C.

Condition 1 Evaluated for a Wide Range of Parameter Values

We compute the difference $\left[W_j(\{j,k\},t+1,y)-W_k(\{j,k\},t+1,y)\right]-\left[W_j(\{j,k\},1,y)-W_k(\{j,k\},1,y)\right]$ for any two options j and k and check when it is non-negative, i.e., when Condition 1 holds. To compute this difference, we follow prior work (Kim, Albuquerque, and Bronnenberg, 2010, 2017; Honka and Chintagunta, 2017; Chen and Yao, 2017; Ursu, 2018) and assume that $F(\cdot)$ represents the normal distribution. More precisely, we consider two options, j and k, with distributions given by $N\left(\mu_j,\sigma_j^2\right)$ and $N\left(\mu_k,\sigma_k^2\right)$. Also consistent with prior work, we parameterize baseline search costs as $c_{j0}=e^{\kappa_j}$ and $c_{k0}=e^{\kappa_k}$. For consistency and to ensure a positive value, we let fatigue equal $\alpha=e^a$. The utility of the best option observed through search y must exceed the value of the outside option, which, consistent with the literature, we set to zero (Weitzman, 1979; Kim, Albuquerque, and Bronnenberg, 2010). We then vary the values of the ten parameters $\left[\mu_j,\mu_k,\sigma_j,\sigma_k,\kappa_j,\kappa_k,a,y,t,\beta\right]$ over a large parameter space, compute the difference $\left[W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)\right]-\left[W_j(\bar{S},1,y)-W_k(\bar{S},1,y)\right]$, and check when Condition 1 is satisfied.

An in-depth description of the simulation studies and detailed results can be found in Section C.1.4 in Web Appendix C; here, we provide a summary. After evaluating more than 1,000,000 parameter combinations (in three different sets of simulations), both far from and close to the values we obtain when estimating our model, we find that Condition 1 is satisfied in the vast majority (86%) of simulations, i.e., the difference $[W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)]-[W_j(\bar{S},1,y)-W_k(\bar{S},1,y)]\geq 0$. Even when Condition 1 is not satisfied, the difference $[W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)]-[W_j(\bar{S},1,y)-W_k(\bar{S},1,y)]$ is mostly a very small negative number.[30] Allowing for very small violations of up to 0.1, Condition 1 holds in more than 93% of simulations.[31] Also, Condition 1 is satisfied in all simulations (100%) when parameter values are close to our coefficient estimates.

Relation between $W_j(\{j,k\},t+1,y)$ and $W_j(\{j,k\},1,y)$

We now turn to analyzing the relation between $W_j(\{j,k\},t+1,y)$ and $W_j(\{j,k\},1,y)$ shown in equation (9). First, note that $W_j(\{j,k\},t+1,y)\leq W_j(\{j,k\},1,y),\forall t,\forall y$, since $W_j(\{j,k\},t+1,y)$ is weakly decreasing in t. And second, note that $W_j(\{j,k\},t+1,y)$ will only depend on t if the consumer chooses to frequently search options without a break. Mathematically, $W_j(\{j,k\},t+1,y)$ will only depend on t if the second

[30] Note that the calculated values of the difference $[W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)]-[W_j(\bar{S},1,y)-W_k(\bar{S},1,y)]$ in the simulations range from -9 to 9.

[31] We define a very small violation as follows: $0<[W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)]-[W_j(\bar{S},1,y)-W_k(\bar{S},1,y)]\leq -0.1$.

terms in each maximum expression in equation (9), marked as II and II′, respectively, are larger than the other two terms, marked as I and III and I′ and III′, respectively.

For example, if fatigue α is small, then the consumer is more likely to search without a break, implying that $W_j(\{j,k\},t+1,y)$ will vary with t. Similarly, if the consumer heavily discounts future searches (small β) or if y is small (e.g., if the consumer is early in her search process), then the consumer will prefer to search without a break and $W_j(\{j,k\},t+1,y)$ will vary with t. In contrast, for large values of α, β or y, the consumer will frequently take breaks from searching and the continuation value will be independent of t, i.e., $W_j(\{j,k\},t+1,y) = W_j(\{j,k\},1,y)$. This result also holds in a variety of other settings. Importantly, we show in Web Appendix C that, for the range of values we obtain when estimating our model (see Section 7), $W_j(\{j,k\},t+1,y) = W_j(\{j,k\},1,y)\ \forall t$ holds in most cases. This comes from the fact that search gaps are very prevalent in our data (see Section 3), implying a large fatigue level and that $W_j(\{j,k\},t+1,y)$ will not vary with t in most cases.[32] Having shown that $W_j(\bar{S},t+1,y) = W_j(\bar{S},1,y)\ \forall t$ when only one option is left to search (because they both equal the continuation value in Weitzman (1979), $W_j(y)$) and when two options are left to search and α, β, and/or y are large, a straightforward induction proof shows that the same result holds for any set $\bar{S}$ of unsearched options (see formal proof in Section C.1.5 in Web Appendix C). If $W_j(\bar{S},t+1,y)$ does not vary with t, then Condition 1 is trivially satisfied.

We illustrate these results using the following simulation exercise. Consider two options, j and k, with distributions given by $N(\mu_j,\sigma_j^2)$ and $N(\mu_k,\sigma_k^2)$, where $\mu_j = 2$, $\mu_k = 1$, $\sigma_j^2 = \sigma_k^2 = 1$. Suppose baseline search costs equal $c_{j0} = c_{k0} = e^{-2}$ and that consumers discount future searches at the rate $\beta = 0.95$. Consistent with the literature, the utility of the best option observed through search y is at least zero (Weitzman, 1979; Kim, Albuquerque, and Bronnenberg, 2010). In Figure 3, we show how the value of $W_j(\{j,k\},t+1,y)$ varies with t for different levels of fatigue α. For small fatigue levels, e.g., $\alpha = e^{-6}$ in Figure 3(a), consumers search an option without a break and, as expected, the continuation value varies with t for relatively small values of y. When y is large, consumers are more likely not to search and the continuation value will not depend on t.

==========
Insert Figure 3 about here
==========

For large fatigue levels, e.g., $\alpha = e^{-1}$ in Figure 3(b), the continuation value is independent of t, i.e.,

[32] Note that this result does not require that consumers take breaks after every option searched. Rather, it says that $W_j(\{j,k\},t+1,y)$ and $W_j(\{j,k\},1,y)$ are good approximations for each other when consumers take frequent breaks.

$W_j(\{j,k\},t+1,y) = W_j(\{j,k\},1,y)\ \forall\ t$, since consumers need to take frequent breaks because of fatigue. The same result holds in our empirical application: in our data, search gaps are very prevalent, implying a large fatigue level and thus that $W_j(\{j,k\},t+1,y)$ will not vary with t in most cases (for more details, see Web Appendix C). Note again that the relation $W_j(\{j,k\},t+1,y) = W_j(\{j,k\},1,y)$ does not only hold when fatigue or y are large. As reported above and in Web Appendix C, in many empirically relevant cases, $W_j(\{j,k\},t+1,y)$ and $W_j(\{j,k\},1,y)$ are good approximations for each other, implying that Condition 1 is satisfied.

Relation between $W_j(\{j,k\},t+1,y)$ and $W_j(y)$

The third set of results we derive based on equation (9) concerns the relation between $W_j(\{j,k\},t+1,y)$ and $W_j(y)$, the continuation values in our and in the Weitzman (1979) problem, respectively. First, it follows from equation (9) that $W_j(\{j,k\},t+1,y)\ \geq\ W_j(y),\ \forall t, \forall y$.[33]

Second, for large (finite) values of y, $F_j(y) \approx 1$, implying that $W_k(y) \approx y$, and thus that the continuation value can be approximated by $W_j(\{j,k\},t+1,y)\ \approx\ \max\left\{y,\ -c_{k0} - \alpha(t+1) + y,\ \beta\left[-c_{j0} + y\right]\right\} = y$ since $\alpha, \beta > 0$, $t \geq 0$, baseline search costs are positive, and y is large and positive. As a result, for large values of y, the continuation value in our problem coincides with the one in Weitzman (1979). And third, for small (finite) values of y, $F_j(y) \approx 0$. In this case, the continuation value can be approximated by $W_j(\{j,k\},t+1,y)\ \approx\ \max\left\{E_{F_j}(u),\ -c_{k0} - \alpha(t+1) + \int_y^\infty W_k(u)\,dF_j(u),\ \beta\left[-c_{j0} + \int_y^\infty W_k(u)\,dF_j(u)\right]\right\}$ since $W_j(y) \approx E_{F_j}(u)$ for small values of y (where $E_{F_j}(u)$ is the expected value with respect to $F_j(\cdot)$). A sufficient condition for the continuation value in our problem to coincide with the one in Weitzman (1979) is that $E_{F_j}(u)\ \geq\ \int_y^\infty W_k(u)\,dF_j(u)\ \geq\ 0$.

We illustrate these results using a simulation exercise. Once again, consider two options, j and k, with distributions given by $N\left(\mu_j, \sigma_j^2\right)$ and $N\left(\mu_k, \sigma_k^2\right)$. In this case, the continuation value in Weitzman is given by $W_j(y)\ =\ \mu_j\sigma_j + \left(y - \mu_j\sigma_j\right)\Phi(m_j) + \sigma_j^2\phi\left(m_j\right)$ with $m_j = \frac{y-\mu_j}{\sigma_j}$. To compare $W_j(\{j,k\},t+1,y)$ with $W_j(y)$, we fix $t = 1$ and illustrate one of the largest differences between the two continuation values (recall that $W_j(\{j,k\},t+1,y)$ is weakly decreasing in t). Figure 4(a) shows the value of $W_j(\{j,k\},t+1,y)$ relative to $W_j(y)$ for $\mu_j = 2$, $\mu_k = 1$, and $\sigma_j = \sigma_k = 1$. As expected, for large values of y, $W_j(\{j,k\},t+1,y) = W_j(y)$. For smaller values of y, we see that $W_j(\{j,k\},t+1,y) \geq W_j(y)$, but the difference is very small. In Figure 4(b), we change the parameter values to $\mu_j = 1$ and $\mu_k = 2$. For large values of y, $W_j(\{j,k\},t+1,y)$

[33] The result follows from the fact that $\max\{a,b,c\} + \max\{d,e,f\} \geq \max\{a+d, b+e, c+f\}$ for any values (a,b,c,d,e,f).

equals $W_j(y)$. For smaller values of y, we observe that $W_j(\{j,k\},t+1,y) \geq W_j(y)$ with a slightly higher difference between the two continuation values.

==========
Insert Figure 4 about here
==========

In Section C.1.4 in Web Appendix C, we perform additional simulation exercises varying more of the parameters of interest and find that the difference between $W_j(\{j,k\},t+1,y)$ and $W_j(y)$ is zero or very small for the vast majority of parameter values (77% or 98%, respectively).

Our Proposed Selection Rule

Motivated by these results, we make the following assumption:[34]

Assumption 1. $W_j(\bar{S},t+1,y) \;=\; W_j(y) \qquad\qquad \forall j \in \{1,\ldots,J\},\ \forall t \geq 0\,.$

Using Assumption 1, we can now characterize the optimal selection rule in our problem.

Theorem 3. *Under Assumption 1, the optimal selection rule coincides with the one in Weitzman (1979).*

Proof: Under Assumption 1, Condition 1 is trivially satisfied. Consumers search any two options in the same order for any $t \geq 0$. When $t = 0$, the search candidate is option $j^* = \arg\max_{j\in\bar{S}} -c_{j0} + W_j(y)$, the same as in Weitzman (1979). □

Theorem 3 describes when the search order in our model coincides with the search order in the Weitzman (1979) model. To paraphrase, Theorem 3 chronicles when consumers in our model search options in a decreasing order of reservation utilities as computed in Weitzman (1979), i.e., as the unique solution z_j to $c_{j0} = W_j\left(z_j\right) - z_j$. This allows us to state the selection rule for our problem as follows:

Selection rule. *Under Assumption 1, if the consumer chooses to search an option at time t, it will be option $j \in \bar{S}$ with the highest reservation utility z_j, where z_j is the unique solution to $c_{j0} = W_j\left(z_j\right) - z_j$.*

5.3.3 Selection and Search Rules

Using the results from the previous section, we now know that the consumer searches the option with the largest reservation utility, i.e., $j^* = \arg\max_{j\in\bar{S}} z_j$, if she decides to continue searching. Thus we

[34] Similar assumptions have been made in recent work modeling consumer search decisions (e.g., Hodgson and Lewis 2021).

can solve the problem in equation (2) in two steps: first, the consumer determines j^*. And second, she solves equation (2) for j^* and determines whether to stop searching and whether to search j^* immediately or after a break. The search problem in the second stage reduces to

$$V(j^*,t,y) = \max_{stop,\, now,\, later} \left\{ y,\ -c_{j^*0} - \alpha t + W_{j^*}(y),\ \beta\left[-c_{j^*0} + W_{j^*}(y)\right] \right\}. \tag{10}$$

The next result on the optimal selection and search rules follows directly from equation (10) and Theorem 3.

Theorem 4. *Under Assumption 1, the following selection and search rules are optimal for $t > 0$:*[35]

1. ***Selection rule:*** *order options in decreasing order of reservation utilities z_j (defined by Weitzman 1979).*

2. ***Search rules:*** *if j is the option with the maximum reservation utility among the options not yet searched $\bar{S}$, then given (t,y)*

   - *search j now if $c_{j0} \leq W_j(y) - y - \alpha t$ and $c_{j0} \leq W_j(y) - \frac{\alpha t}{1-\beta}$;*
   - *search j later if $c_{j0} \leq W_j(y) - \frac{y}{\beta}$ and $c_{j0} > W_j(y) - \frac{\alpha t}{1-\beta}$;*
   - *stop searching if $c_{j0} > W_j(y) - y - \alpha t$ and $c_{j0} > W_j(y) - \frac{y}{\beta}$.*

Proof: As shown in Theorem 3, the selection rule above is optimal. The statements describing the search rules follow directly from equation (10). □

For the product searched first, $t = 0$ and $V(\bar{S}, 0, u_0) = \max\{u_0, \max_{j\in\bar{S}} -c_{j0} + W_j(u_0)\}$, so the consumer chooses between searching the product now or stopping search. Search occurs if the condition that $c_{j0} \leq W_j(u_0) - u_0$ holds for product j with the largest reservation utility.

Using Theorem 4, we can now further characterize the solution to our problem. Recall that $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$. Taking appropriate limits, we find that $W_j(\infty) = \infty$ and $W_j(-\infty) = E_{F_j}(\cdot)$. Further, we can show that $\partial W_j(\cdot)/\partial y = F_j(y) \geq 0$. As such, $W_j(y)$ is a continuous and monotonically increasing function of y. It follows (as shown by Weitzman 1979) that $W_j(y) - y$ is a continuous and monotonically decreasing function of y. Therefore, there exists a unique solution z_j to the equation

$$c_{j0} = W_j\left(z_j\right) - z_j, \tag{11}$$

[35] We break ties as follows: the consumer prefers to search now if choosing between any of the three options, and prefers to search later rather than to stop.

which represents the reservation utility of an option (Weitzman 1979). Similarly, there exist unique solutions to each of the inequalities describing the consumer search rules above. More precisely, because search costs $c_{j0} > 0$ and the value of αt is constant in y, while $W_j(y) - y$ is continuous and monotonically decreasing in y, there exists a unique solution $z_j^1(t)$ to the equation

$$c_{j0} = W_j\left(z_j^1(t)\right) - z_j^1(t) - \alpha t. \tag{12}$$

Further, whenever $\beta \neq 0$, there exists a unique solution z_j^2 to the equation

$$c_{j0} = W_j\left(z_j^2\right) - \frac{z_j^2}{\beta} \tag{13}$$

And finally, whenever $\beta < 1$, there exists a unique solution $z_j^3(t)$ to the equation

$$c_{j0} = W_j\left(z_j^3(t)\right) - \frac{\alpha t}{1-\beta}. \tag{14}$$

5.4 Our Proposed Solution

Based on the results derived in the previous subsection, our proposed solution is as follows:

Theorem 5. *Under Assumption 1, a consumer's optimal search strategy for $t > 0$ is:*

1. ***Selection rule:*** *order options in decreasing order of reservation utilities z_j (defined by Weitzman 1979).*

2. ***Search rules:*** *if j is the option with the maximum reservation utility among options not yet searched $\bar{S}$, then given (t, y)*

   - *search j now if* $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ *and* $z_j^3(t) < y$ *;*
   - *search j later if* $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ *and* $z_j^3(t) \geq y$ *;*
   - *stop searching if* $z_j^1(t) < y$ *and* $z_j^2 < y$ *.*

3. ***Choice rule:*** *upon stopping, purchase the option with the largest realized value among those searched, $y = \max_{j \in S} u_j$, or choose the outside option of not purchasing.*

Proof: As shown in Theorem 3, the selection rule above is optimal. Also, the choice rule is not affected by the presence of fatigue, and is therefore optimal as per Weitzman (1979). The search rules follow directly from equation (10) and Theorem 4. A complete derivation of the specific functions describing these search rules is presented in Section C.2 in Web Appendix C. □

Note that the consumer searches fewer options for larger values of fatigue, α. To see this, observe that $z_j^1(t)$ decreases in α in equation (12). Therefore, larger fatigue levels (weakly) decrease the number

of searched options, since the condition that $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ is less likely to hold. Finally, for the product searched first, we know that $t = 0$ and search occurs if $z_j^1(0) = z_j \geq u_0$ for product j with the largest reservation utility.

6 Empirical Application

6.1 Empirical Model

We take the theoretical model presented in the previous section to data using the following empirical specification: we model consumers as searching across websites (e.g., zalando.nl or nike.com). Specifically, consumer $i = 1, \ldots, N$ seeks to purchase from website $j = 1, \ldots, J$ or to choose the outside option of not purchasing (denoted by $j = 0$). Consumer i's utility for website j is given by

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij} \\ &= w_j + \gamma X_{ij} + \eta_{ij} + \epsilon_{ij} \end{aligned} \tag{15}$$

where v_{ij} denotes the information the consumer has about a website before searching it, while ϵ_{ij} denotes the information she searches for. Before searching, the consumer knows individual websites' values which are denoted by website intercepts w_j. In each subcategory, we estimate separate website intercepts for the 10 most searched websites (accounting for approximately 65% of clicks in each subcategory) and group all other websites into a composite reference website called "Other." Although we scraped prices and other product features from the URLs provided in our data, these features do not generally vary over time or across consumers in the apparel industry (see also Section 4.4). Therefore, after controlling for website intercepts, the effects of such features cannot be separately estimated. Nevertheless, we include two additional controls X_{ij} in the utility function: (i) the number of times the consumer has previously searched a given website (across all product subcategories) to measure her loyalty for and knowledge of a website, and (ii) an indicator for whether the consumer visited a price discount page to partially capture her price sensitivity.

Next, η_{ij} is the part of the utility that is observed by the consumer (prior and post search), but not the researcher (neither prior nor post search). It captures deviations from website features that the consumer may be aware of before starting her search. Consumers search sequentially to resolve uncertainty about their match values ϵ_{ij}, i.e., consumers do not observe ϵ_{ij} prior to search but they do post search. The researchers neither observes ϵ_{ij} prior nor post search. ϵ_{ij} captures everything the

consumer learns by visiting the website, e.g., available (actual) product styles, sizes, colors, customer reviews, photos, etc. η_{ij} and ϵ_{ij} are both standard normally distributed. The outside option does not require searching and is modeled as $u_{i0} = q_0 + \eta_{i0}$, where q_0 is an intercept denoting the value of not purchasing.

Searching to resolve uncertainty about ϵ_{ij} is costly to consumers. Search costs (per search) are given by

$$\begin{aligned} c_i &= c_0 + \alpha_i t \\ &= \exp(\kappa_0) + \exp(\lambda_0 + \lambda_1 Age_i)\, t \end{aligned} \tag{16}$$

where c_0 are baseline search costs and $t \geq 0$ captures the number of searches performed without a break. To ensure that search costs are positive, we operationalize search costs as exponential functions (see, e.g., Honka 2014; Chen and Yao 2017; Ursu 2018). Lastly, consistent with our results from Table 5, we use age as a shifter of fatigue.

6.2 Estimation

We use the search rules from Section 5.3.3 to construct the likelihood of consumers' search and purchase decisions. These rules translate into the following restrictions on preferences, search costs, and fatigue parameters. Suppose a consumer i searched a number of options s of the J websites available and she chose j after stopping her search (including the outside option). With a slight abuse of notation, order websites by their reservation utilities and let n denote the website with the nth largest reservation utility. Also, let t_n denote the number of websites the consumer searched since the previous gap and before searching n.

Since consumers search websites in a decreasing order of their reservation utilities, according to the *selection rule*, it must be that

$$z_{in} \geq \max_{k=n+1}^{J} z_{ik} \qquad \forall n \in \{1, \ldots, J-1\}. \tag{17}$$

After searching the first website, $t_n > 0 \;\; \forall n > 1$ and the *search rules* describe consumer behavior. For searched website n we know that

$$\max\left\{z_{in}^1(t_n), z_{in}^2\right\} \geq \max_{k=0}^{n-1} u_{ik} \qquad \forall n \in \{2, \ldots, s\}. \tag{18}$$

For the website searched first (for which $t_1 = 0$), the search rules require that its reservation utility

exceeds the utility of the outside option, i.e., $z_{i1} \geq u_{i0}$. Since all consumers in our data search at least once, consistent with prior work (e.g., Honka 2014; Honka and Chintagunta 2017), we assume that the first search is free. Note that we allow for the possibility of no search in our Monte Carlo simulation in Section 6.4.

What separates our model from previous work is that we additionally capture a consumer's decision of when to search an option (with or without a break). In particular, all websites, except the one searched first, may be searched with or without a gap. Thus, if n was searched without a break, i.e., $t_n = t_{n+1} - 1$, we know that

$$z_{in}^3(t_n) \; < \; \max_{k=0}^{n-1} \; u_{ik} \qquad \forall n \in \{2, \ldots, s\}, \tag{19}$$

while, if n was searched after a break, i.e., if $t_n \neq t_{n+1} - 1$, it must be that

$$z_{in}^3(t_n) \; \geq \; \max_{k=0}^{n-1} \; u_{ik} \qquad \forall n \in \{2, \ldots, s\}. \tag{20}$$

For all options m that were not searched, it must be that

$$z_{im}^1(t_m) \; < \; \max_{k=0}^{s} \; u_{ik} \qquad \forall m \in \{s+1, \ldots, J\}, \tag{21}$$

$$z_{im}^2 \; < \; \max_{k=0}^{s} \; u_{ik} \qquad \forall m \in \{s+1, \ldots, J\}. \tag{22}$$

with $t_m = t_s + 1$.

Finally, consistent with the *choice rule*, if the consumer chooses j (including the outside option), her utility from this choice exceeds that of all searched websites, i.e.,

$$u_{ij} \; = \; \max_{k=0}^{s} \; u_{ik} \qquad \forall j \in \{0, 1, \ldots, s\}. \tag{23}$$

If the consumer searches using the rules described above, then she makes search, search gap, and purchase decisions jointly. Thus, the probability of observing a certain outcome for consumer i is characterized by the joint probability of equations (17)–(23) holding. This probability is given by

$$L_i = Pr(\text{ Selection rule}_i, \text{ Search rule}_i, \text{ Choice rule}_i \text{ }). \tag{24}$$

Because consumers make these decisions jointly, the likelihood function does not have a closed-form solution. We use a simulated maximum likelihood esitmation (SMLE) approach to infer the parameters of the model. In choosing the simulation method, we follow McFadden (1989), Honka

(2014), Honka and Chintagunta (2017), Ursu (2018), and Ursu, Wang, and Chintagunta (2020) and use the logit-smoothed AR simulator. The implementation details are discussed in Web Appendix D.

An advantage of our proposed method lies in its ease of estimation due to its similarity to the Weitzman (1979) model: consumers search in decreasing order of reservation utilities z_j and also make search and purchase decisions based on threshold values of the best alternative observed so far. The main difference consists of computing the values of $\left[z_j^1(t), z_j^2, z_j^3(t)\right]$ in addition to that of z_j. We describe how to compute these values in Web Appendix D.

We estimate our model using data from the two most commonly purchased subcategories in our data: "shirts, tops, & blouses" and "shoes" to demonstrate that our results are not limited to a specific subcategory. Details on the construction of the estimation samples are provided in Web Appendix A.

6.3 Identification

The set of model parameters is composed of the utility parameters w_j and γ, baseline search cost c_0 (parameterized by κ_0), search fatigue α (parameterized by λ_0 and λ_1), and the discount factor β. As is well-known, the discount factor is typically not identified in dynamic discrete choice models without further restrictions (Rust 1994; Magnac and Thesmar 2002). This is the case in our model as well. Prior work either fixes the discount factor to a value close to 1 (e.g., 0.995 per week in Erdem and Keane 1996 and 1 in Kim, Albuquerque, and Bronnenberg 2010) or makes additional model assumptions and estimates discount factors that are close to 1 (e.g., 0.992 per week in Hotz and Miller 1993 or 0.98 per week in Ackerberg 2003). However, using a series of field studies, Yao et al. (2012) show that implied discount rates are generally lower, on average 0.9 per week. Since the median search gap in our data is approximately 4 days long, the 0.9 weekly discount factor reported in Yao et al. (2012) corresponds to a discount factor of 0.94 for 4 days. Based on this result, we fix the discount factor to 0.95 in our main estimation.[36,37] Given these considerations, the set of parameters to be estimated is $\theta = (w_j, \gamma, \kappa_0, \lambda_0, \lambda_1)$.

As is standard in consumer search models, utility parameters are identified from search and purchase frequencies observed in the data. For example, websites that are searched and purchased

[36] We evaluate the robustness of our results with respect to the discount factor by re-estimating our model under three alternative discount factors: 0.85, 0.90, and 0.99. The results are displayed in Table F-2 in Web Appendix F. Utility and baseline search cost parameters are very similar across all specifications. The fatigue parameter is estimated to be higher for lower values of the discount factor.

[37] Consumers may also vary in the degree to which they discount future utility. As noted, the discount factor is not identified in our model. Nevertheless, by estimating our model with discount factors in the range [0.85, 0.99], we recover parameter estimates under multiple discount factors.

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.

Similar to prior work, search costs do not affect purchase decisions and are identified from the number of websites that consumers search. More precisely, the search rules impose an upper and a lower bound on search cost c_0 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 reservation utility expressions derived in equations (12), (13), and (14).

By observing search gaps, i.e., consumers' decisions of *when* to search each website, we can additionally identify consumer fatigue levels. In other words, conditional on an observed number of searches, search gaps identify the fatigue parameter λ_0. Finally, deviations in search gaps across consumers attributable to age identify λ_1.

6.4 Monte Carlo Simulation

To show that our estimation procedure can recover the model parameters, we perform the following Monte Carlo simulation exercise. We generate a data set of 5,000 consumers making choices among five options (one outside option and four websites). Consumers value each website differently. Search costs have two components: a baseline level of search costs and fatigue. The true values of the utility and both search cost parameters are similar to those from a preliminary estimation of our model.

For estimation, we follow the steps described in Section 6.2 and use 200 draws from the distribution of the utility error terms (both η_{ij} and ϵ_{ij}) for each consumer-website combination. We simulate 50 different data sets using the same true parameters but different seeds for the utility error terms and repeat the estimation for each data set.

Our Monte Carlo simulation results are displayed in Table 6. In column (i), we present the true parameters; in column (ii), we show the mean of the estimated parameters across the 50 simulations and the standard deviation of the mean across these simulations. Our proposed estimation procedure recovers the model parameters well. In addition, in column (iii), we also report results from the Weitzman model that ignores search gaps. We obtain these results by estimating the model on the same data, but assuming no gaps occurred in the data.[38] We find that the Weitzman model performs

[38] See Web Appendix D for details on the Weitzman (1979) model estimation.

well in recovering the true utility parameters, but overestimates baseline search costs. We discuss the estimation bias observed here after presenting our estimation results in the next section.[39]

==========
Insert Table 6 about here
==========

7 Results

7.1 Model Estimates

We show the estimation results for the "shirts, tops, & blouses" and "shoes" subcategories in Table 7. For each subcategory, the first two columns display results from our model that accounts for search gaps, while the third column reports results from the Weitzman model that ignores search gaps.[40] The results indicate that consumers derive positive utility from the outside option, consistent with the empirical observation that most consumers do not make a purchase. Next, we find that Zalando, the largest online retailer in the Netherlands, is among the most preferred websites in both subcategories, together with C&A in the "shirts, tops, & blouses" subcategory and Schurrman Shoenen in the "shoes" subcategory.[41] As expected, previous visits to a website increase a consumer's utility for that website. Although we are not able to account for the effect of prices, we show consumers' price sensitivity by reporting that consumers who visited the price discount page (e.g., the "sale" or "clearance" page) of a website derive higher utility.

==========
Insert Table 7 about here
==========

The baseline search cost estimate is relatively small compared to the fatigue estimate. To put it differently, fatigue has a large effect on consumer search decisions, equivalent to increasing baseline search costs at least tenfold with every searched website. Also, the larger number of search gaps in the "shoes" subcategory reveals a larger fatigue level (as a proportion of baseline search costs) for consumers in that subcategory (the ratio of fatigue to baseline search costs is 14 in the "shirts, tops, & blouses" subcategory and it is 18 in the "shoes" subcategory).[42] Finally, consistent with our reduced-form results in Table 5, older consumers have even larger fatigue costs.

[39] The difference in the reported log-likelihood values comes from the difference in the likelihood functions and the number of parameters to be estimated. As detailed in Section 7.3, log-likelihood values may not be reliable measures of model fit.

[40] We also estimated a version of our model in which we set $\alpha = 0$ and assume no gaps occurred in the data. The results are very similar to those from the Weitzman model and available from the authors upon request.

[41] Recall that utility estimates for the top 10 websites are relative to a reference website comprising all other websites.

[42] Calculation follows after dividing the fatigue constant by the baseline search cost parameter, e.g., in the first column, $\exp(-2.4436)/\exp(-5.0740) \approx 14$.

7.2 Comparison with Weitzman (1979)

By comparing our results to those from the Weitzman (1979) model that ignores search gaps, we provide insights into the estimation bias that arises when search gaps are ignored.[43] In particular, both our estimation and our simulation results show that the utility parameter estimates are similar in our and in the Weitzman (1979) model, but that baseline search costs are overestimated in the latter one.[44] The intuition behind this difference is as follows: recall that the Weitzman model ignores search gaps and assumes that search costs are independent of the number of previously searched options. Thus, when estimated on the same data set as a model in which fatigue affects search costs, the Weitzman model rationalizes the same number of searched products by inflating baseline search costs.

Although the differences in intercept estimates for a particular website in our and in the Weitzman (1979) model are small, theoretically the intercepts for a particular website in both models are different. Whether this difference in intercepts is positive or negative depends on whether a particular website is predominantly searched before or after search gaps. In our empirical application, the Weitzman (1979) model provides directionally larger estimates of website intercepts than our model. This finding is consistent with the observation that many websites are searched after a break in our data.

Our model makes two changes to the Weitzman (1979) framework: (i) allows for search gaps; and (ii) models the effect of fatigue on search costs. To better isolate the effect of each model change on parameter estimates, we also estimate a (variation of the) Weitzman (1979) model with increasing search costs (due to fatigue) but without search gaps, i.e., we only make one change to the Weitzman (1979) framework. To the best of our knowledge, such a variation of the Weitzman (1979) model has not been studied by previous literature. As in our model, there is no known optimal search rule. However, in Web Appendix F, we describe how the solution we developed for our model can be used to derive an optimal search rule for the Weitzman (1979) model with increasing search costs.

The estimation results are displayed in Table F-1 in Web Appendix F. Not accounting for search gaps leads to an overestimation of the baseline search cost, although the bias is smaller when at least fatigue is taken into account. In addition, a Weitzman (1979) model with increasing search cost underestimates the importance of fatigue compared to a model that accounts for search gaps. The intuition for this result is as follows: when breaks are allowed, the number of previously searched

[43] We use the term "bias" to describe the difference in estimates based on our results.

[44] Note that, if we had estimated the Weitzman (1979) model at the session level rather than at the spell level, similar to some of the literature that does not observe search gaps (e.g., Ursu 2018), then our results would have shown an even larger difference.

options is equal or smaller than the number of previously searched options when breaks are not allowed. Thus, when breaks are not allowed, it appears as if fatigue played a smaller role in the consumer's decision to search, since the consumer chooses to continue searching despite a higher fatigue level, resulting in a smaller fatigue parameter estimate. In contrast, when breaks are allowed, fatigue causing the consumer to take a break from searching (in addition to increasing search costs) reveals the larger importance of fatigue.

7.3 Model Fit

Since the likelihood functions are different, we cannot rely on the log-likelihood measures reported in Table 7 to compare our model to the Weitzman (1979) model. Instead, to understand which model better captures consumer behavior, we show both predicted consumer decisions in the three areas (search gaps, searches, and purchases) and calculate the root mean squared error (RMSE). For these calculations, we use the first set of estimation results reported in Table 7 for each subcategory for our model and the estimation results for the Weitzman (1979) model. The predictions and RMSE results are displayed in Table 8.[45]

==========

Insert Table 8 about here

==========

Our model recovers the number of search gaps well, while the Weitzman (1979) model (by construction) cannot predict any such decisions. When comparing the search shares of each website, our model more accurately predicts which websites consumers search more frequently. Finally, although the utility estimates in the two models are very similar, the Weitzman (1979) model better predicts market shares. In part, this difference also arises from the fact that our model explains three rather than only two decisions that consumers make.

8 Counterfactuals

8.1 The Impact of Fatigue

In this counterfactual, we measure the impact of fatigue (and thus search gaps) on market outcomes such as consumer search and purchases and how it compares to the effect of baseline search costs. To

[45] The RMSEs for the Weitzman (1979) model with increasing search costs are as follows: for "shirts, tops, & blouses," 58 and 44 for search and market shares, respectively. For "shoes," 71 and 52 for search and market shares, respectively. RMSE for search gaps is the same as in the Weitzman (1979) model.

accomplish this goal, for each of the two apparel subcategories, we employ our model and coefficient estimates to simulate consumer decisions regarding searches, search gaps, and purchases in four scenarios (holding everything else constant): (i) when fatigue is reduced by 50% and 90%;[46] and (ii) when baseline search costs are reduced by 50% and 90%. Consumers' simulated decisions in these scenarios are then compared to the current setting in which no such change occurs. To integrate over the distribution of unobserved utility shocks in the model, we repeat the simulation 50 times and report the mean results.

We present our findings in Table 9. Columns (i) and (ii) display the effects of fatigue reductions by 50% and 90%, respectively. The effects are similar in both apparel subcategories although they are larger in the "shoes" subcategory for which we found higher fatigue levels. In particular, we find that decreasing fatigue by 50% increases the number of searched websites by $1-4\%$, increases transactions by $0.5-1.2\%$, and lowers search gaps by $11-22\%$. A fatigue reduction by 90% magnifies the effect sizes by a factor of 3 to 14. For comparison, columns (iii) and (iv) present the effects of baseline search cost reductions by 50% and 90%, respectively, a policy that most prior work focused on (e.g., Seiler 2013, Honka 2014, Moraga-González, Sándor, and Wildenbeest 2018, Yavorsky, Honka, and Chen 2021). The effects of baseline search cost reductions are smaller than those of fatigue reductions. Further, in contrast to the effect of reducing fatigue, decreasing baseline search costs increases search gaps: the more options the consumer searches, the higher the chances of her taking a break from searching, and thus the more gaps.

==========

Insert Table 9 about here

==========

Next, we investigate whether and how the effects of a fatigue reduction vary across websites. For this analysis, we utilize our results from the case in which fatigue is decreased by 90% to better highlight the magnitude of the effects. Our results for both subcategories are displayed in Figure 5. The websites are ordered based on their predicted search or predicted market shares, respectively (the rankings are very similar to the ones shown in Table 3). Not surprisingly, all websites benefit from a reduction in fatigue. However, while the impact on consumer search is relatively equal across websites (in %), smaller and less popular websites benefit relatively more in terms of transactions than larger and more popular websites. The reason for this finding is that a reduction in fatigue leads to

[46] Improvements in website design to increase readability or an increase in a website's loading speed represent two potential avenues to reduce consumer fatigue based on the evidence in Table 5.

additional searches and purchases that would not have occurred for smaller websites if fatigue levels had been higher.

==========
Insert Figure 5 about here
==========

To summarize, we find that fatigue has a large impact on market outcomes such as search and purchase shares. In our empirical context, the effect of fatigue is larger than the effect of baseline search cost. And lastly, while all websites suffer from high fatigue levels, larger and more popular website are less negatively affected than smaller and less popular websites.

8.2 When Breaks Do Not Decrease Fatigue

Although search costs due to fatigue reset after a break in our model, it is not a forgone conclusion that consumers can always lower their fatigue levels during a break. On the contrary, several recent articles talk about the consequences of consumers being constantly stimulated by marketing activities[47] and of the Covid-19 pandemic keeping consumer fatigue levels high for a prolonged period.[48] In such cases, an important question is how an inability to decrease fatigue levels affects consumer decisions.

In our second counterfactual, we measure the effects of consumers' inability to reduce their fatigue levels via a search gap. Our simulation procedure is similar to the one described for the first counterfactual; however, the analyzed scenario differs. More precisely, we simulate consumer search and purchase decisions for the case in which the fatigue level α does not reset to zero during a break. In this case, consumers always prefer to search now rather than to delay their search, since a delay only leads to a discounted value of the same expected utility they would get if they searched now. Therefore, the problem consumers solve in this case reduces to that of deciding whether to search now or whether to stop. The optimal decision rule follows after inspecting equations 2 and 3 without the option to search later. In particular, the consumer continues to search website j if $z_j^1(t) \geq y$ and stops otherwise.

Not being able to reset fatigue during a break leads to a significant reduction in the number of searched and purchased products: the number of searches decreases by 21% and 23% and purchases decline by 8% and 6% in the "shirts, tops, and blouses" and "shoes" subcategories, respectively. Both

[47] This information is available in the following articles: https://hbr.org/to-keep-your-customers-keep-it-simple and https://www.nytimes.com/do-you-suffer-from-decision-fatigue.

[48] See, e.g., https://www.delish.com/grocery-shopping-brand-loyalty-reason/, https://hbr.org/how-to-combat-zoom-fatigue, and https://hbr.org/coping-with-fatigue-fear-and-panic-during-a-crisis.

of these effects are large, e.g., in comparison to those obtained in the previous counterfactual. We display the website-specific predictions for both subcategories in Figure 6. The websites are ordered based on their predicted search or predicted market shares, respectively (the rankings are very similar to the ones shown in Table 3). Although all websites suffer in this setting, different websites are affected differently, i.e., the effect is not "neutral." Consumers not being able to reduce fatigue during breaks hurts larger and more popular websites less than smaller ones. In other words, website prominence becomes more important. Consumers prefer to search and choose more often websites they are familiar with.

==========
Insert Figure 6 about here
==========

9 Conclusion

In this paper, we document that consumers frequently decide to take breaks during their search process and provide model-free evidence that such breaks are related to fatigue. To rationalize such behavior, we develop a model of sequential search that extends the Weitzman (1979) framework and allows for search gaps due to fatigue. We quantify the effect of fatigue on consumer search and purchase decisions and show the possible estimation bias in search costs when search gaps are ignored. Finally, we illustrate the impact of search gaps and consumer fatigue on market outcomes through a series of counterfactuals.

There are several limitations and potentially useful extensions to our approach. First, it would be interesting to explore other drivers of search gaps across a variety of product categories. Our model can provide a starting point for formalizing the mechanism behind search gaps in such settings. Second, future work could model the length of search gaps in addition to their occurrence. Third, fatigue levels from a website visit might depend on the number of investigated products on a website. We leave such a model extension for future research. And finally, extending our model to account for brand-specific fatigue could provide another interesting avenue for researchers to explore. We leave these and other related topics to future research.

References

Abaluck, J. and J. Gruber (2013), "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 101 (4), 1180–1210.

Ackerberg, D. (2003), "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44 (3), 1007–1040.

Anderson, S. and R. Renault (2018), "Firm Pricing with Consumer Search," in L. Corchón and M. Marini (Editors), "Handbook of Game Theory and Industrial Organization," Edward Elgar, volume 2, 177–224.

Arbatskaya, M. (2007), "Ordered Search," *RAND Journal of Economics*, 38 (1), 119–126.

Armstrong, M., J. Vickers, and J. Zhou (2009), "Prominence and Consumer Search," *RAND Journal of Economics*, 40 (2), 209–233.

Baye, M., J. Morgan, and P. Scholten (2006), "Information, Search, and Price Dispersion," *Handbook on Economics and Information Systems*, 1, 323–377.

Bednar, S. and K. Rouse (2019), "Does Taking a Break Make Students (Un)Productive?" Working Paper.

Bronnenberg, B., J. Kim, and C. Mela (2016), "Zooming in on Choice: How Do Consumers Search for Cameras Online?" *Marketing Science*, 35 (5), 693–712.

Brown, D. and J. Smith (2013), "Optimal Sequential Exploration: Bandits, Clairvoyants, and Wildcats," *Operations Research*, 61 (3), 644–665.

Carlin, B. and F. Ederer (2019), "Search Fatigue," *Review of Industrial Organization*, 54 (3), 485–508.

Carpenter, S. and C. Yoon (2011), "Aging and Consumer Decision Making," *Annals of the New York Academy of Sciences*, 1235 (1), E1–E12.

Cavallo, A. (2017), "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers," *American Economic Review*, 107 (1), 283–303.

Chen, Y. and S. Yao (2017), "Sequential Search with Refinement: Model and Application with Click-stream Data," *Management Science*, 63 (12), 4345–4365.

Dang, C., R. Ursu, and P. Chintagunta (2020), "Search Revisits," Working Paper.

De los Santos, B. and S. Koulayev (2017), "Optimizing Click-Through in Online Rankings with Endogenous Search Refinement," *Marketing Science*, 36 (4), 542–564.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012), "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102 (6), 2955–80.

Dhar, R. (1997), "Context and Task Effects on Choice Deferral," *Marketing Letters*, 8 (1), 119–130.

Dong, X., I. Morozov, S. Seiler, and L. Hou (2020), "Estimation of Preference Heterogeneity in Markets with Costly Search," *Marketing Science*, forthcoming.

Ellison, G. and S. Ellison (2009), "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77 (2), 427–452.

Ellison, G. and A. Wolitzky (2012), "A Search Cost Model of Obfuscation," *RAND Journal of Economics*, 43 (3), 417–441.

Erdem, Tülin and Michael P Keane (1996), "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1–20.

Gittins, J., K. Glazebrook, and R. Weber (2011), *Multi-Armed Bandit Allocation Indices*, John Wiley and Sons Inc., 2nd edition.

Greenleaf, E. and D. Lehmann (1995), "Reasons for Substantial Delay in Consumer Decision Making," *Journal of Consumer Research*, 22 (2), 186–199.

Harada, C., M. Natelson Love, and K. Triebel (2013), "Normal Cognitive Aging," *Clinics in Geriatric Medicine*, 29 (4), 737–752.

Hodgson, C. and G. Lewis (2021), "You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Consumer Search," Working Paper.

Honka, E. (2014), "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry," *RAND Journal of Economics*, 45 (4), 847–884.

Honka, E. and P. Chintagunta (2017), "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *Marketing Science*, 36 (1), 21–40.

Honka, E., A. Hortaçsu, and M. Vitorino (2017), "Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry," *RAND Journal of Economics*, 48 (3), 611–646.

Honka, E., A. Hortascu, and M. Wildenbeest (2019), "Empirical Search and Consideration Sets," in JP Dube and P Rossi (Editors), "Handbook on the Economics of Marketing," North Holland.

Hotz, J. and R. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60 (3), 497–529.

Kim, J., P. Albuquerque, and B. Bronnenberg (2010), "Online Demand under Limited Consumer Search," *Marketing Science*, 29 (6), 1001–1023.

— (2017), "The Probit Choice Model under Sequential Search with an Application to Online Retailing," *Management Science*, 63 (11), 3911–3929.

Kim, S., D. Goldstein, L. Hasher, and R. Zacks (2005), "Framing Effects in Younger and Older Adults," *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 60 (4), 215–218.

Koulayev, S. (2014), "Search for Differentiated Products: Identification and Estimation," *RAND Journal of Economics*, 45 (3), 553–575.

Lambert-Pandraud, R., G. Laurent, and E. Lapersonne (2005), "Repeat Purchasing of New Automobiles by Older Consumers: Empirical Evidence and Interpretations," *Journal of Marketing*, 69 (2), 97–103.

Levav, J., M. Heitmann, A. Herrmann, and S. Iyengar (2010), "Order in Product Customization Decisions: Evidence from Field Experiments," *Journal of Political Economy*, 118 (2), 274–299.

Magnac, T. and D. Thesmar (2002), "Identifying Dynamic Discrete Choice Processes," *Econometrica*, 70 (2), 801–816.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57 (5), 995–1026.

Moraga-González, J., Z. Sándor, and M. Wildenbeest (2018), "Consumer Search and Prices in the Automobile Market," Working Paper.

Morgan, P. and R. Manning (1985), "Optimal Search," *Econometrica*, 53 (4), 923–944.

Novemsky, N., R. Dhar, N. Schwarz, and I. Simonson (2007), "Preference Fluency in Choice," *Journal of Marketing Research*, 44 (3), 347–356.

Ozdenoren, E., S. Salant, and D. Silverman (2012), "Willpower and the Optimal Control of Visceral Urges," *Journal of the European Economic Association*, 10 (2), 342–368.

Peters, E. (2010), "Age Related Changes in Decision Making," in A. Drolet, N. Schwarz, and C Yoon (Editors), "The Aging Consumer: Perspectives from Psychology and Economics," Taylor and Francis, 75–101.

Petrikaite, V. (2018), "Consumer Obfuscation by a Multiproduct Firm," *RAND Journal of Economics*, 49 (1), 206–223.

Rust, J. (1994), "Structural Estimation of Markov Decision Processes," in R. Engle and D. McFadden (Editors), "Handbook of Econometrics," Amsterdam: North-Holland, volume 4, 3081–3143.

Schmittlein, D., D. Morrison, and R. Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33 (1), 1–24.

Seiler, S. (2013), "The Impact of Search Costs on Consumer Behavior: A Dynamic Approach," *Quantitative Marketing and Economics*, 11 (2), 155–203.

Stigler, G. (1961), "The Economics of Information," *Journal of Political Economy*, 69 (3), 213–225.

Stiglitz, J. (1987), "Competition and the Number of Firms in a Market: Are Duopolies More Competitive than Atomistic Markets?" *Journal of Political Economy*, 95 (5), 1041–1061.

Tanius, B., S. Wood, Y. Hanoch, and T. Rice (2009), "Aging and Choice: Applications to Medicare Part D," *Judgement and Decision Making*, 4 (1), 92–101.

Ursu, R. (2018), "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 37 (4), 530–552.

Ursu, R. and D. Dzyabura (2020), "Retailers' Product Location Problem with Consumer Search," *Quantitative Marketing and Economics*, 18 (3), 125–154.

Ursu, R., A. Simonov, and E. An (2021), "Online Advertising as Passive Search," Working paper.

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

Vohs, K., R. Baumeister, B. Schmeichel, J. Twenge, N. Nelson, and D. Tice (2008), "Making Choices Impairs Subsequent Self-Control: A Limited-Resource Account of Decision Making, Self- Regulation, and Active Initiative," *Journal of Personality and Social Psychology*, 94 (5), 883–898.

Weitzman, M. (1979), "Optimal Search for the Best Alternative," *Econometrica*, 47 (3), 641–654.

Whittle, P. (1980), "Multi-Armed Bandits and the Gittins Index," *Journal of the Royal Statistical Society Series B (Methodological)*, 42 (2), 143–149.

Wu, C., H. Che, T. Chan, and X. Lu (2015), "The Economic Value of Online Reviews," *Marketing Science*, 34 (5), 739–754.

Yao, S., C. Mela, J. Chiang, and Y. Chen (2012), "Determining Consumers' Discount Rates with Field Studies," *Journal of Marketing Research*, 49 (6), 822–841.

Yavorsky, D., E. Honka, and K. Chen (2021), "Consumer Search in the U.S. Auto Industry: The Role of Dealership Visits," *Quantitative Marketing and Economics*, 19 (1), 1–52.

Figures and Tables

Figure 1: Example of a Search Process with Definitions

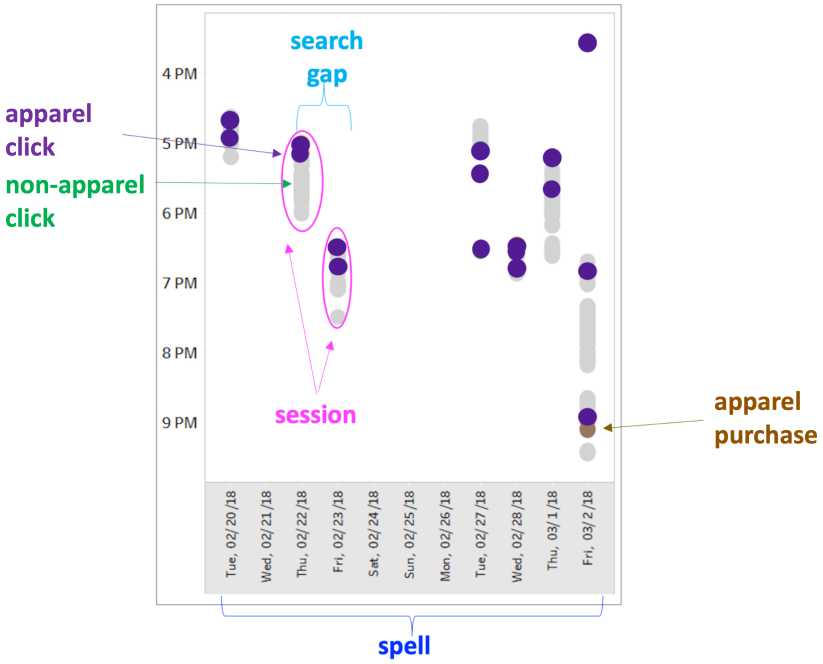
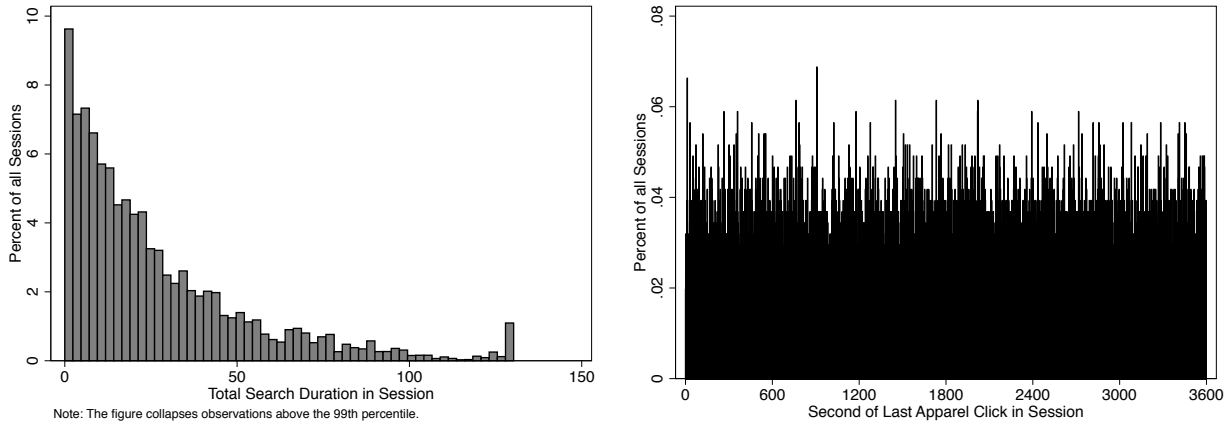
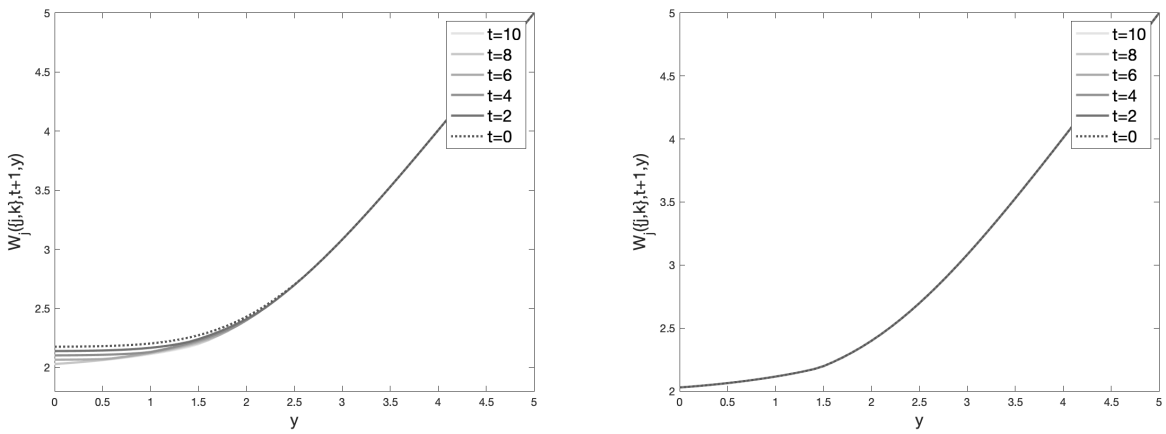


Figure 2: Duration and End of Apparel Search Sessions



(a) Apparel Search Session Duration in Minutes **(b) Second of Last Apparel Search Session Click**

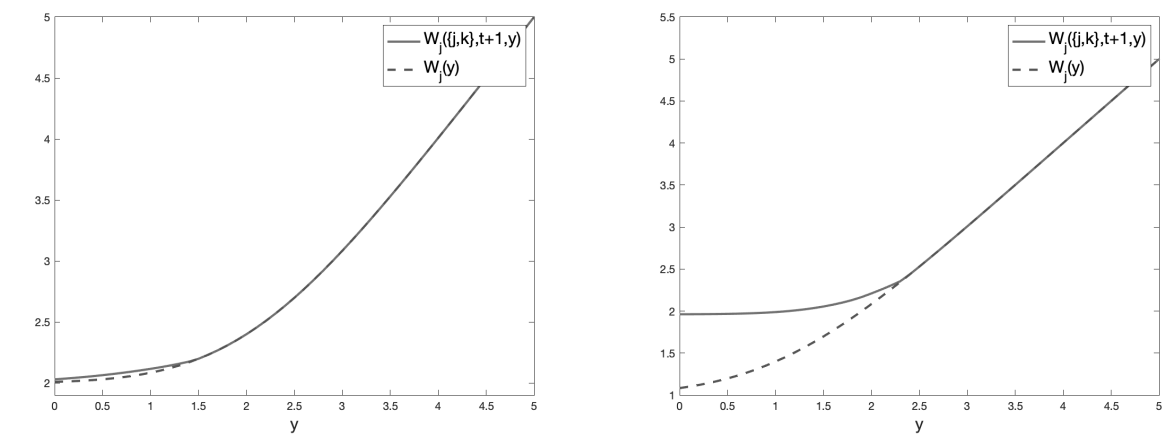
Figure 3: The Continuation Value $W_j(\{j,k\},t+1,y)$ for Different Levels of Fatigue



(a) Small Fatigue $\alpha = e^{-6}$
($\beta = 0.95$, $\mu_j = 2$, $\mu_k = 1$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-2}$)

(b) Large Fatigue $\alpha = e^{-1}$
($\beta = 0.95$, $\mu_j = 2$, $\mu_k = 1$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-2}$)

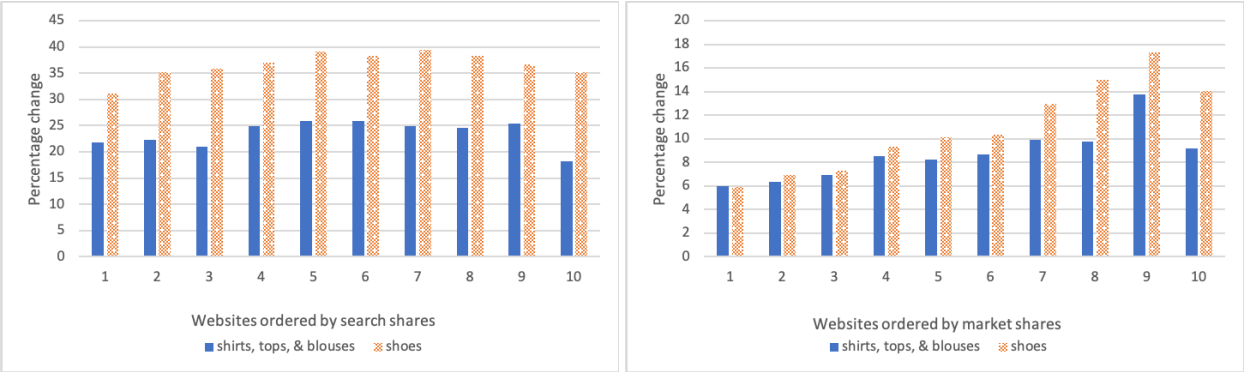
Figure 4: Comparing Continuation Values $W_j(\{j,k\},t+1,y)$ and $W_j(y)$



(a) $\mu_j = 2$ **and** $\mu_k = 1$
($\alpha = e^{-1}$, $t = 1$, $\beta = 0.95$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-2}$)

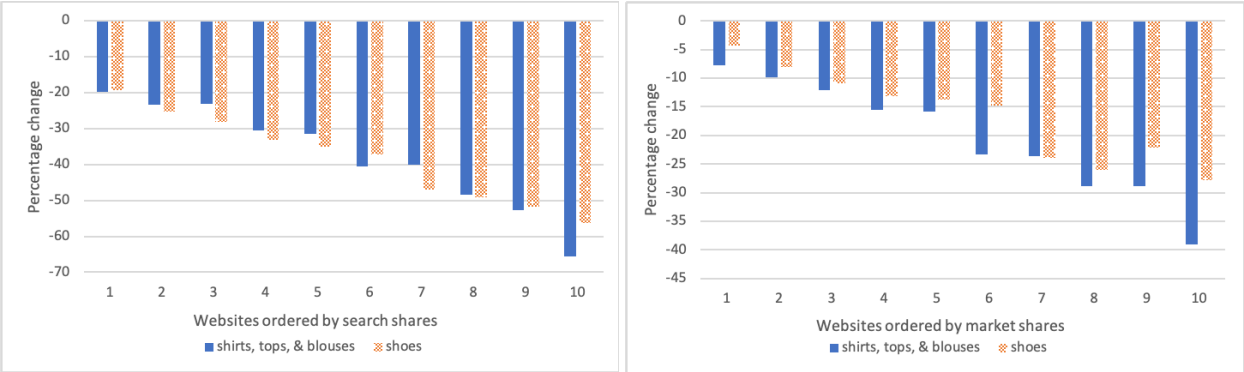
(b) $\mu_j = 1$ **and** $\mu_k = 2$
($\alpha = e^{-1}$, $t = 1$, $\beta = 0.95$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-2}$)

Figure 5: Effects of 90% Fatigue Reduction by Website



(a) Search shares **(b) Market shares**

Figure 6: Effects of No Breaks by Website



(a) Search shares **(b) Market shares**

Table 1: Session Characteristics

	Mean	Median	Std. Dev.	Min	Max
Session: All clicks					
Number of clicks	189.07	116.00	250.85	1	9,108
Number of visited websites	29.38	20.00	28.44	1	309
Duration (in minutes)	95.58	64.25	106.11	0	2,043
Session: Apparel clicks					
Number of clicks	10.52	3.00	26.34	1	2,407
Number of visited websites	1.73	1.00	1.44	1	29
Duration (in minutes)	5.59	1.03	12.21	0	508
Number of subcategories	1.14	1.00	1.29	0	9

Table 2: Most Popular Apparel Websites and Product Subcategories

	Popular Websites		Ranked Subcategories
Ordered by	*Search share*	*Market share*	*Market share*
	zalando.nl	zalando.nl	Shirts, tops, & blouses
	hm.com	hm.com	Shoes
	c-and-a.com	c-and-a.com	Pants & jeans
	debijenkorf.nl	your-look-for-less.nl	Underwear
	missetam.nl	esprit.nl	Dresses & skirts
	your-look-for-less.nl	debijenkorf.nl	Children's clothes
	vente-exclusive.com	missetam.nl	Jackets & vests
	esprit.nl	vente-exclusive.com	Accessories
	vanharen.nl	hunkemoller.nl	
	schuurman-shoenen.nl	vanharen.nl	

Table 3: Top 10 Websites for the "Shirts, Tops, & Blouses" and "Shoes" Subcategories

	"Shirts, tops, & blouses"		"Shoes"	
Ordered by	*Search share*	*Market share*	*Search share*	*Market share*
	c-and-a.com	zalando.nl	zalando.nl	zalando.nl
	debijenkorf.nl	hm.com	schuurman-shoenen.nl	vanharen.nl
	zalando.nl	your-look-for-less.nl	vanharen.nl	adidas.com
	hm.com	c-and-a.com	adidas.com	debijenkorf.nl
	aboutyou.com	esprit.nl	spartoo.nl	nelson.nl
	esprit.nl	peterhahn.nl	nike.com	nike.com
	your-look-for-less.nl	aboutyou.com	omoda.nl	omoda.nl
	msmode.nl	debijenkorf.nl	nelson.nl	schuurman-shoenen.nl
	peterhahn.nl	jbfo.nl	debijenkorf.nl	spartoo.nl
	jbfo.nl	msmode.nl	ziengs.nl	ziengs.nl

Table 4: Search Gaps within Apparel Subcategories

	Shirts, Tops, & Blouses	Shoes	Pants & Jeans	Under-wear	Sweaters	Dresses & Skirts	Children's Clothes	Jackets & Vests	Acces-sories
Proportion of spells with ≥ 1 search gap	0.50	0.56	0.44	0.36	0.37	0.39	0.45	0.38	0.41
Av. number of search gaps if ≥ 1 search gap	2.88	5.14	2.65	2.22	2.43	2.90	2.77	2.54	2.81
Av. length of search gaps (in days)	7.35	5.86	8.02	8.90	8.21	7.24	8.33	7.70	7.75
Med. length of search gaps (in days)	3.70	2.10	3.23	4.38	3.71	2.98	4.04	3.17	3.77
Av. length of spell (in days)	10.54	16.91	9.32	7.05	7.38	8.14	10.52	7.38	8.85
Av. time between spells (in days)	11.52	10.57	12.12	13.80	13.01	13.71	12.84	14.23	14.47

Table 5: Effects of Fatigue Proxies on Search Gaps

	Dependent variable:						
	Number of search gaps in a spell[a]			Search gap indicator			
	(i)	(ii)		(iii)	(iv)	(v)	(vi)
Subcategory	"Shirts, tops, & blouses"	"Shoes"		"Shirts, tops, & blouses"		"Shoes"	
Age	0.0035***	0.097***	Cumulative number	0.3213***		0.0832***	
	(0.0009)	(0.0011)	of searched websites	(0.0277)		(0.0117)	
Slower speed score[a]	0.0881***	0.0584***	Total time spent searching		0.0151***		0.0057***
	(0.0127)	(0.0132)			(0.0018)		(0.0011)
Readability (SMOG)[a]	0.0383	0.8334***					
	(0.0629)	(0.0692)					
Controls							
Gender indicator	Yes	Yes	Spell FEs	Yes	Yes	Yes	Yes
Number of images	Yes	Yes	Website FEs	Yes	Yes	Yes	Yes
Number of words	Yes	Yes	Session with				
Number of searches	Yes	Yes	Transactions Indicator	Yes	Yes	Yes	Yes
R^2	0.10	0.26		0.46	0.30	0.34	0.32
Number of Observations	2,315	2,435		7,102	7,102	15,554	15,554

Standard errors in parentheses.

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

[a] Operationalized on a logarithmic scale.

Table 6: Monte Carlo Simulation Results

	(i)	(ii)		(iii)	
		Our model		*Weitzman*	
	True values	Estimates	Std. Dev.	Estimates	Std. Dev.
Utility					
Website 1	-1.0	-0.91	(0.03)	-0.93	(0.02)
Website 2	-0.5	-0.48	(0.03)	-0.48	(0.03)
Website 3	-0.3	-0.30	(0.04)	-0.29	(0.02)
Outside option	0.5	0.54	(0.03)	0.23	(0.04)
Search cost (exp)					
Baseline	-3.5	-3.53	(0.12)	-2.18	(0.04)
Fatigue	-3.0	-2.67	(0.03)		
Log-likelihood		-19,299		-17,939	
Number of Observations		25,000		25,000	

Table 7: Estimation Results

	(i) "Shirts, tops, & blouses"				(ii) "Shoes"		
	Accounting for search gaps		Ignoring search gaps		Accounting for search gaps		Ignoring search gaps
	Our model	*Our model*	*Weitzman*		*Our model*	*Our model*	*Weitzman*
Utility				*Utility*			
aboutyou.com	-1.4368***	-1.3084***	-1.4061***	adidas.com	-1.3895***	-1.3979***	-1.2570***
	(0.0440)	(0.0337)	(0.0463)		(0.0507)	(0.0259)	(0.0423)
c-and-a.com	-0.8623***	-0.7662***	-0.8250***	debijenkorf.nl	-1.9646***	-2.1189***	-1.8366***
	(0.0372)	(0.0288)	(0.0397)		(0.0625)	(0.0467)	(0.0513)
debijenkorf.nl	-1.0226***	-1.3782***	-1.0022***	nelson.nl	-1.9438***	-1.8941***	-1.7978***
	(0.0375)	(0.0406)	(0.0408)		(0.0578)	(0.0416)	(0.0532)
esprit.nl	-1.7118***	-1.7419***	-1.6847***	nike.com	-1.5095***	-1.5162***	-1.3740***
	(0.0496)	(0.0459)	(0.0517)		(0.0540)	(0.0285)	(0.0433)
hm.com	-1.3601***	-1.3584***	-1.3119***	omoda.nl	-1.8658***	-1.8540***	-1.7174***
	(0.0412)	(0.0373)	(0.0444)		(0.0581)	(0.0373)	(0.0491)
jbfo.nl	-2.6664***	-2.4721***	-2.7581***	schuurman-shoenen.nl	-1.1034***	-1.0502***	-0.9748***
	(0.1448)	(0.1261)	(0.1455)		(0.0500)	(0.0211)	(0.0373)
msmode.nl	-2.0188***	-1.9307***	-2.0093***	spartoo.nl	-1.4858***	-1.4437***	-1.3663***
	(0.0653)	(0.0573)	(0.0638)		(0.0514)	(0.0269)	(0.0419)
peterhahn.nl	-2.1520***	-2.0599***	-2.1132***	vanharen.nl	-1.2167***	-1.2168***	-1.0641***
	(0.0816)	(0.0708)	(0.0776)		(0.0466)	(0.0242)	(0.0392)
your-look-for-less.nl	-1.7526***	-1.6579***	-1.7228***	zalando.nl	-0.8434***	-0.9784***	-0.6502***
	(0.0501)	(0.0416)	(0.0512)		(0.0442)	(0.0218)	(0.0344)
zalando.nl	-1.0483***	-1.1357***	-0.9856***	ziengs.nl	-2.0965 ***	-2.1204***	-1.9532***
	(0.0381)	(0.0325)	(0.0410)		(0.0722)	(0.0546)	(0.0578)
Number of previous		0.1407***		Number of previous		0.1849 ***	
website visits		(0.0109)		website visits		(0.0168)	
Visit to a		1.3058***		Visit to a		0.9307***	
price discount page		(0.0411)		price discount page		(0.0442)	
Outside option	1.5823***	1.4050***	1.9583***	Outside option	1.3288***	1.2714***	2.1653***
	(0.0304)	(0.0194)	(0.0512)		(0.0334)	(0.0185)	(0.0558)
Search cost (exp)				*Search cost (exp)*			
Baseline	-5.0740***	-4.8991***	-3.1639***	Baseline	-5.5267***	-5.6355***	-4.2142***
	(0.3024)	(0.2507)	(0.1066)		(0.2950)	(0.2290)	(0.1354)
Fatigue constant	-2.4436***	-2.4793***		Fatigue constant	-2.6379***	-2.6422***	
	(0.0243)	(0.0164)			(0.0395)	(0.0207)	
Fatigue age (≥ 50)		0.3337***		Fatigue age (≥ 50)		0.1131***	
		(0.0268)				(0.0265)	
Number of Observations	27,924	27,600	27,924		27,756	27,264	27,756
LL	-9,356	-8,643	-8,642		-12,179	-11,671	-10,985

Standard errors in parentheses.

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

Table 8: Model Fit Comparison

	(i) "Shirts, tops, & blouses"			(ii) "Shoes"		
	Data	*Our model*	*Weitzman*	*Data*	*Our model*	*Weitzman*
RMSE						
Number of search gaps		242	763		44	1,368
Website search shares		39	52		76	80
Website market shares		71	45		105	49
Data and Predictions						
Number of search gaps	763	1005	N/A	1,368	1,412	N/A
Number of searches	3,334	3,346	3,049	4,254	3,784	3,725
Number of purchases	309	717	529	248	873	486

Table 9: Effects of Fatigue and Search Cost Reductions

	(i)	(ii)	(iii)	(iv)
	Fatigue reduction by		Search cost reduction by	
	50%	90%	50%	90%
"Shirts, tops, & blouses"				
Percent change				
Number of searches	1.33	18.54	2.02	3.66
Number of purchases	0.53	4.51	0.37	0.65
Number of search gaps	-11.53	-64.37	6.62	11.76
"Shoes"				
Percent change				
Number of searches	4.12	28.18	1.32	1.98
Number of purchases	1.20	4.99	0.18	0.28
Number of search gaps	-22.46	-66.97	3.44	5.13

Web Appendix A: Construction of the Final Data Sample

The raw data contain information on the user (including demographics such as age and gender), session, and time of the click, as well as the website name and the entire URL address of the visited website. Furthermore, GfK coded the transaction funnel, identifying a website visit, a product view, a basket addition, a checkout, or an order confirmation.

Data Augmentation: Scraping

Using the full URLs, we attempted to scrape the top 50 apparel websites (ranked by the number of clicks) and were successful in scraping 44 of them. We were able to scrape information from the most popular websites, such as Zalando and H&M, obtaining product information from the top websites accounting for more than 57% of all apparel clicks. This data collection stage occurred within one month of the last observation day in our sample to prevent changes in the web pages. The information we gathered by scraping contains: price, price promotion (if any), page title, brand name, product name, and, if available, product color, reviews, star rating, number of photos, product description, shipping information, speed score of the website (page loading speed), as well as word counts, sentiment on the page, and reading ease.

Data Classification: List, Product, and Other Pages

Although GfK coded the transaction funnel, we performed an additional step in classifying clicks into list, product, or other pages, to ensure that we correctly identify the product (if any) that was purchased. An example of a list page is "https://www.adidas.com/us/women-originals-shoes", where consumers can see a list of shoes along with a photo, the product name and its price. If a consumer clicks on a product in this list, she navigates to that product's page. An example of such a product page is "https://www.adidas.com/us/adidas-sleek-shoes/EE4723.html" for a consumer who clicked on the product "adidas sleek shoes" on the list page. The product page contains more detailed information about the product, such as a product description, additional photos, reviews, etc. Of the 428,651 apparel clicks in our data, 172,536 clicks are on list pages and 93,463 clicks are on product pages. We labeled the remaining clicks as "others" to represent clicks to the homepage of a website, account pages, or any transaction-related pages, such as the cart page.

To categorize clicks into either product, list, or other pages, we performed the following steps.

First, during data scraping, we identified list and product pages by examining whether there was any product information available on the page and, if so, how many products were available on the page. Second, we used the following rules:

- 'Other' pages:
  1. Pages labeled as 'Add to Basket', 'Start Checkout' and 'Order Confirmation' by GfK
  2. The homepage of a website such as 'www.zara.com'
- 'Product' page:
  1. Pages labeled as 'Product View' by GfK
  2. Pages from which we can scrape a single product's information
  3. URLs that contains product SKU or product IDs (rules differ for each website)
  4. URLs with specific keywords such as 'product-view' or 'shop-by-item'
- 'List' page:
  1. Pages from which we can scrape information on multiple products
  2. Pages that display search results, e.g., 'https://www.adidas.com/us/search?q=redshoes'
  3. URLs with specific keywords that indicate their function as a list page, such as 'shop-per-categorie', 'page=', or 'category='
  4. URLs with specific keywords that indicate the sorting or filtering function available on the page, such as 'price_max=', 'productsoort', 'pagenumber' or 'filter='

We further categorized URLs labeled as 'others' by manually checking them and hired an RA to independently manually check our categorization.

Data Classification: Product Categories

To identify the searched product category, we used URLs, page titles, and the scraped information (e.g., product description) to search for keywords identifying nine broad categories (as defined on the most popular website in our data, Zalando): accessories, children's, dresses & skirts, jackets & vests, pants & jeans, shirts & tops and blouses, shoes, sweater and underwear. The keywords used to identify the product categories include but are not limited to:

- **accessories**: 'accesso', 'sjaal', 'lippen', 'earr', 'necklace', 'jewelry', 'bracelet', 'bag', 'eastpak', 'hals', 'banden'
  - with exception of: 'brand', 'bracelet', 'braad', 'brax', 'dirk', 'brace', 'overnachtingen', 'aangebrachte'
- **children's**: 'jongens', 'kinder', 'meisjes', 'baby', 'tiener', 'kids', 'boys' 'girls'
- **dresses & skirts**: 'roecke', 'jurken', 'dress', 'jumpsuit', 'jurkje', 'jurk', 'rok'

- **jackets & vests**: 'trench', 'jack', 'fleece', 'blazer', 'mantel', 'coat', 'parka', 'tussenjas', 'winterjas', 'jas'
- **pants & jeans**: 'hosen', 'broek', 'jogger', 'tights', 'shorts', 'sweatpant', 'pants', 'pantalon', 'leggin', 'trouser', 'tregging', 'jegging'
  - with exception of: 'brand', 'bracelet', 'braad', 'brax', 'dirk', 'brace', 'overnachtingen', 'aangebrachte'
- **shirts, tops & blouses**: 'top', 'hemden', 'langarm', 'kurzarm', 'blusen', 'shirt', 'singlet', 'blouse', 'blouson', 'polo-', '-polo', 'polos', 'longsleeve', 'overhemd', 'onderhemd'
  - with exception of: 'topseller', 'topic', 'topbox', 'topgear', 'topdeals', 'topper', 'topman', 'sniztop', 'marc-c-polo', 'topcom', 'topbloemen', 'laptop'
- **shoes**: 'schuhe', 'stiefel', 'schoen', 'shoe', 'sneaker', 'sandal', 'birkenstock', 'fitflop', 'teva', 'footwear', 'e-walk', 'ecco', 'gabor', 'instappe', 'pumps'
  - with exception of domain names that contain the word 'shoe'
- **sweater**: 'parka', 'hoodie', 'poncho', 'westen', 'trui', 'capuchon', 'pullover', 'tuniek', 'vest', 'cardigan', 'sweater', 'jumper'
- **underwear**: 'thong', 'nightwear', 'bra', 'lingerie', 'sleep', 'swim', 'badpak', 'ondergoed', 'underwear', 'panties', 'sock', 'sok', 'bustier', 'push-up', 'boxer', 'badmode', 'bikini', 'tanga', 'tankini'

Data Classification: Activities

GfK classifies clicks into activities such as 'apparel,' 'social networking,' or 'web search.' We used this classification as well as the following rules to further identify the type of online activity the consumer engaged in. This process resulted in ten categories.

1. Apparel:
   - GfK's classification as 'Fashion'
2. Search engine:
   - GfK's classification as 'Web Search'
   - URLs that contain the keyword 'search' when the website visited is Google, Yahoo or Bing.
   - URLs where the website is 'ask.com'
3. Email
   - GfK's classification as 'Communication'
   - URLs that contain keywords 'mail.google', 'outlook', and 'webmail'.
   - URLs that contain the keyword 'mail' when the visited website is Google, Yahoo or Bing
   - URLs that contain the keyword 'messenger' when the visited website is Yahoo
4. Social Networking
   - GfK's classification as 'Social Networking'

- The visited website is one of the 5 major social media platforms: facebook, pinterest, twitter, instagram, linkedin

5. Banking
   - GfK's classification as 'Money Management'
   - The visited website is or contains 'rabobank.nl', 'abnamro.nl', 'bank', 'achmea' or 'vanlanschot'
6. Cashback
   - The visited website is one of: 'geldrace.nl', 'geldkoffer.info', 'geldwolf.info', 'zinngeld.nl', 'mailbeurs.nl', 'extraeuro.nl', 'centmail.nl', 'cashhier.nl', 'spaar4cash.nl', 'snelverdienen.nl', 'ipay.nl', 'spaaractief.nl', 'nucash.nl', 'myflavours.nl', 'directverdiend.nl', 'dieselmail.nl', 'spaar4cash.nl', 'dutcheuro.nl', 'extraeuro.nl', 'cashparadijs.nl', 'sneleuro.nl', 'myclics.nl', 'spaar-voor-euries.nl', 'jiggy.nl', 'qlics.nl', 'quidco'
7. Surveys
   - The visited website is one of: 'gfk.com', 'ssisurveys.com', 'focusvision.com', 'opinionbar.com', 'globaltestmarket.com'
8. Media
   - GfK's classification as 'Media Broadcasting' or 'Media On-Demand'
   - URLs that contain the keyword 'tvgids'
9. Google exclude
   - URLs from 'google.com' that are not classified as search engine or email related (this includes Google Drive, Maps, etc.)
10. Gaming
    - GfK's classification as 'Gaming'
    - URLs containing the keywords: 'casino', 'game', 'unibet', 'nederlandseloterij.nl'

Data Cleaning: Removing Non-search Activity

The raw data contains 9,531,448 observations. To obtain the final data set, we removed observations in the following cases:

- Consumers use a web browser to open local files on their computers rather than browse the Internet
- A new tab is opened but no webpage is visited on that tab
- Consumers open web browsers' extensions
- Any URL that does not contain 'ttp'
- Duplicates at the session-time level: the same URL is clicked more than once at the same time or two different URLs are clicked at the same time. In both cases, we only kept a record of the first click

- Spells during which sessions overlap in time (one instance)
- Spells with a transaction but no clicks on product pages observed (in these rare cases, websites likely offer the option of adding a product to the cart directly from the list page)
- Spells with a transaction and observed product clicks but no product added to the cart
- Spells that end within the first week of our observation period, i.e., before February 23rd, 2018, since it is likely that these observations are left truncated

These changes resulted in a data set with 7,877,551 observations. In addition, among the apparel clicks (437,659 observations), we dropped sessions and their corresponding spells that only contained clicks that were unrelated to product search activity, such as clicks to log in or out of an account, to track a shipment, to find a store location, to access customer service, to manage a subscription or to create a password. Of the original 437,659 observations, we were left with 428,651 observations.

Estimation Samples

We constructed our estimation samples as follows. We focused on the two most commonly purchased product subcategories in our data: "shirts, tops, & blouses" and "shoes." We removed search revisits (i.e., revisits to the same website beyond the first visit that are unrelated to the actions required to make a purchase, such as logging into an account) from the sample (approximately 30% of observations). To address concerns about right truncation, we removed spells that did not end in a transaction and that had a search session within the last two days of our observation period.[49] We focused on spells with at most one transaction (more than 99.3% of spells) and we removed spells with clicks that might have occurred after the consumer saw an ad on social media, email, newsletter or through retargeting (fewer than 20% of spells).[50] For each subcategory, we determined the top 10 most searched websites (accounting for approximately 65% of clicks in each subcategory), for which we estimate website intercepts. All other websites were grouped together into a composite website which we call "Other."

The resulting estimation samples have 27,924 observations and 2,327 spells in the "shirts, tops,

[49] Recall that we addressed left truncation concerns by removing spells that ended within the first week of our observation period. To further address potential concerns about left and right truncation, we also conducted a robustness check in which we dropped spells (i) that contained searches within the first week of our observation period and (ii) that did not end in a transaction but had a search session within the last week of our observation period. The estimation results are shown in Table F-4 in Web Appendix F and are very similar to those from our main specification (see Table 7), with slightly higher search costs since removing spells performed in two of the 10 weeks in our data selects consumers with fewer searches. Note that in the first apparel subcategory, two of the top 10 websites are different than in our main sample, given the significant drop in search spells.

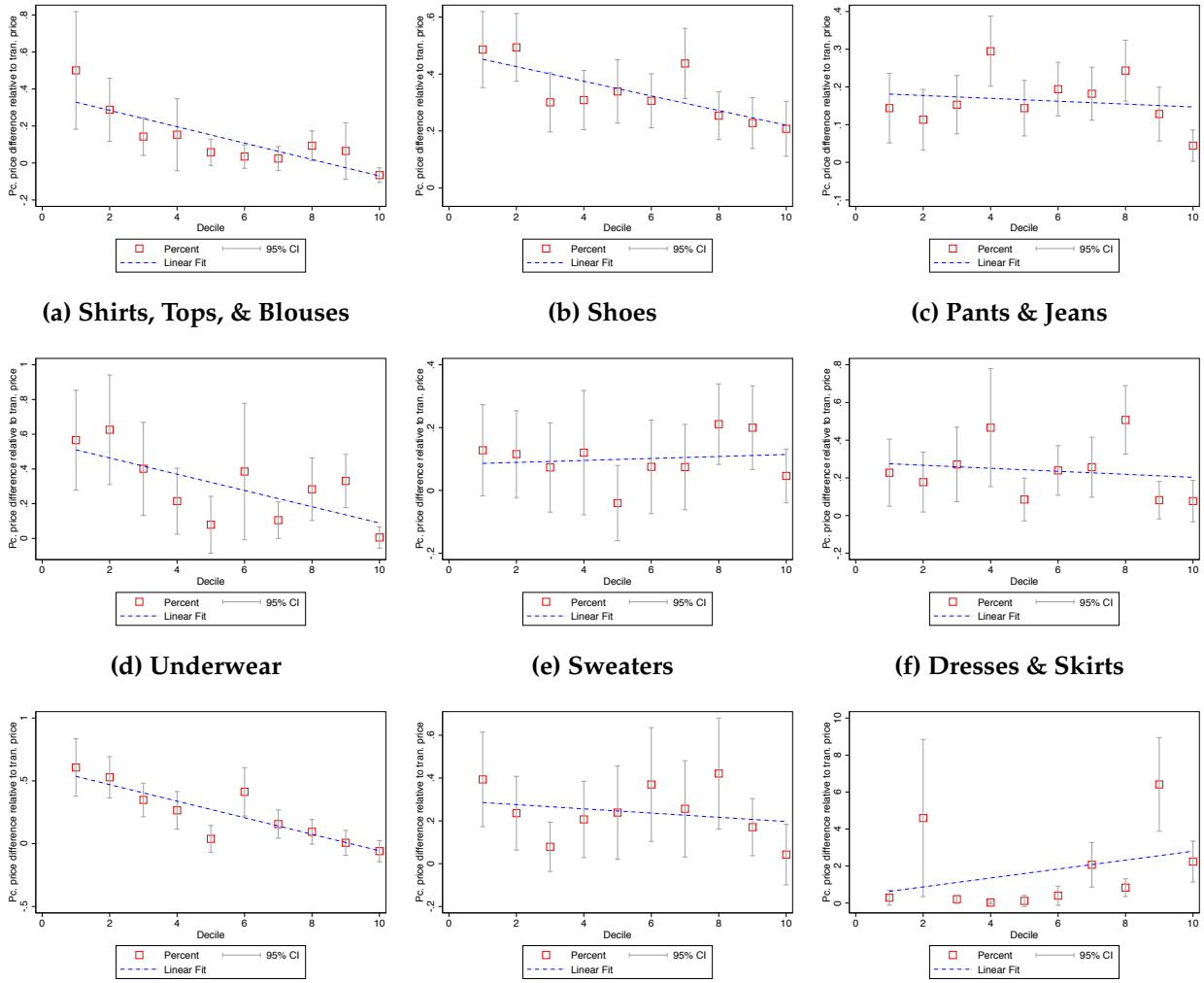
[50] Note that ads may encourage consumer to start a new search session, but cannot explain why consumers stopped their search in a previous session, and therefore cannot explain search gaps. More robustness checks estimating our model on spells without any ad clicks can be found in Web Appendix F.

& blouses" subcategory, and 27,756 observations and 2,313 spells in the "shoes" subcategory.[51] Consumers made 309 and 248 transactions in each subcategory, respectively. Further, there are 763 and 1,368 search gaps in each subcategory, respectively, with 586 and 818 spells with at least one search gap.

[51] To ensure that we capture all searches performed in a subcategory, we attribute searches that do not have a subcategory labeling but that occur in between searches labeled as a given subcategory, to that subcategory.

Web Appendix B: Supporting Data Patterns

Figure B-1: Price Convergence across Apparel Subcategories



(a) Shirts, Tops, & Blouses **(b) Shoes** **(c) Pants & Jeans**

(d) Underwear **(e) Sweaters** **(f) Dresses & Skirts**

(g) Children's Clothes **(h) Jackets & Vests** **(i) Accessories**

Table B-1: Revisit Patterns

	Shirts, Tops, & Blouses	Shoes	Pants & Jeans	Under-wear	Sweaters	Dresses & Skirts	Children's Clothes	Jackets & Vests	Acces-sories
Spells with No Revisits	0.67	0.57	0.69	0.77	0.75	0.75	0.69	0.75	0.73
Spells with Revisits & No Purchase	0.28	0.38	0.28	0.20	0.22	0.23	0.28	0.23	0.26
Spells with Revisits & Purchases	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.02	0.01
Conditional on Purchasing									
% of Spells with Revisits	0.34	0.39	0.29	0.23	0.34	0.25	0.42	0.24	0.34

Notes: Percentages for the first three rows in each column may not add up to 1 due to rounding.

Web Appendix C: Formal Details on Model Solution

C.1. Supplementary Analyses for Section 5.3

We have two goals in this section: (i) to provide a more formal theoretical treatment of the analysis performed in Section 5.3 for the specific empirical application of our model; and (ii) to perform more simulations to investigate the generality of the results in Section 5.3. More specifically, we show that, for a large range of parameter values, our results in Section 5.3 related to Condition 1 and the continuation value are robust.

C.1.1. Formal Analysis for Empirical Application

In Section 5.3, we showed that, for the general case when $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$, as y grows large, $W_j(y)$ approaches y, while for small values of y, $W_j(y)$ approaches $E_{F_j}(u)$. In this subsection, we aim to make this analysis more concrete by imposing the distributional constraints of our empirical application and thereby showing additional results. In our empirical application, we assume utilities are normally distributed with $N(\mu_j, \sigma_j^2)$. In this case, the continuation value $W_j(y)$ can be expressed as

$$W_j(y) = \mu_j\sigma_j + (y - \mu_j\sigma_j)\Phi(m_j) + \sigma_j^2\phi(m_j), \tag{C1}$$

where $m_j = \frac{y-\mu_j}{\sigma_j}$, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and the pdf of the standard normal distribution. In what follows, we assume $\sigma_j = 1$ (as in our empirical application).

So far, we have been able to characterize the behavior of $W_j(y)$ for large and small values of y. Here, we will make the statements "large" and "small" more precise. To achieve this goal, we first ask: for what values of y is $W_j(y) = \mu_j$? It is obvious that $m_j\Phi(m_j) + \phi(m_j) = 0$ only if both $\Phi(\cdot) = 0$ and $\phi(\cdot) = 0$. Noticing that, e.g., $-2\Phi(-2) + \phi(-2) = 0.0085$ and $-3\Phi(-3) + \phi(-3) = 3.8215e-04$, we find that $W_j(y) \approx \mu_j$ for $y - \mu_j < -3$. Second, we ask: for what values of y is $W_j(y) = y$? This is equivalent to asking for what values of m_j is $m_j\Phi(m_j) + \phi(m_j) - m_j = 0$. Once again, we notice that $3\Phi(3) + \phi(3) - 3 = 3.8215e-04$, suggesting that $W_j(y) \approx y$ for $y - \mu_j > 3$. Finally, we rely on the function $W_j(y) = \mu_j + (y - \mu_j)\Phi(m_j) + \phi(m_j)$ for $y - \mu_j \in (-3, 3)$.

Given these results, we now return to describing the continuation value $W_j(\bar{S}, t+1, y)$. From

equation (9), we know that

$$
\begin{aligned}
W_j(\{j,k\},t+1,y) &= \max\{y,\ -c_{k0}-\alpha(t+1)+W_k(y),\ \beta[-c_{k0}+W_k(y)]\}F_j(y) \\
&+ \max\left\{\int_y^\infty u dF_j(u),\ \int_y^\infty [-c_{k0}-\alpha(t+1)+W_k(u)]dF_j(u),\ \int_y^\infty \{\beta[-c_{k0}+W_k(u)]\}dF_j(u)\right\} \\
&\leq \max\left\{yF_j(y)+\int_y^\infty u dF_j(u),\ -c_{k0}-\alpha(t+1)+M_j(y),\ \beta\Big[-c_{k0}+M_j(y)\Big]\right\} \\
&= \max\left\{W_j(y),\ -c_{k0}-\alpha(t+1)+M_j(y),\ \beta\Big[-c_{k0}+M_j(y)\Big]\right\}
\end{aligned}
\tag{C2}
$$

where $M_j(y) = W_k(y)F_j(y) + \int_y^\infty W_k(u)\,dF_j(u)$. The inequality follows from the fact that $\max\{a,b,c\} + \max\{d,e,f\} \geq \max\{a+d, b+e, c+f\}$ for any values (a,b,c,d,e,f). Using the expression for $W_j(y)$ from equation (C1) above, we can further simplify $W_j(\{j,k\},t+1,y)$. Given the complexity of the expression for $W_j(\{j,k\},t+1,y)$, we next turn to additional simulation results to describe its relation to $W_j(y)$.

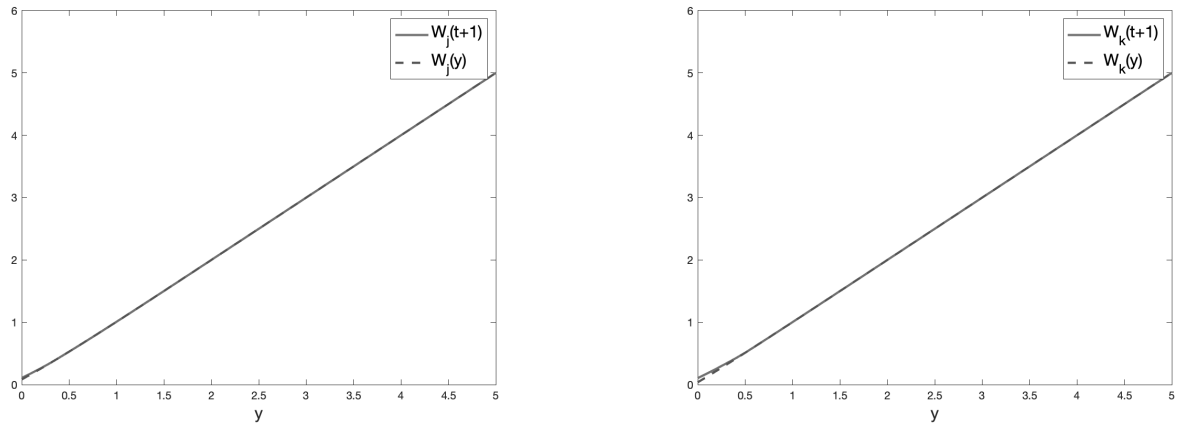
C.1.2. Simulations for Empirical Application – Relating $W_j(\{j,k\},t+1,y)$ and $W_j(y)$

In Section 5.3, we showed the relation between $W_j(\{j,k\},t+1,y)$ and $W_j(y)$ for two cases: (ii) $\mu_j = 2$, $\mu_k = 1$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-2}$, $\alpha = e^{-1}$, $\beta = 0.95$, and $t = 1$; and (ii) $\mu_j = 1$, $\mu_k = 2$, and all other parameters equal to the ones in case (i). Here, we describe the relation between $W_j(\{j,k\},t+1,y)$ and $W_j(y)$ for the parameter values we obtain in the estimation (see Section 7). For simplicity, we consider results from the case in which only website intercepts affect consumer utility (first column for each subcategory in Table 7). All website intercepts across the two apparel subcategories lie in the range $(-0.86, -2.66)$, but most lie in the range $(-1.0, -1.4)$. In our simulation, we compare $W_j(\{j,k\},t+1,y)$ and $W_j(y)$ for two sets of mean utilities for two websites, j and k: (i) common values $(\mu_j = -1.0, \mu_k = -1.4)$ and (ii) extreme values $(\mu_j = -0.86, \mu_k = -2.66)$. In the estimation, utility error terms are standard normally distributed, so we set $\sigma_j = \sigma_k = 1$ in our simulation. Baseline search costs are approximately $c_0 = e^{-5}$ and fatigue is $\alpha = e^{-2.5}$. The utility of the best option observed through search y must exceed the value of the outside option, which, consistent with the literature, we assume is at least zero (Weitzman, 1979; Kim, Albuquerque, and Bronnenberg, 2010). Finally, as in Section 5.3, we set $\beta = 0.95$ and $t = 1$ (to obtain the largest possible difference between $W_j(\{j,k\},t+1,y)$ and $W_j(y)$, since for larger values of t our results will continue to hold).

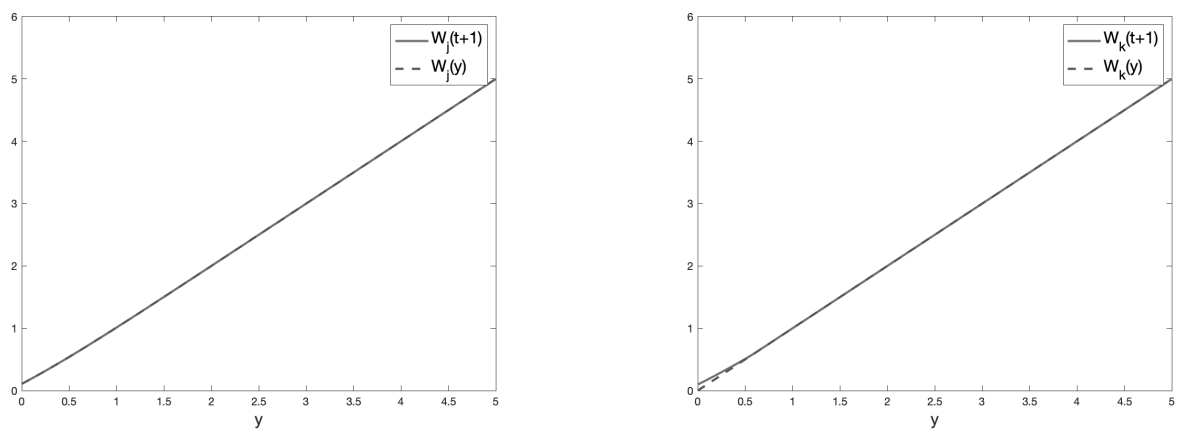
The results are shown in Figure C-1. In Figure C-1(a), we present results for the case of common values $(\mu_j = -1, \mu_k = -1.4)$. The graph on the left presents the difference between $W_j(\{j,k\},t+1,y)$

and $W_j(y)$ for different values of y, while the graph on the right shows the difference between $W_k(\{j,k\}, t+1, y)$ and $W_k(y)$. In both cases, the difference is relatively small, allowing us to conclude that the continuation value in our problem is very close to the one in the Weitzman (1979) model for most of the relevant parameter values in our estimation. Similar conclusions are derived from Figure C-1(b).

Figure C-1: Comparing Continuation Values for Parameter Estimates Obtained in Empirical Application



(a) Case (i): Common values obtained in estimation
($\alpha = e^{-2.5}$, $\beta = 0.95$, $\mu_j = -1$, $\mu_k = -1.4$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-5}$, $t = 1$)

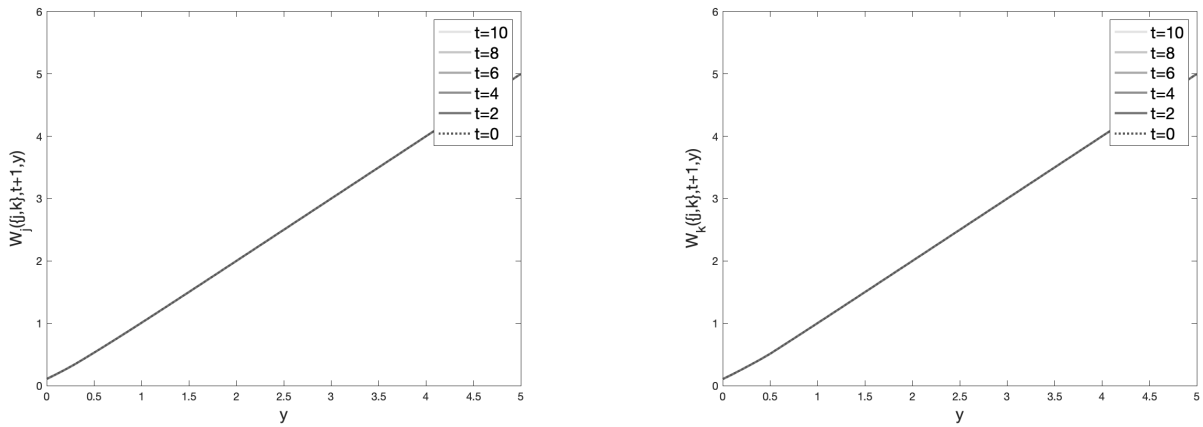


(b) Case (ii): Extreme values obtained in estimation
($\alpha = e^{-2.5}$, $\beta = 0.95$, $\mu_j = -0.86$, $\mu_k = -2.66$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-5}$, $t = 1$)

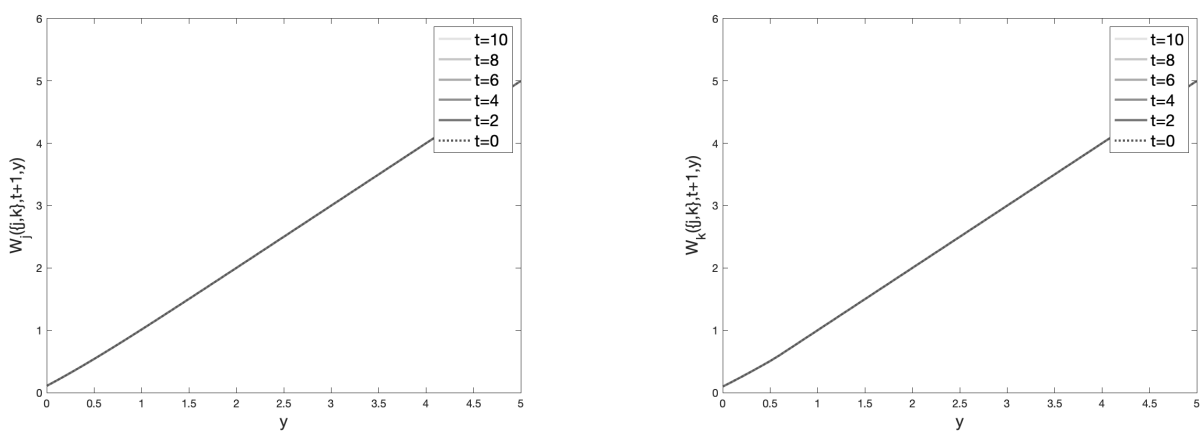
C.1.3. Simulations for Empirical Application – Relating $W_j(\{j,k\},t+1,y)$ and $W_j(\{j,k\},1,y)$

We follow the same steps outlined above to perform another simulation illustrating $W_j(\{j,k\},t+1,y)$ using the parameter values we obtain when estimating our model (see Section 7) and the functional form for $W_j(\{j,k\},t+1,y)$ we derived in equation 9. Our results in Figure C-2 below confirm that $W_j(\{j,k\},t+1,y)$ is independent of t for the parameter estimates we obtain in our empirical application.

Figure C-2: The Continuation Value $W_j(\{j,k\},t+1,y)$ for Parameter Estimates Obtained in Empirical Application



(a) Case (i): Common values obtained in estimation
($\alpha = e^{-2.5}$, $\beta = 0.95$, $\mu_j = -1$, $\mu_k = -1.4$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-5}$)



(b) Case (ii): Extreme values obtained in estimation
($\alpha = e^{-2.5}$, $\beta = 0.95$, $\mu_j = -0.86$, $\mu_k = -2.66$, $\sigma_j = \sigma_k = 1$, $c_{j0} = c_{k0} = e^{-5}$)

C.1.4. Evaluating Condition 1 and Assumption 1 for a Wide Range of Parameter Values

The previous two subsections in this web appendix and Section 5.3 in the paper provided evidence that the continuation values in our problem coincide or are very similar to the ones in the Weitzman model for parameter estimates close to those we obtain in our empirical application. These results were used to support the notion that Condition 1 and Assumption 1 are satisfied in our empirical application. Here, we evaluate how often Condition 1 and Assumption 1 are satisfied for parameter values far beyond those we obtain with our data.

Condition 1: We evaluate the relation that formally defines Condition 1, $W_j(\{j,k\},t+1,y) - W_k(\{j,k\},t+1,y) \geq W_j(\{j,k\},1,y) - W_k(\{j,k\},1,y)$, for any two options j and k and for a large set of parameter values. To do so, we write it as a difference, i.e., $\left[W_j(\{j,k\},t+1,y) - W_k(\{j,k\},t+1,y)\right] - \left[W_j(\{j,k\},1,y) - W_k(\{j,k\},1,y)\right]$, and check when this difference is non-negative. We use the functional form for $W_j(\{j,k\},t+1,y)$ we derived in equation 9 to precisely compute this difference and thus determine when Condition 1 is satisfied. Also, we maintain the same normality assumptions on the utility distribution as prior work, $N\left(\mu_j,\sigma_j^2\right)$ and $N\left(\mu_k,\sigma_k^2\right)$ (Kim, Albuquerque, and Bronnenberg, 2010, 2017; Honka and Chintagunta, 2017; Chen and Yao, 2017; Ursu, 2018).

We implement three sets of simulations. In the first set, we consider a large range of values with larger increments. More precisely, set 1 is described as follows:

- Set 1: "Vary All Values": 1,056,000 parameter combinations
  - Vary mean utility parameters, μ_j and μ_k, in the range $[-10,10]$ in increments of 5, resulting in the values $(-10,-5,0,5,10)$
  - Vary search cost parameters, $c_j = exp(\kappa_j)$ and $c_k = exp(\kappa_k)$, for values of κ_j (κ_k) in the range $[-6,0]$ in increments of 2, resulting in the values $(-6,-4,-2,0)$
  - Vary fatigue parameter, $\alpha = exp(a)$ for values of a in the range $[-6,0]$ in increments of 2, resulting in the values $(-6,-4,-2,0)$
  - Vary the uncertainty parameter, σ, in the range $[0.5,1.5]$ in increments of 0.5, resulting in the values $(0.5,1,1.5)$
  - Vary the discount factor, β, in the range $[0.8,1]$ in increments of 0.05, resulting in the values $(0.8,0.85,0.9,0.95,1)$
  - Vary t, in the range $[1,10]$ in increments of 3, resulting in the values $(1,4,7,10)$
  - Vary the utility of the best option observed though search, y, in the range $[0,10]$ in increments of 1, resulting in the values $(0,1,2,3,4,5,6,7,8,9,10)$

The next two sets consider a narrower set of parameter values, but with smaller increments, each focusing on a subset of the cases from set 1. In set 2, we vary utility and uncertainty parameters, as well as the discount factor. In set 3, we display results for parameter values close to the values estimated in our empirical application.

- Set 2: "Vary Utility, Uncertainty, and Discount Factor": 53,900 parameter combinations
  - Vary mean utility parameters, μ_j and μ_k, in the range $[-3,3]$ in increments of 1, resulting in the values $(-3,-2,-1,0,1,2,3)$
  - Vary the uncertainty parameter, σ, in the range $[0.5,1.5]$ in increments of 0.25, resulting in the values $(0.5,0.75,1,1.25,1.5)$
  - Vary the discount factor, β, in the range $[0.8,1]$ in increments of 0.05, resulting in the values $(0.8,0.85,0.9,0.95,1)$
  - Vary t, in the range $[1,10]$ in increments of 3, resulting in the values $(1,4,7,10)$
  - Vary the utility of the best option observed though search, y, in the range $[0,10]$ in increments of 1, resulting in the values $(0,1,2,3,4,5,6,7,8,9,10)$
  - Set $c_j = exp(-5)$, $c_k = exp(-5)$, and $\alpha = exp(-2.5)$
- Set 3: "Values Close to Estimates": 550 parameter combinations
  - Vary the discount factor, β, in the range $[0.8,1]$ in increments of 0.05, resulting in the values $(0.8,0.85,0.9,0.95,1)$
  - Vary t, in the range $[1,10]$ in increments of 1, resulting in the values $(1,2,3,4,5,6,7,8,9,10)$
  - Vary the utility of the best option observed though search, y, in the range $[0,10]$ in increments of 1, resulting in the values $(0,1,2,3,4,5,6,7,8,9,10)$
  - Set $\mu_j = -1$, $\mu_k = -4$, $c_j = exp(-5)$, $c_k = exp(-5)$, $\alpha = exp(-2.5)$, and $\sigma = 1$

We display our results in two different ways. First, we summarize our results for the three sets of simulations in Table C-1. We find that Condition 1 is satisfied in the vast majority of simulations and for wide as well as narrow ranges of the parameter space. For example, in set 1, 86% of the more than 1,000,000 parameter combinations indicate that Condition 1 is satisfied. Further, Condition 1 is satisfied for 92% of the parameter combinations in set 2 and 100% of those in set 3. Even when Condition 1 is not satisfied, the difference $\left[W_j(\bar{S},t+1,y) - W_k(\bar{S},t+1,y)\right] - \left[W_j(\bar{S},1,y) - W_k(\bar{S},1,y)\right]$ is mostly a very small negative number.[52] Allowing for very small violations of up to 0.1, Condition 1 holds in more than 93% of simulations in set 1,[53] supporting the notion that the value $\left[W_j(\bar{S},t+1,y) - W_k(\bar{S},t+1,y)\right]$ is a good approximation for the value $\left[W_j(\bar{S},1,y) - W_k(\bar{S},1,y)\right]$.

[52] Note that the calculated values of the difference $\left[W_j(\bar{S},t+1,y) - W_k(\bar{S},t+1,y)\right] - \left[W_j(\bar{S},1,y) - W_k(\bar{S},1,y)\right]$ in the simulations range from -9 to 9.

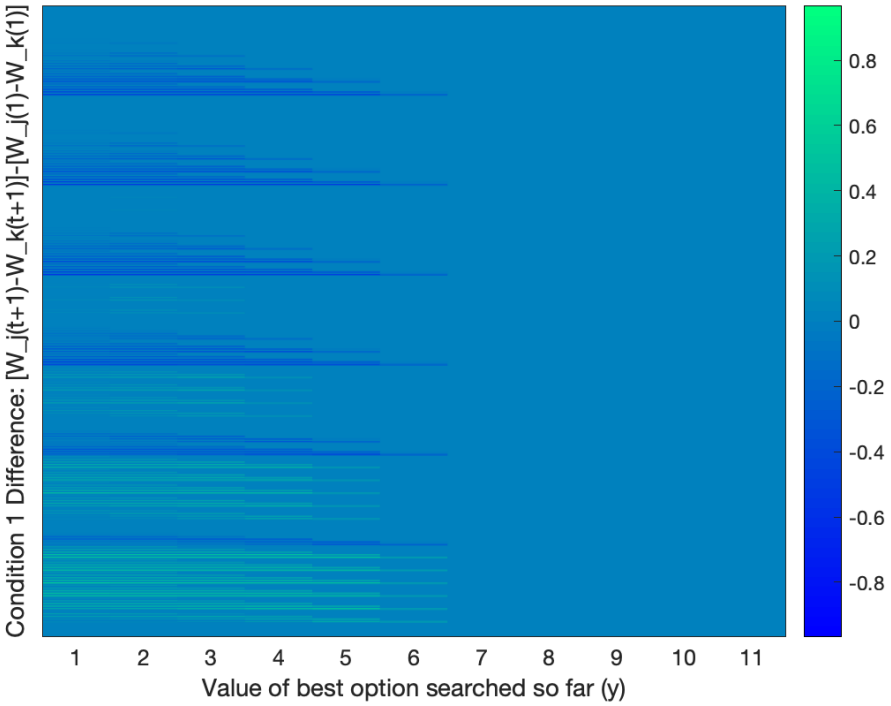
[53] We define a very small violation as follows: $0 < \left[W_j(\bar{S},t+1,y) - W_k(\bar{S},t+1,y)\right] - \left[W_j(\bar{S},1,y) - W_k(\bar{S},1,y)\right] \leq -0.1$.

Table C-1: Simulation Studies to Evaluate Condition 1

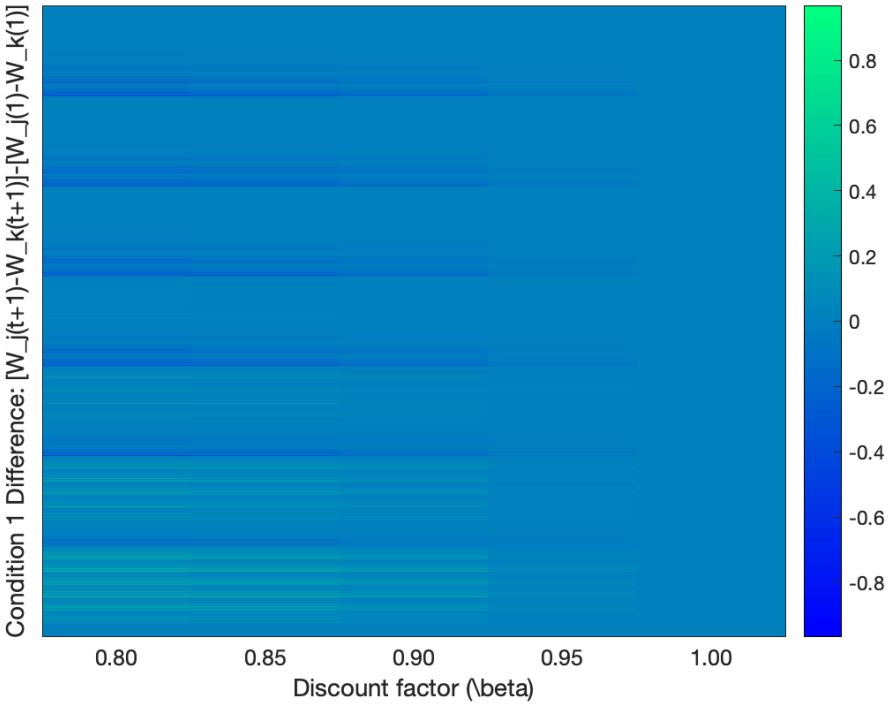
	(i) Num. of Parameter Combinations	(ii) Percent of Simulations in which Condition 1 is Satisfied	(iii) Percent of Simulations in which Condition 1 Holds Allowing for Violations up to 0.1
Set 1: "Vary All Values"	1,056,000	85.80	93.19
Set 2: "Vary Utility, Uncertainty, and Discount Factor"	53,900	92.32	96.32
Set 3: "Values Close to Estimates"	550	100.00	100.00

And second, we also display the results for simulation set 2 using heat maps in Figure C-3. Set 1 considers more than 1,000,000 parameter value combinations, and is thus too large to be displayed in a heat map, while set 3 shows no variation since Condition 1 is satisfied for all parameter combinations. For set 2, we plot the heat map first as a function of y (Figure C-3(a)), and then as a function of the discount factor β (Figure C-3(b)) to emphasize the impact of the values of these two parameters on the validity of Condition 1. The blue base color indicates a value of zero, lighter colors indicate positive values, and darker colors indicate negative values. In the vast majority of simulations, the difference $\left[W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)\right]-\left[W_j(\bar{S},1,y)-W_k(\bar{S},1,y)\right]$ is zero or positive, meaning that Condition 1 is satisfied. Only a small proportion of parameter values leads to a failure of Condition 1. Even if Condition 1 fails, the violations are mostly very close to zero. In Figure C-3(a), we show that larger values of y make it more likely that Condition 1 holds (for a discussion of why this is the case, see Section 5.3.2 in the paper). In Figure C-3(b), we show that larger values of the discount factor β (consistent with our empirical application in which the search gap length is usually short, approximately 4 days) also make it more likely that Condition 1 holds.

Figure C-3: Evaluating Condition 1 Using Heat Maps



(a) Case 1 - As Function of y



(b) Case 2 - As Function of β

Table C-2: Parameter Values and their Implications for Condition 1

	Condition 1 more likely to hold when
α	large
β	large
σ	small
t	small
y	large

To better understand which parameter values lead to a failure of Condition 1, we investigated the results more closely. Most failures of Condition 1 (61% of failures in set 1) happen when the uncertainty parameter σ takes on values that are different from 1. Since this parameter is hard to pin down empirically (indeed it is commonly normalized to 1 in empirical consumer search models, see discussion in Yavorsky, Honka, and Chen 2021), we consider the fact that Condition 1 holds in the vast majority of simulations when $\sigma = 1$ as encouraging. In the remaining simulations, larger values of t, large differences between mean utilities and search costs, as well as small fatigue levels are also more likely to lead to a failure of Condition 1. We summarize the main relations in Table C-2. However, once more, our main result highlights the fact that Condition 1 holds for the vast majority of considered parameter values, values that are beyond those obtained with our data.

Assumption 1: We use a similar approach to the one described above to evaluate when Assumption 1 is satisfied. To do so, we compute the difference between the continuation value in our model, $W_j(\bar{S}, t+1, y)$, and the continuation value in Weitzman, $W_j(y)$. Using the parameter value combinations from set 1 above, we find that Assumption 1 exactly holds, i.e., $W_j(\bar{S}, t+1, y) = W_j(y)$, in 77% of the simulations. When Assumption 1 is not satisfied with equality, the difference between the two continuation values is nearly always very small:[54] in 98% of cases, the difference in continuation values is smaller than 0.1, i.e., $\left| W_j(\bar{S}, t+1, y) - W_j(y) \right| \leq 0.1$. Given that Assumption 1 is made to aid our empirical specification, we judge these results as encouraging.

C.1.5. Proof by Induction – Relating $W_j(\bar{S}, t+1, y)$ and $W_j(\bar{S}, 1, y)$ for any $\bar{S}$

In Section 5.3, we showed that $W_j(\bar{S}, t+1, y) = W_j(\bar{S}, 1, y)$ when only one option is left to search. Suppose the same holds when $n > 1$ options are left to search. For simplicity, denote the set of n options left to search by $\bar{S}_n$. To complete the induction proof, we need to show that, if

[54] Note that the calculated values of the difference $W_j(\bar{S}, t+1, y) - W_j(y)$ in the simulations range from -17 to 17.

$W_j(\bar{S}_1, t+1, y) = W_j(\bar{S}_1, 1, y), \forall t \geq 0$ and $W_j(\bar{S}_n, t+1, y) = W_j(\bar{S}_n, 1, y), \forall t \geq 0$ (by the induction hypothesis), then $W_j(\bar{S}_{n+1}, t+1, y) = W_j(\bar{S}_{n+1}, 1, y), \forall t \geq 0$ will also hold. From Section 5.3, we know that this statement will not always hold, but that it will hold under fairly general conditions (that are met in our empirical application), such as a high fatigue level. Thus, our goal is to show that this statement holds under those same conditions. Using equation 9, we can write $W_j(\bar{S}_{n+1}, t+1, y)$ as

$$\begin{aligned} W_j(\bar{S}_{n+1}, t+1, y) &= V(\bar{S}_n, t+1, y) F_j(y) + \int_y^\infty V(\bar{S}_n, t+1, u) dF_j(u) \\ &= \max\left\{ y, \max_{k \in \bar{S}_n} \{ \max\{ -c_{k0} - \alpha(t+1) + W_k(\bar{S}_n, t+2, y), \beta[-c_{k0} + W_k(\bar{S}_n, 1, y)]\}\} \right\} F_j(y) \\ &+ \max\left\{ \int_y^\infty u dF_j(u), \max_{k \in \bar{S}_n} \left\{ \max\left\{ \int_y^\infty [-c_{k0} - \alpha(t+1) + W_k(\bar{S}_n, t+2, u)] dF_j(u), \int_y^\infty \beta[-c_{k0} + W_k(\bar{S}_n, 1, u)] dF_j(u) \right\} \right\} \right\}, \end{aligned} \tag{C3}$$

where for simplicity we let $\bar{S}_{n+1} \setminus j = \bar{S}_n$. Since by the induction hypothesis, $W_j(\bar{S}_n, t+1, y) = W_j(\bar{S}_n, 1, y), \forall t \geq 0$, it clear that, if α, β and/or y are large, then $W_j(\bar{S}_{n+1}, t+1, y)$ will not depend on t, and thus it will equal $W_j(\bar{S}_{n+1}, 1, y)$ (using a similar argument as we did for the case where $n = 2$ in Section 5.3). This is the same condition we observed when looking at two options left to search. Thus, our statement follows.

C.2. Deriving the Search Rules of our Proposed Solution in Theorem 5

Here, we derive the specific functions characterizing the search rules in Theorem 5. Suppose j is the alternative with the maximum reservation utility among alternatives not yet searched $\bar{S}$. Then given (t, y), the consumer will

- search j *now* if $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ and $z_j^3(t) < y$;
- search j *later* if $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ and $z_j^3(t) \geq y$;
- *stop* searching if $z_j^1(t) < y$ and $z_j^2 < y$.

The expressions defining (z_j^1, z_j^2, z_j^3) were derived in equations (12, 13, 14). For ease of exposition, we replicate them here:

$$\begin{aligned} c_{j0} &= W_j\left(z_j^1(t)\right) - z_j^1(t) - \alpha t; \\ c_{j0} &= W_j\left(z_j^2\right) - \frac{z_j^2}{\beta}; \\ c_{j0} &= W_j\left(z_j^3(t)\right) - \frac{\alpha t}{1-\beta}. \end{aligned} \tag{C4}$$

We know from Theorem 4 and the results derived in Section 5.3.3 that the consumer will

- search j now if $c_{j0} \leq \underbrace{W_j(y) - y - \alpha t}_{A}$ and $c_{j0} \leq \underbrace{W_j(y) - \frac{\alpha t}{1-\beta}}_{B}$;
- search j later if $c_{j0} \leq \underbrace{W_j(y) - \frac{y}{\beta}}_{D}$ and $c_{j0} > \underbrace{W_j(y) - \frac{\alpha t}{1-\beta}}_{B}$;
- stop searching if $c_{j0} > \underbrace{W_j(y) - y - \alpha t}_{A}$ and $c_{j0} > \underbrace{W_j(y) - \frac{y}{\beta}}_{D}$.

To simplify exposition, we denote the three terms that search cost c_{j0} is compared to by A, B, and D. Note that $c_{j0} > A$ if $y > z_j^1$ (based on equation C4 and the monotonicity results of the function $W_j(\cdot)$ derived in Section 5.3.3). Similarly, $c_{j0} > D$ if $y > z_j^2$, and $c_{j0} < B$ if $y > z_j^3$.

We proceed by enumerating all possible relations between c_{j0} and A, B, and D, and demonstrate that they imply the relations we need to show.

1. $c_{j0} > A$, $c_{j0} > B$, $c_{j0} > D$
2. $c_{j0} > A$, $c_{j0} \leq B$, $c_{j0} > D$
3. $c_{j0} > A$, $c_{j0} > B$, $c_{j0} \leq D$
4. $c_{j0} \leq A$, $c_{j0} > B$, $c_{j0} \leq D$
5. $c_{j0} \leq A$, $c_{j0} \leq B$, $c_{j0} > D$
6. $c_{j0} \leq A$, $c_{j0} \leq B$, $c_{j0} \leq D$
7. $c_{j0} > A$, $c_{j0} \leq B$, $c_{j0} < D$
8. $c_{j0} \leq A$, $c_{j0} > B$, $c_{j0} \leq D$

In cases 1 and 2, it is immediately clear that the consumer will want to stop searching, since conditions $c_{j0} > A$ and $c_{j0} > D$ are satisfied. These conditions are equivalent to stating that the consumer will stop searching when $z_j^1(t) < y$ and $z_j^2 < y$, as required.

In cases 3 and 4, the consumer will decide to search j after a break, since conditions $c_{j0} > B$ and $c_{j0} \leq D$ are satisfied. Note that the consumer will want to search j later regardless of the relation

between c_{j0} and A. This results in the equivalent statement that the consumer will search j later whenever the conditions $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ and $z_j^3(t) \geq y$ hold, as we needed to show (inequalities follow from our tie-breaking rules, defined in footnote 35). Similarly, in cases 5 and 6, we can show that the consumer will want to search j now because conditions $c_{j0} \leq A$ and $c_{j0} \leq B$ hold. These conditions are equivalent to those we need to show, i.e., that the consumer will search j now if $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ and $z_j^3(t) < y$.

Finally, cases 7 and 8 involve contradictions and thus do not impact the search rules we derived. Consider first case 7. If $c_{j0} > A$ and $c_{j0} \leq B$, it follows that $A < B$. Plugging in for the values of A and B and simplifying, we obtain the condition $\frac{y}{\beta} > \frac{\alpha t}{1-\beta}$. Also, if $c_{j0} > A$ and $c_{j0} \leq D$, then after simplification, we obtain the condition $\frac{y}{\beta} < \frac{\alpha t}{1-\beta}$, which contradicts the previous statement. A similar contradiction can be derived in case 8 as well. Here, the fact that $c_{j0} \leq A$ and $c_{j0} > B$ implies that $\frac{y}{\beta} < \frac{\alpha t}{1-\beta}$. However, the fact that $c_{j0} \leq A$ and $c_{j0} > D$ implies that $\frac{y}{\beta} > \frac{\alpha t}{1-\beta}$, which is a contradiction. We conclude that cases 7 and 8 cannot occur, so we do not need to consider them in describing consumers' optimal search rules.

Web Appendix D: Estimation Details

D.1. Our Model

The estimation using the logit-smoothed AR simulator involves the following steps:

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 $\left[z_j^d,\, z_j^{1d}(t),\, z_j^{2d},\, z_j^{3d}(t)\right]$
3. Calculate the following expressions for each draw d:
   (a) $v_1^d = z_{in}^d - \max_{k=n+1}^{J} z_{ik}^d \qquad \forall n \in \{1, \ldots, J-1\}$
   (b) $v_2^d = z_{i1}^d - u_{i0}^d$
   (c) $v_3^d = \max\left\{z_{in}^{1d}(t_n), z_{in}^{2d}\right\} - \max_{k=0}^{n-1} u_{ik}^d \qquad \forall n \in \{2, \ldots, s\}$
   (d) $v_4^d = \max_{k=0}^{n-1} u_{ik}^d - z_{in}^{3d}(t_n)$ if $t_n = t_{n+1} - 1 \qquad \forall n \in \{2, \ldots, s\}$
   (e) $v_5^d = z_{in}^{3d}(t_n) - \max_{k=0}^{n-1} u_{ik}^d$ if $t_n \neq t_{n+1} - 1 \qquad \forall n \in \{2, \ldots, s\}$
   (f) $v_6^d = z_{im}^{1d}(t_m) - \max_{k=0}^{s} u_{ik}^d \qquad \forall m \in \{s+1, \ldots, J\}$
   (g) $v_7^d = z_{im}^{2d} - \max_{k=0}^{s} u_{ik}^d \qquad \forall m \in \{s+1, \ldots, J\}$
   (h) $v_8^d = u_{ij}^d - \max_{k=0}^{s} u_{ik}^d \qquad \forall j \in \{0, 1, \ldots, s\}$
4. Compute $V^d = \frac{1}{1+M^d}$ for each draw d, where

$$M^d = e^{-v_1^d/\rho_1} + e^{-v_2^d/\rho_2} + (e^{-v_3^d/\rho_3} + e^{-v_4^d/\rho_3}) + (e^{-v_3^d/\rho_4} + e^{-v_5^d/\rho_4}) + (e^{-v_6^d/\rho_5} + e^{-v_7^d/\rho_5}) + e^{-v_8^d/\rho_6}, \qquad \text{(D1)}$$

   where the terms in parentheses represent the values of searching now, searching later, and of stopping, respectively, and where ρ_k are scaling parameters, chosen using Monte Carlo simulations (Honka 2014; Ursu 2018; Ursu, Wang, and Chintagunta 2020).
5. The average of V^d over the D draws and over consumers and websites gives the simulated likelihood function.

Similar to Ursu, Wang, and Chintagunta (2020), we use different scaling values ρ_k for each for the decisions consumers make. Using our Monte Carlo simulation that closely resembles the estimation data, we determined that the following scaling parameters recover the data well: $\rho = [-10, -3, -3, -10, -10, -5]$. Therefore, we estimate our model with the same set of scaling values. We also notice that estimating the model with all scaling parameters set to -3 also recovers the parameters well.[55]

[55] The analysis is available from the authors upon request.

D.2. Weitzman (1979) Model

In the Weitzman (1979) model, consumers search products in decreasing order of their reservation utilities and stop searching when the best observed utility so far exceeds the reservation utility of any unsearched option. The estimation procedure using the logit-smoothed AR simulator follows that in Honka and Chintagunta (2017) and Ursu (2018) and involves the following steps:

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.
3. Calculate the following expressions for each draw d:
   (a) $v_1^d = z_{in}^d - \max_{k=n+1}^{J} z_{ik}^d \qquad \forall n \in \{1, \ldots, J-1\}$
   (b) $v_2^d = z_{in}^d - \max_{k=0}^{n-1} u_{ik}^d \qquad \forall n \in \{1, \ldots, s\}$
   (c) $v_3^d = \max_{k=0}^{s} u_{ik}^d - z_{im}^d \qquad \forall m \in \{s+1, \ldots, J\}$
   (d) $v_4^d = u_{ij}^d - \max_{k=0}^{s} u_{ik}^d \qquad \forall j \in \{0, 1, \ldots, s\}$
4. Compute $V^d = \frac{1}{1+M^d}$ for each draw d, where

$$M^d = \sum_{k=1}^{4} e^{-v_k^d/\rho_k}, \tag{D2}$$

   where ρ_k are scaling parameters, chosen using Monte Carlo simulations (Honka 2014; Ursu 2018; Ursu, Wang, and Chintagunta 2020).
5. The average of V^d over the D draws and over consumers and websites gives the simulated likelihood function.

To ensure consistency across the estimation of the two models, we use the same scaling parameters, adjusted for the fact that, in the Weitzman model, the likelihood function is made up of only four components (rather than eight as in our model). The set of scaling parameters is given by $\rho = [-10, -3, -3, -5]$.

Web Appendix E: Calculating Reservation Utilities

An advantage of our proposed method lies in its ease of estimation due to its similarity to the Weitzman (1979) model: consumers search in decreasing order of reservation utilities z_j and also make search and purchase decisions based on threshold values of the best alternative observed so far. The main difference consists of computing the values of $\left[z_j^1(t), z_j^2, z_j^3(t)\right]$ in addition to that of z_j. We start by describing the method developed in Kim, Albuquerque, and Bronnenberg (2010) to compute z_j. Then we describe our method to compute $\left[z_j^1(\cdot), z_j^2, z_j^3(\cdot)\right]$.

Recall that $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$ and that the reservation utility z_j is the solution to $c_{j0} = W_j\left(z_j\right) - z_j$ (see equation (11)). From Kim, Albuquerque, and Bronnenberg (2010), we know that, under the assumption that ϵ_j is standard normally distributed,

$$B\left(m_j\right) \;=\; W_j\left(z_j\right) - z_j \;=\; \phi\left(m_j\right) + m_j\Phi\left(m_j\right) - m_j$$

with $m_j = z_j - \mu_j$, and $\phi(\cdot)$ and $\Phi(\cdot)$ representing the pdf and the cdf of the standard normal distribution. Given that a unique solution to $c_{j0} = B\left(m_j\right)$ exists (see Weitzman 1979 or our discussion in Section 5.3.3), one can invert the relation, solve for m_j, and then compute z_j from the relation $z_j = m_j + \mu_j$. Following prior work, we create a look-up table relating c_{j0} to m_j according to function $B\left(m_j\right)$, which we can use to solve for z_j for any search cost value.

To compute $z_j^1(t)$, we use a similar method. Recall that $z_j^1(t)$ is the solution to $c_{j0} + \alpha t = W_j\left(z_j^1(t)\right) - z_j^1(t)$ (see equation (12)). Since the additional term affecting search costs is constant in j, we can similarly create a look-up table relating $c_{j0} + \alpha t$ to $m_j^1(t)$ according to the function $B\left(m_j^1(t)\right)$ for the observed value of t (same function $B(\cdot)$ as above), and then solve for $z_j^1(t)$ from $z_j^1(t) = m_j^1(t) + \mu_j$.

To compute $\left[z_j^2, z_j^3(t)\right]$, we use a different method. Note that $W_j(z) = \phi(m) + m\Phi(m) + \mu$ for $m = z - \mu$ if $W_j(z) - z = \phi(m) + m\Phi(m) - m$. Next, recall that z_j^2 is the solution to $c_{j0} = W_j\left(z_j^2\right) - \frac{z_j^2}{\beta} = \phi\left(z_j^2 - \mu_j\right) + \left(z_j^2 - \mu_j\right)\Phi\left(z_j^2 - \mu_j\right) + \mu_j - \frac{z_j^2}{\beta}$. To solve for z_j^2, we create a look-up table relating c_{j0} and μ_j to values of z_j^2 by (numerically) solving the stated equation for all relevant values of c_{j0} and μ_j.

To solve for $z_j^3(t)$, we use a similar approach. From equation (14), we know $z_j^3(t)$ is the solution to $c_{j0} = W_j\left(z_j^3(t)\right) - \frac{\alpha t}{1-\beta}$, which can be rewritten as $c_{j0} + \frac{\alpha t}{1-\beta} = \phi\left(z_j^3(t) - \mu_j\right) + \left(z_j^3(t) - \mu_j\right)\Phi\left(z_j^3(t) - \mu_j\right) + \mu_j$. Using a look-up table relating $c_{j0} + \frac{\alpha t}{1-\beta}$ and μ_j to values of $z_j^3(t)$, after (numerically) solving the stated equation, we can compute $z_j^3(t)$.

Web Appendix F: Robustness Checks

F.1. Weitzman (1979) Model with Increasing Search Cost

Our model makes two changes to the Weitzman (1979) framework: (i) allows for search gaps; and (ii) models the effect of fatigue on search costs. To better isolate the effect of each change on parameter estimates, we also estimate a variation of the Weitzman (1979) model with increasing search costs (due to fatigue) but without search gaps, i.e., we only make one change to the Weitzman (1979) framework. Technically, this involves removing the option to search after the break from equation (3), but continuing to assume that fatigue affects search costs. To the best of our knowledge, this variation of the Weitzman (1979) model has not been studied by previous literature. As in our problem, there is no known optimal search rule. However, the solution we developed for our model can be used to derive an optimal search rule for this variation of the Weitzman (1979) model. Using the arguments we made in Section 5.3.1, the consumer searches products in the same order as in the original Weitzman (1979) model, i.e., in decreasing order of their reservation utilities z_j. And, the consumer stops searching when she encounters a product j for which $z_j^1(t)$ is smaller than the best option searched so far, and continues searching j otherwise.

Table F-1: Estimation Results for Weitzman Model with Increasing Search Costs

	(i) "Shirts, tops, & blouses" *Adapted Weitzman*		(ii) "Shoes" *Adapted Weitzman*
Utility		*Utility*	
aboutyou.com	-1.3920***	adidas.com	-1.2832***
	(0.0465)		(0.0408)
c-and-a.com	-0.8181***	debijenkorf.nl	-1.8562***
	(0.0400)		(0.0488)
debijenkorf.nl	-0.9952***	nelson.nl	-1.8205***
	(0.0410)		(0.0493)
esprit.nl	-1.6627***	nike.com	-1.3993***
	(0.0519)		(0.0415)
hm.com	-1.3008***	omoda.nl	-1.7360***
	(0.0447)		(0.0467)
jbfo.nl	-2.7536***	schuurman-shoenen.nl	-0.9964***
	(0.1486)		(0.0360)
msmode.nl	-2.0303***	spartoo.nl	-1.3912***
	(0.0643)		(0.0402)
peterhahn.nl	-2.1501***	vanharen.nl	-1.0869***
	(0.0787)		(0.0378)
your-look-for-less.nl	-1.7080***	zalando.nl	-0.6744***
	(0.0514)		(0.0329)
zalando.nl	-0.9790***	ziengs.nl	-1.9833***
	(0.0412)		(0.0556)
Outside option	1.9572***	Outside option	2.1079***
	(0.0475)		(0.0481)
Search cost (exp)		*Search cost (exp)*	
Baseline	-3.8829***	Baseline	-4.3036***
	(0.0043)		(0.0823)
Number of previous searches	-3.9986***	Number of previous searches	-7.7467***
	(0.0229)		(1.0988)
Number of Observations	27,924		27,756
LL	-8,644		-10,990

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

F.2. Empirical Results with Different Discount Factors

Table F-2: Empirical Results with Different Discount Factors

	(i) "Shirts, tops, & blouses"				(ii) "Shoes"		
	$\beta = 0.85$	$\beta = 0.90$	$\beta = 0.99$		$\beta = 0.85$	$\beta = 0.90$	$\beta = 0.99$
Utility				*Utility*			
aboutyou.com	-1.4088***	-1.4264***	-1.3887***	adidas.com	-1.3909***	-1.4067***	-1.3528***
	(0.0495)	(0.0452)	(0.0463)		(0.0518)	(0.0523)	(0.0388)
c-and-a.com	-0.8514***	-0.8654***	-0.8284***	debijenkorf.nl	-2.0139***	-2.0119***	-1.9721***
	(0.0442)	(0.0393)	(0.0411)		(0.0629)	(0.0635)	(0.0508)
debijenkorf.nl	-1.0103***	-1.0263***	-0.9881***	nelson.nl	-1.9855***	-1.9857***	-1.9520***
	(0.0439)	(0.0395)	(0.0419)		(0.0601)	(0.0581)	(0.0503)
esprit.nl	-1.6930***	-1.7031***	-1.6738***	nike.com	-1.5078***	-1.5281***	-1.4749***
	(0.0531)	(0.0516)	(0.0521)		(0.0557)	(0.0525)	(0.0383)
hm.com	-1.3300***	-1.3484***	-1.3119***	omoda.nl	-1.8981***	-1.9050***	-1.8739***
	(0.0462)	(0.0427)	(0.0456)		(0.0585)	(0.0579)	(0.0433)
jbfo.nl	-2.7833***	-2.7816***	-2.8667***	schuurman-shoenen.nl	-1.1165***	-1.1262***	-1.0827***
	(0.1673)	(0.1582)	(0.1785)		(0.0502)	(0.0493)	(0.0349)
msmode.nl	-2.0313***	-2.0354***	-1.9701***	spartoo.nl	-1.4880***	-1.4976***	-1.4437***
	(0.0695)	(0.0656)	(0.0656)		(0.0516)	(0.0502)	(0.0383)
peterhahn.nl	-2.1860***	-2.1891***	-2.1192***	vanharen.nl	-1.2211***	-1.2346***	-1.1910***
	(0.0828)	(0.0820)	(0.0834)		(0.0470)	(0.0484)	(0.0348)
your-look-for-less.nl	-1.7387***	-1.7456***	-1.7039***	zalando.nl	-0.8560***	-0.8707***	-0.8300***
	(0.0538)	(0.0502)	(0.0535)		(0.0437)	(0.0428)	(0.0311)
zalando.nl	-1.0287***	-1.0445***	-1.0062***	ziengs.nl	-2.1824***	-2.1667***	-2.1361***
	(0.0447)	(0.0400)	(0.0416)		(0.0679)	(0.0711)	(0.0534)
Outside option	1.5751***	1.5634***	1.8672***	Outside option	1.2884***	1.2679***	1.6514***
	(0.0379)	(0.0344)	(0.0255)		(0.0356)	(0.0366)	(0.0234)
Search cost (exp)				*Search cost (exp)*			
Baseline	-5.1029***	-4.9558***	-4.8938***	Baseline	-5.3588***	-5.0183***	-5.7978***
	(0.4244)	(0.3286)	(0.0198)		(0.2261)	(0.2267)	(0.0433)
Fatigue constant	-1.3755***	-1.7889***	-3.8558***	Fatigue constant	-1.6268***	-2.0332***	-4.0541***
	(0.0407)	(0.0342)	(0.0057)		(0.0522)	(0.0512)	(0.0189)
Number of Observations	27,924	27,924	27,924		27,756	27,756	27,756
LL	-9,331	-9,333	-9,393		-12,147	-12,157	-12,278

Standard errors in parentheses.

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

F.3. Spells without Ads

Ads may encourage consumers to start a new search session, but cannot explain why consumers stopped their search in a previous session, and therefore cannot explain search gaps.[56] Nevertheless, in our main estimation sample, we removed spells with clicks on social media, email, newsletter or retargeting ads. Here we re-estimate our model on data without any clicks on advertisements (except those occurring on search engines which are initiated by a consumer query). As expected, in the resulting data we see a considerable number of search gaps (322 in subcategory 1, 300 in subcategory 2, for 1,645 and 1,261 spells, respectively). Our results are robust to this change.

Table F-3: Estimation results on spells with no clicks on advertisements

	(i) "Shirts, tops, & blouses"		(ii) "Shoes"
Utility		*Utility*	
aboutyou.com	-2.0508***	adidas.com	-1.4470***
	(0.0725)		(0.0581)
c-and-a.com	-1.0616***	debijenkorf.nl	-1.6096***
	(0.0479)		(0.0655)
debijenkorf.nl	-1.7544***	nelson.nl	-1.6456***
	(0.0586)		(0.0621)
esprit.nl	-1.7026***	nike.com	-1.4520***
	(0.0576)		(0.0592)
hm.com	-1.3115***	omoda.nl	-1.5864***
	(0.0485)		(0.0650)
jbfo.nl	-2.6900***	schuurman-shoenen.nl	-1.6438***
	(0.1663)		(0.0636)
msmode.nl	-2.0443***	spartoo.nl	-2.1941***
	(0.0683)		(0.1106)
peterhahn.nl	-2.1122***	vanharen.nl	-1.2640***
	(0.0876)		(0.0526)
your-look-for-less.nl	-1.7283***	zalando.nl	-0.4356***
	(0.0552)		(0.0434)
zalando.nl	-1.0406***	ziengs.nl	-1.7926***
	(0.0469)		(0.0750)
Outside option	1.6364***	Outside option	-1.6035***
	(0.0255)		(0.0310)
Search cost (exp)		*Search cost (exp)*	
Baseline	-4.5373***	Baseline	-4.4997***
	(0.0899)		(0.1901)
Fatigue constant	-2.5503***	Fatigue constant	-2.7705***
	(0.0122)		(0.0309)
Number of Observations	19,740		15,132
LL	-5,628		-5,400

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

[56] Ursu, Simonov, and An (2021) provide a more in-depth analysis of the relation between online ads and search behavior.

F.4. Accounting for Left and Right Truncation

Table F-4: Estimation Results when Accounting for Truncation

	(i) "Shirts, tops, & blouses"		(ii) "Shoes"
Utility		*Utility*	
aboutyou.com	-1.5193***	adidas.com	-1.3502***
	(0.0793)		(0.0504)
c-and-a.com	-0.8584***	debijenkorf.nl	-1.7209***
	(0.0544)		(0.0608)
debijenkorf.nl	-1.2452***	nelson.nl	-1.7053***
	(0.0587)		(0.0605)
esprit.nl	-1.5623***	nike.com	-1.5013***
	(0.0673)		(0.0508)
hm.com	-1.1676***	omoda.nl	-1.6281***
	(0.0586)		(0.0578)
jbfo.nl	-2.7509***	schuurman-shoenen.nl	-1.3517***
	(0.6523)		(0.0523)
missetam.nl	-1.6608***	spartoo.nl	-1.5692***
	(0.0722)		(0.0561)
ullapopken.nl	-1.8274***	vanharen.nl	-1.1056***
	(0.0777)		(0.0472)
your-look-for-less.nl	-1.6001***	zalando.nl	-0.5207***
	(0.0750)		(0.0347)
zalando.nl	-0.9018***	ziengs.nl	-1.8203***
	(0.0497)		(0.0677)
Outside option	1.6461***	Outside option	1.5173***
	(0.0289)		(0.0203)
Search cost (exp)		*Search cost (exp)*	
Baseline	-4.1789***	Baseline	-4.6790***
	(0.0894)		(0.1101)
Fatigue constant	-2.5000***	Fatigue constant	-2.7427***
	(0.0143)		(0.0172)
Number of Observations	15,696		14,268
LL	-5,008		-5,544

Standard errors in parentheses.

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