

Honka, Elisabeth; Seiler, Stephan; Ursu, Raluca

Working Paper

Consumer Search: What Can We Learn from Pre-Purchase Data?

CESifo Working Paper, No. 10786

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Honka, Elisabeth; Seiler, Stephan; Ursu, Raluca (2023) : Consumer Search: What Can We Learn from Pre-Purchase Data?, CESifo Working Paper, No. 10786, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/282474>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Consumer Search: What Can We Learn from Pre-Purchase Data?

Elisabeth Honka, Stephan Seiler, Raluca Ursu

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Consumer Search: What Can We Learn from Pre-Purchase Data?

Abstract

Researchers are increasingly able to observe consumers' behavior prior to a purchase, such as their navigation through a store or website and the products they consider. Such pre-purchase (or search) data can be valuable to researchers in a variety of ways: as an additional source of information to estimate consumer preferences, to understand how firms can influence the search process through marketing mix variables, and to analyze how limited information about products shapes market outcomes. We provide an overview of these three areas with a particular emphasis on online and offline retailing.

JEL-Codes: D430, D830, L130.

Keywords: consumer search, limited information, consideration sets, retailing.

Elisabeth Honka
University of California Los Angeles / USA
elisabeth.honka@anderson.ucla.edu

Stephan Seiler
Imperial College London / United Kingdom
stephan.a.seiler@gmail.com

Raluca Ursu
NYU Stern / USA
rmu208@stern.nyu.edu

This draft: November 5, 2023

None of the authors received external funding for this paper. All errors are our own.

1 Introduction

The marketing literature has long recognized that consumers typically do not choose from all available products, but, instead, form consideration sets out of which they decide which product to buy (see, e.g., Howard and Sheth 1969, Gensch 1987, Hauser and Wernerfelt 1990, Roberts and Lattin 1991, Shocker et al. 1991).¹ Due to technological advancements, the consumer search process has increasingly become observable. For example, RFID tags and geo-location data enable firms and researchers to track consumers’ movements in a brick-and-mortar retail store. In online retail, browsing data permit researchers to observe which products a consumer inspects before making a purchase. In this article, we outline how researchers can use such pre-purchase data combined with a search model framework to gain new insights into consumer and firm behavior with a special focus on the retail sector. We discuss, in particular, how search data can be used to better understand consumer preferences, how firms can influence the search process through various marketing variables, and how information frictions shape market outcomes.

The research areas we cover are informed by two empirical patterns that have been documented across a variety of different markets. First, there is substantially more data on search than purchase behavior in many settings because some consumers search but do not make a purchase and some consumers search multiple products. For example, Ursu et al. (2023c) document 9.5 more search than purchase incidences in their apparel data, and Zhang et al. (2023) report 45 times more search than purchase incidences for shoes on a mobile app. Second, consumers typically search very few products. Previous research documents average consideration set sizes of 2.5 for savings accounts (Honka et al. 2017), 1.4 for cosmetics (Morozov et al. 2021), and 1.7 for home improvement products (Amano et al. 2022).²

These two patterns suggest several research areas where search data can be particularly valuable. In Section 3, we cover how the richness of search relative to purchase data allows researchers to better estimate consumer preferences. We discuss how these preference estimates can then be used to analyze substitution patterns between products and to improve firms’ targeting policies. In Section 4, we examine how firms can influence consumer search through marketing tools such as advertising, rankings, and store design and, in Section 5, we discuss equilibrium outcomes in markets with information frictions. Because consumers tend to search relatively few products, it is important for firms to influence search behavior, and search frictions likely play an important role in shaping market outcomes.

Before we turn to these three substantive research areas, we provide some general background on different types of search data that researchers typically have access to as well as model frameworks that have been used in the literature. We regard this content as a foundation for the discussion of substantive research areas in the later sections of this paper and will frequently refer back to it. We

¹We use the terms “consideration set” and “search set” interchangeably in this paper.

²Additional examples of consumers’ typically small consideration set sizes are 2.4 for auto insurance (Honka 2014), 2.8 - 6.4 for digital cameras (Bronnenberg et al. 2016), 2.3 for online used cars (Gardete and Hunter 2020), 1.1 for new car purchases (Yavorsky et al. 2021), and 1.9 for shoes (Zhang et al. 2023).

focus on the use of search data in a retailing context and differentiate between brick-and-mortar and online retailing due to differences in available data and in relevant marketing variables. We note that we do not provide an exhaustive overview of the search literature, but, instead, our aim is to highlight a set of broad research areas where search data is particularly useful. We refer the interested reader to Honka et al. (2019) for an overview of the search literature.

2 Search Data and Models

In this section, we discuss the types of pre-purchase data that are typically available to researchers. Then, we turn to the model frameworks which have been employed to analyze such data starting with the Weitzman (1979) sequential search model, which has become the most common search framework used in empirical work. However, the sequential search model is not well-suited to answer some research questions and to accommodate certain types of data. We therefore discuss a series of extensions and conclude the section by providing guidance on how to select an appropriate model framework as a function of the available data and the research question that one is trying to address.

2.1 Common Types of Pre-Purchase Data

Pre-purchase data can, in principle, comprise any activity that a consumer engages in to gather information about products before making a purchase. For example, a typical pre-purchase journey of a consumer on an online retail platform might consist of entering a search query, filtering by product characteristics such as price or star rating, and finally visiting the product detail pages for some products and purchasing one of them. In the following, we first discuss common types of pre-purchase data found in online retail and then in brick-and-mortar retail settings.

The most commonly used type of data containing information on consumers’ online pre-purchase behavior is clickstream data. The unit of observation in clickstream data is a URL of the webpage that a consumer visited, and, typically, the researcher is able to also observe the time of the visit. In other words, clickstream data allows the researcher to observe the identities and the sequence of products a consumer searched. Therefore, most of the existing literature on consumer search in the online context focuses on the identities and the order of searched products and ignores other steps in the process (see, e.g., Ursu 2018, Jiang et al. 2021, Morozov et al. 2021). Sometimes, researchers have access to more detailed data that also enable them to observe how consumers scroll and navigate within a given webpage. In such cases, other consumer decisions such as sorting and/or filtering of search results (see, e.g., Chen and Yao 2017, Koulayev, De los Santos and Koulayev 2017), scrolling decisions on a webpage (Korganbekova and Zuber 2023), and the search queries that consumers employ to generate search results (Padilla et al. 2019) can be modeled.

Another source of both online and offline search data is based on eye-tracking. In a physical store, eye-tracking glasses enable researchers to gather pre-purchase data (Xu et al. 2023). However, eye-tracking data gathered from consumers who wear glasses may suffer from the fact that

consumers know that they are being observed, thus leading to an unnatural shopping experience. In the online environment, collecting eye movement data using a computer camera represents a less obtrusive option (Martinovici et al. 2021; Ursu et al. 2022). Similar to clickstream data, eye-tracking data permits researchers to observe the identities and sequence of products a consumer searches.³ However, eye-tracking data has the potential for an even more granular level of observation by recording which product attributes a consumer looked at (Ursu et al. 2022) and for how long (Ursu et al. 2020).

Except for eye-tracking, data on pre-purchase behavior in brick-and-mortar retail tends to be different in nature from online data, which affects how it can be analyzed and what kind of research questions can be addressed based on such data. A first type of offline pre-purchase data is path-tracking data that permits the researcher to observe how consumers walk through a physical store based on radio-frequency identification (RFID) tags (Hui et al. 2009, Seiler and Pinna 2017, Seiler and Yao 2017) or video tracking (Jain et al. 2020). Path-tracking data contains information that is less detailed than typical online search data and does not allow the researcher to observe which specific products a consumer considers. Instead, the researcher can only track which areas of the store were visited, when they were visited, and how much time a consumer spent in front of a particular category. The latter can be seen as a measure of category-level search intensity (Seiler and Pinna 2017), but it is a cruder measure than the detailed information on product page visits typically available in online data.

One disadvantage of both RFID and video data is that it is usually restricted to one store. This disadvantage can be overcome using a different type of offline data, namely, cellphone tracking data where a consumer’s geo-location is recorded based on the location of her phone. These data enable researchers to track consumers across retail outlets. However, this type of data is usually not precise enough to record consumer movements within a store. Yavorsky et al. (2021) use geo-location data to study consumers’ shopping process for new cars. The authors use the cellphone location to determine which dealerships consumers visit. Because search is modeled at the dealership level, the data in Yavorsky et al. (2021) is akin to online search data in the sense that it permits the researchers to observe the identities and sequence of visited dealerships.

2.2 The Weitzman (1979) Sequential Search Model

Weitzman (1979) proposed a solution to the sequential search model for differentiated products that has emerged as the workhorse model for empirical work using consumer search data. While we do not exclusively focus on studies using this particular framework, it is the most common model of search behavior and many alternative models can be seen as extensions or modifications to the Weitzman (1979) framework. In this section, we describe the mechanics of the model at a relatively high level to provide the necessary background needed for the applications we discuss in

³In the case of eye-tracking data, one needs to make an assumption about what duration of eye fixation on a product constitutes a consumer considering a product. The industry standard is that an eye fixation is defined as a period of 200-400 milliseconds during which the eye is relatively fixed on an area. In online settings, the click decision more cleanly identifies when a product was searched.

the following sections. We refer the interested reader to the more detailed technical overview of the sequential search model in Ursu et al. (2023a).

Similar to the well-established discrete choice modeling approach based on purchase data (see Dubé 2019 for an overview), we assume that consumer i receives utility u_{ij} from purchasing product j :

$$u_{ij} = \delta_{ij} + \varepsilon_{ij} . \quad (1)$$

We let δ_{ij} denote the part of utility which is known by the consumer prior to search. Consumers know the distribution of ε_{ij} prior to search and need to incur a search cost c_{ij} to learn its realization for a specific product. Consumers search products sequentially until they decide to stop searching and purchase one of the searched products or choose the outside option of not purchasing. Often, ε_{ij} is thought of as a “match value” which reflects an idiosyncratic taste shock that consumers learn after getting more detailed information about a product. In general, the unknown component ε_{ij} can also reflect product characteristics such as price (Honka 2014) or combinations of characteristics (Yao et al. 2017, Compiani et al. 2023).

If the post-search utility component ε_{ij} is independently distributed across products for a given consumer, then the payoff from searching a specific product does not depend on which other products were searched before. Weitzman (1979) shows that, under this assumption, the search process can be summarized by a simple set of rules which depend on the so-called reservation value (or reservation utility) z_{ij} that summarizes the value of searching a particular product. A product’s reservation value is equal to the value of the current utility that would make a consumer indifferent between searching product j or not searching it. Under the assumptions laid out above, the reservation utility is given by an additively separable expression which depends on search costs and pre-search utility:

$$z_{ij} = \delta_{ij} + g(c_{ij}), \quad (2)$$

where $g(c_{ij})$ is a known function which monotonically decreases in search cost c_{ij} and only depends on search costs and the distribution of the post-search component ε_{ij} .⁴

The optimal search sequence consists of a consumer searching products in decreasing order of their reservation utilities (“search rule”) until the maximum realized utility among the searched products exceeds the maximum reservation utility among the unsearched products (“stopping rule”). After stopping to search, the consumer chooses the highest-utility option from the searched products, including the outside option of not purchasing (“purchase rule”). These are the three rules that describe optimal consumer search and purchase behavior and that are often called the “Weitzman rules.” Importantly, the chosen option might not coincide with the utility-maximizing product in the full assortment because the consumer does not learn the full utility of all options.

⁴The function ranges from negative infinity (if search costs go to infinity) to infinity (if search costs are zero). See Kim et al. (2010) for an example of the expression for $g(c_{ij})$ when ε_{ij} is normally distributed.

Utility and search costs are usually parameterized in a way that resembles standard perfect information choice models:

$$\begin{aligned}\delta_{ij} &= \mathbf{X}'_j \boldsymbol{\beta}_i - \alpha_i p_j + \mu_{ij} \\ c_{ij} &= \mathbf{Z}'_j \boldsymbol{\gamma}_i,\end{aligned}$$

where p_j denotes price and \mathbf{X}_j and \mathbf{Z}_j denote vectors of product characteristics that enter preferences and/or search costs. Both sets of variables can include product fixed effects, physical product characteristics, and variables that capture saliency such as product ranking on a webpage or advertising.

A natural parameterization consists of including physical product characteristics as part of preferences whereas variables that do not procure utility but affect the visibility and salience of a product are included in the set of variables that shift search costs. It is also possible to include variables as part of both search costs and utility and to test whether a given variable affects search costs and/or preferences (see Ursu et al. 2023a). Intuitively, a variable that lowers search costs will lead to the product being search more often and earlier, whereas a variable that increases utility will lead to higher conversion conditional on search.⁵

Two properties of the search model are particularly important for the research areas we cover later in the paper. First, the order of search, when search stopped, as well as the purchase decision are informative about preferences. Therefore, search data provide additional information about consumer preferences beyond solely purchase data, that is used in traditional demand estimation approaches. Moreover, as we discussed in the introduction, search data are often much more abundant than purchase data. Therefore, search data are likely a valuable source of information to learn about consumer preferences. In Section 3, we discuss how researchers can use search data to better understand preferences.

Second, search data combined with a model of consumer search behavior enables the researcher to model the impact of variables that shift search costs by making products more salient. Many marketing decisions such as product rankings on a webpage or shelf placement in a physical store are best thought of as search cost shifters and a search framework therefore provides a natural lens to study the impact these marketing activities. In Section 4, we discuss how firms can influence consumers' search decisions and how we can use data to learn about the impact of marketing on search.

2.3 Beyond the Weitzman (1979) Search Model

The Weitzman (1979) search model allows the researcher to model the sequence of searches and the identities of the considered products. However, other parts of the search process, such as the usage of sorting or filtering tools or the search queries a consumer enters, cannot be easily modeled using

⁵See Ursu et al. (2023a) for a formal argument regarding the separate identification of preferences and search costs.

the standard Weitzman framework. As we discussed in Section 2.1, researchers sometimes have access to additional pre-purchase behaviors, but only a small number of papers has incorporated information beyond the sequence of searched products. We discuss some of these papers in the following.

Chen and Yao (2017) amend the Weitzman (1979) model to incorporate consumers’ decisions to refine search results. The authors estimate a sequential search model in which consumers are uncertain about product characteristics and sorting and filtering changes the distribution of characteristics in the list of search results.⁶ Intuitively, a consumer who chooses to sort products by price will tend to be more price sensitive and hence the sorting decision provides valuable information about preference parameters. Koulayev (2014), De los Santos and Koulayev (2017), and Gu and Wang (2022) similarly incorporate sorting and filtering into a sequential search model.

Using a version of the search and discovery model developed by Greminger (2021), Zhang et al. (2023) allow consumers to take different search routes (such as navigating through the main category page or the sales page of the website) to reach the same product. Similar to the consumer decision to refine search results, the choice of a search route also provides information about consumer preferences. The ways in which consumers are allowed to navigate a webpage are set by companies and constitute a major decision through which firms can impact consumers’ search and purchase behavior. We return to this issue in Section 4.2.

Search queries can also contain information about preferences because consumers might mention specific brands or product attributes that they are looking for in a category. Due to the high-dimensional and open-ended nature of search queries, they are inherently more difficult to analyze than the binary decision of whether to search a product or to use sorting and filtering options. The closest work dealing with search queries is that by Liu and Toubia (2018), Liu and Toubia (2020), and Liu et al. (2021) who analyze consumers’ content preferences based on their search queries. All three papers analyze search queries in isolation, do not combine search query data with purchase data or other information about consumer search behavior, and do not use a search framework. Padilla et al. (2019) is the only paper we are aware of that combines search queries and search data to analyze the customer’s journey. The authors do not estimate a search model, but rather develop a statistical model of the consumer’s purchase journey that can be used to predict consumer choices based on their pre-purchase behavior.

Finally, several papers incorporate additional information about the products a consumer searched. Ursu et al. (2020) amend the sequential search model by incorporating the decision of how long to search a product in the sequential search model. Equipped with eye-tracking, Ursu et al. (2022) observe not only which products a consumer looked at, but also which product attributes a consumer inspected. The authors use this additional information to estimate a model of search at the brand-attribute level. Gardete and Hunter (2020) also study attribute-level search behavior using browsing data. They are able to observe attribute-level search because, in their

⁶The paper also assumes that search costs are higher for products with higher ranks (i.e., products that are displayed further down on the results page). Therefore, re-sorting products with desirable characteristics towards the top will tend to lower search costs.

data, different information about products is located on different webpages.

2.4 Developing a Search Model for an Empirical Application

Next, we discuss the decisions a researcher needs to make to develop a suitable empirical framework for a specific project. The goal of this section is to provide guidance on how to set up a search model framework based on the research question, the available data, and the institutional features of a particular setting.

One of the most important decisions the researcher has to make is to decide which information consumers are searching for, i.e., the researcher has to choose whether a price-search, a match-value-search or a multiple-characteristics-search model is most appropriate in her empirical context.⁷ To this end, the researcher needs to assess which product attributes are immediately observable to consumers and which product attribute(s) are hidden or can only be accessed with additional effort. For example, when estimating demand for an infrequently purchased product sold on an online retail platform, price information is typically easily observable as it is usually displayed on product list pages. However, the consumer can only learn information on product material, country of origin, size or review content when she visits a product detail page. In this empirical context, a match-value search model is likely appropriate where the match value would capture all the information a consumer would learn when visiting the product detail page. In another empirical context, a price-search model might make more sense. For example, when modeling demand for a frequently purchased product, such as cereal or detergent, in an offline context, consumers typically know the relevant product attributes but need to search for prices due to promotions. Similarly, a price-search model is more appropriate if a consumers has decided which product to buy but needs to visit different retailers to find the best price.

Another decision the researcher has to make is to determine whether observed consumer behavior is compatible with the Weitzman (1979) sequential search model or whether an amended sequential search model or a different model altogether would better describe it. For example, the Weitzman (1979) model requires the assumption that consumers have rational expectations about the distribution of uncertainty. Such an assumption is appropriate when the consumer has prior experience with a category, but less so when the consumer is purchasing for the first time (or after a long break). Another critical building block of the Weitzman (1979) model is the assumption of independence of draws from the uncertainty distribution. It implies that the consumer only learns the value of the draw from the searched product, but does not learn anything about any unsearched products or their uncertainty distribution. For example, this assumption does likely not hold when products share common characteristics that consumers learn about through search (see, e.g., Gardete and Hunter 2020 on how to model consumer search in such a situation).

A third decision the researcher has to make is which institutional details to incorporate in the

⁷The multiple-characteristics-search model requires the researcher to assume that consumers know the joint distribution of multiple characteristics. Arguably, this is a stronger assumption than presuming that consumers know the distribution of one characteristic.

model and which ones to ignore. The decision should be guided by their importance for the research question and their impact on the consumer’s shopping and purchase process. For example, online retailers typically show products as a ranked list on a screen. Taking rankings into account is important on large-screen devices where 30+ products are shown on a screen. However, one could argue that rankings can be ignored in mobile commerce where typically only 2 - 6 products are shown on a screen. A second example is a paper by Lam (2023) who studies how the dual role of Amazon as both a platform and a retailer affects consumers and third-party sellers. To answer this question, Lam (2023) needs to model the impact of the “BuyBox” (which assigns one seller as the default) on consumer search because this is how Amazon directly competes with third-party sellers for the sale of a product when both entities are selling it.

To summarize, the Weitzman (1979) model provides a natural starting point when the researcher is interested in estimating demand with consumer search. However, contrary to demand estimation under the assumption of perfect information, the researcher needs to make an assumption about what consumers learn while searching. The researcher also needs to assess whether the empirical setting is compatible with the Weitzman (1979) framework or whether a more complicated model is required. Finally, the researcher has to decide which market features need to be modeled explicitly and how to incorporate them in the search model framework.

3 Using Search Data to Learn Preferences

A key input into many marketing decisions is an understanding of consumer preferences. At the most basic level, a researcher or manager will need to understand how aggregate demand reacts to changes in prices or advertising to optimally set those marketing variables. At a more granular level, marketers are often interested in understanding heterogeneity in consumer preferences to target specific marketing activities to specific consumers (Rossi and Allenby 1993). To gain an understanding of consumer preferences, the most common approach is to estimate some form of demand model using historic data on consumers’ purchases. In this section, we re-cap how demand estimation is typically implemented using only purchase data and then turn to outlining the additional information that can be gleaned from search data.

3.1 Background: Choice Models

Before delving into how search data can provide additional information, it is useful to briefly re-cap what we can learn from purchase data alone. Typically, a researcher would posit that a consumer i obtains a given level of utility when purchasing product j :⁸

$$u_{ij} = \mathbf{X}'_j \boldsymbol{\beta}_i - \alpha_i p_j + \mu_{ij}$$

⁸If a researcher has access to panel data with repeated observation for each consumer, a time subscript t can be added to the utility function.

where p_j denotes price and \mathbf{X}_j denotes a vector of product characteristics and can include product fixed effects and physical product characteristics. Finally, μ_{ij} denotes a taste shock that is iid across consumers and products. Integrating out over the taste shocks⁹ (and possibly other stochastic elements of utility) yields expressions for choice probabilities which are then used to form the likelihood. The model is estimated by finding the parameters that generate predicted choice probabilities that closely match the empirical realizations, i.e., maximize the likelihood function. While there are many settings in which researchers estimate more complicated models of demand that deal with quantity choice or dynamic considerations such as stockpiling, we focus on the workhorse static discrete choice model because of its wide usage and because of the close relationship to models of consumer search which we describe below.

Depending on the research question or marketing decision one wants to address and the type of data available, a specific parameterization of the utility function can be estimated. We will focus on two applications: an understanding of substitution patterns between products and the estimation of heterogeneous preferences across consumers that form the basis for targeted marketing actions. Turning to substitution patterns first, a typical approach is to allow for heterogeneous preferences over characteristics which leads to substitution patterns where products that are more similar in terms of the observed characteristics \mathbf{X}_j have larger cross-price elasticities. Berry et al. (2004) note that secondary choice data, typically obtained from surveys, is particularly helpful for estimating heterogeneous preferences because it permits the researcher to observe along which dimensions the consumer’s most preferred and second-most preferred products are similar. As we show below, search data analyzed through the lens of the sequential search model provides information about preferences akin to the information contained in second-choice data. A second application of choice models is to focus specifically on estimating preferences at the individual consumer level. Allenby and Rossi (1999) make the case that such an analysis is particularly important for many applications in marketing because individual-level estimates are a prerequisite for designing targeted marketing strategies such as targeted coupons, which are common in retailing.

3.2 Consumer Search & Substitution Patterns

Consumer search data allow the researcher to observe which products a consumer considered before making a purchase. As described in Section 2.2, a researcher typically assumes a utility of the following form:

$$\begin{aligned} u_{ij} &= \delta_{ij} + \varepsilon_{ij} \\ &= (\mathbf{X}'_j \boldsymbol{\beta}_i - \alpha_i p_j + \mu_{ij}) + \varepsilon_{ij} \end{aligned}$$

which differs from the perfect information choice model only due to the presence of the additional component ε_{ij} which is discovered after search. Moreover, a consumer will search products in

⁹Taste shocks are usually assumed to be normally (extreme value) distributed which leads to probit (logit) expressions for choice probabilities.

decreasing order of reservation utility, where the reservation utility is given by

$$z_{ij} = (\mathbf{X}'_j \boldsymbol{\beta}_i - \alpha_i p_j + \mu_{ij}) + g(c_i).$$

When search costs are constant across products, reservation utilities are equal to the pre-search utility, δ_{ij} , plus a term that does not vary across products and therefore does not impact the ranking of reservation utilities. Therefore, consumers will search products in decreasing order of their pre-search utility. Depending on the length of the search spells the researcher will observe the first, second, third, etc., highest-utility options for each consumer. Through the lens of the sequential search model, data on the search sequence therefore takes exactly the same form as second-choice data in a perfect information choice model.¹⁰

It follows that the arguments with regards to second-choice data and the information they provide about preferences directly carry over to search data. The intuition is that consumers who strongly care about a specific product characteristic will only search products with high values of that characteristic.¹¹ Bronnenberg et al. (2016) provide empirical support for the usefulness of search data as second-choice data by showing that products in a consumers' search set tend to be similar in terms of their characteristics. Search data has also been used in a more descriptive fashion as a predictor of market-structure by using co-search patterns as a basis for conceptual maps (Kim et al. 2011, Ringel and Skiera 2016). While the conceptual map approach is not based on an explicit model of optimal consumer search behavior, it does indirectly leverage the search model insight that frequently co-searched products are products with a high degree of substitutability.

Although the idea that search patterns are informative about substitution appears in some early empirical search papers (e.g., Kim et al. 2010), the focus of most search papers has not been to estimate cross-price elasticities. Only recently, a few papers have started to use search data specifically with the goal of estimating flexible substitution patterns. Amano et al. (2022) use search data from a retailer selling home improvement products and estimate product-pair-specific correlations in search propensities. Contrary to a characteristics-based approach, the model in Amano et al. (2022) can therefore capture substitutability along dimensions that are not observed by the researcher.¹² Armona et al. (2021) apply a related idea to the market for hotels. They use a search model to estimate latent characteristics over which consumers have heterogeneous preferences. The logic behind their approach is similar to the second-choice logic described above. Namely, if two products are often searched together, the model will rationalize the co-search behavior by assigning a high value to a latent characteristic present in both products so that consumers who value this characteristics will tend to search both products. Contrary to Amano et al. (2022), the Armona et al. (2021) paper does not estimate product-level correlations, but instead estimates a

¹⁰We note that this is only true when search costs are homogenous across products.

¹¹An important assumption is that the characteristic is known prior to search. A characteristic whose value is only discovered after search can, by definition, not influence the search order.

¹²The paper uses a consideration set framework, rather than a model of consumer search, but the general idea could, in principle, be implemented within a sequential search framework. However, a two-stage consideration-then-choice framework is computationally lighter.

lower-dimensional representation of those correlations via the latent characteristics structure. This approach can be especially valuable when assortments are large, because it lowers the number of parameters that need to be estimated and can increase efficiency. In a second step, Armona et al. (2021) then uses the inferred characteristics as an input in a discrete choice demand model.

In summary, search data analyzed through the lens of a sequential search model or a related framework plays a similar role to second-choice data which is known to help estimate heterogeneous preferences over product characteristics in perfect information demand models. Moreover, because search data tends to be significantly more abundant than purchase data, it has recently been used to estimate similarity along *unobserved* dimensions either by estimating a latent characteristics model or an unconstrained correlation structure.

3.3 Using Search Data for Targeted Marketing

Apart from its value for understanding the shape of the aggregate demand function and, in particular, cross-price elasticities, search data can also be helpful for identifying individual-level parameters. Similar to the estimation of consumer-level parameters from purchase data, these estimates can then be used to derive optimal targeted marketing strategies.

Jiang et al. (2021) use search data to derive better retargeting strategies for consumers who searched in a category but did not make a purchase. One of the empirical patterns we discussed in the introduction – the fact that conversion rates in many online markets are relatively low – is a key motivation for the study. The authors show that information on the set of products a consumer searched before leaving the platform without a purchase can be useful to improve a firm’s retargeting strategy. They also show that information on search behavior is more valuable than information on consumer demographics.

Morozov et al. (2021) use panel data on consumer search behavior from an online cosmetics retailer and are, to the best of our knowledge, the first paper to observe consumers repeatedly searching and purchasing in the same category. The authors show that modeling search behavior leads to lower estimates of preference heterogeneity across consumers. Intuitively, search cost can lead to high state dependence (repeated purchases of the same product) because consumers do not search beyond their most preferred product. In a model without search, high persistence in choice instead needs to be rationalized by strong heterogeneity in preferences. Apart from removing bias, search data also leads to more precise estimates of consumer-level parameters and therefore to more profitable targeting strategies. In their setting, the authors show that profits from targeted pricing increase by 9%, on average, and can increase by as much as 15% for some brands when search data is used.

Finally, Padilla et al. (2019) use search data to predict consumer demand for flights. They argue that, in many settings such as hotels and flights, firms only have access to sparse purchase information and additional information on the customer’s journey can therefore help firms better understand consumer preferences. The paper emphasizes the importance of context heterogeneity in the market for flights such as heterogeneity for business versus leisure flights. The paper also

makes use of search query data which contains information about the destination, date, and other trip characteristics. The authors show that additional information from product searches and queries leads to an improvement in model predictions with regards to purchase incidence and the characteristics of the purchased product.

The above papers demonstrate that search data can provide helpful information for targeted marketing. This is especially true in settings in which consumers purchase infrequently, such as in many online settings, and where consumers often engage with a category by searching products but ultimately do not purchase.

4 How Firms Can Influence Search Behavior

A second area where consumer search data and a search model framework are particularly useful is the analysis of marketing tools that can affect consumers' search behavior. As we described in the introduction, most empirical studies on consumer search find that consumers only consider a very limited set of products. Therefore, understanding how firms can impact search behavior to make sure that consumers consider their products is likely of first-order importance to companies. Before turning to specific tools that firms can use to influence search behavior, we first outline how changes in search costs influence behavior in a search model, how the impact of marketing decisions on search costs can be identified, and how the sequential search model framework can be used to analyze counterfactual marketing strategies.

To understand how a firm's marketing decisions can affect consumer search behavior, it is instructive to look at the main equations governing search behavior in the sequential search model outlined in Section 2.2. Consumers will search products in descending order of reservation utility z_{ij} :

$$z_{ij} = \delta_{ij} + g(c_{ij}), \quad (3)$$

where δ_{ij} denotes pre-search utility and $g(c_{ij})$ is a decreasing function of search cost. Therefore, an increase in a product's search costs will lower its reservation utility and make it less likely that the product will be searched. Recall that utility and search costs are typically parameterized as follows:

$$\begin{aligned} u_{ij} &= \delta_{ij} + \varepsilon_{ij} \\ &= (\mathbf{X}'_j \boldsymbol{\beta}_i - \alpha_i p_j + \mu_{ij}) + \varepsilon_{ij} \\ c_{ij} &= \mathbf{Z}'_j \boldsymbol{\gamma}_i, \end{aligned}$$

where p_j denotes price and \mathbf{X}_j and \mathbf{Z}_j denote vectors of product characteristics that enter preferences and/or search costs. Both sets of variables can include product fixed effects, physical

product characteristics, and other marketing variables, such as product rankings on a webpage or advertising, that are under a firm’s control.

One key issue that comes up in this setting is how to separate the impact of a particular marketing variable on search costs versus preferences. For example, advertising could in principle be seen as shifting search costs by making certain products more salient to the consumer or as shifting preferences by increasing the consumer’s perceived utility from the product.¹³ Because search costs and preferences have a different impact on search and purchase decisions, one can separately identify the impact of a given marketing variable on preferences and search costs. The reservation utility formula in equation (3) shows that a decrease in search costs and an increase in (pre-search) utility both shift the reservation utility in the same direction. However, after stopping to search, the consumer will choose the highest-utility product solely based on its utility. Search costs are not relevant at this point of the choice process. Intuitively, increasing utility will increase the search and purchase probability, whereas lowering search costs will increase search but affect purchases less. Ursu et al. (2023a) provide a formal identification argument for the separate identification of search costs and preferences. Ursu (2018) implements a test that follows a similar logic and shows that rankings impact search decisions, but not conversion conditional on search. She therefore argues that rankings should enter search costs and not utility.

Regardless of whether a given variable enters utility or search costs, one needs random variation to estimate its causal impact. The concerns around endogeneity are not unique to search models and are similar in spirit to the issue of price endogeneity in a standard demand model settings. For instance, the marketing variables of interest might be correlated with product quality in specific ways: higher quality products tend to be ranked higher on a webpage because they are deemed more relevant to consumers. In such a situation, the estimated effect of rank on search costs might be biased upwards because high-ranked products are searched earlier partly because of their rank but also partly because of their higher quality. Similar arguments can be made for all the other marketing variables discussed below that are under the firm’s control. This represents a challenge for researchers as they are either left with quantifying correlational relationships, having to model the firm’s decision-making or must find exogenous/experimental variation.

Once the effect of a specific marketing action on search costs has been inferred, it is easy to use the search model framework to compute relevant counterfactuals. For example, counterfactual purchase probabilities can be calculated and, together with information on mark-ups, used to compute counterfactual profits. Similarly, it is possible to compute welfare for different values of the marketing variables. Contrary to perfect information demand models, welfare calculations also need to take search costs under different counterfactual scenarios into account. Consumer surplus in the search model context is given by

¹³ Additionally, marketing variables can also indirectly affect consumer search via changing consumers’ awareness sets. We discuss these effects in detail in the context of advertising in Section 4.3.

$$\mathbb{E}(CS_i) = \frac{1}{\alpha_i} \times \int \int \left(u_{ij}^* - \sum_{k \in S_i} c_{ik} \right) \times \phi(\boldsymbol{\varepsilon}) \phi(\boldsymbol{\mu}) d\boldsymbol{\varepsilon} d\boldsymbol{\mu},$$

where u_{ij}^* denotes the utility of the chosen option and the second term denotes the total search costs the consumer incurs by searching the set of products S_i . We integrate out over the stochastic elements of utility, i.e., the pre- and post-search taste shocks μ_{ij} and ε_{ij} . Search data combined with a structural framework like the sequential search model therefore allow researchers to analyze the impact of counterfactual marketing strategies on both firm profits and consumer welfare.

Based on the conceptual framework outlined above, we next turn to specific marketing tools that have been studied using a model of consumer search behavior. We note that the search model framework is very general and can be used to analyze any variable that shifts search costs. Below we focus on the most commonly studied marketing tools.

4.1 Rankings

One of the marketing tools that has received most attention are product rankings on online platforms. An early paper that analyzes rankings through the lens of a search model is Ursu (2018). The author uses data from a field experiment ran by the online travel agent Expedia that exposed a sample of consumers to hotels displayed in a randomized order. Using this experimental data, Ursu (2018) shows that rankings impact search decisions, but not conversion conditional on search. She therefore argues that rankings should enter search costs and not utility. To the best of our knowledge, most papers that study rankings now take it as given that rankings should enter search costs but not utility.

In terms of magnitude, Ursu (2018) finds that rankings have a large effect on which products consumers search and, by increasing the probability of search, also positively impact purchase probabilities. She documents that the click probability decreases from 12% to 9% between the first and second rank position and further decreases to around 4.5% at position 10. Such large ranking effects are consistent with the empirical fact that consumers tend to search relatively little and therefore are most likely to click on products that are salient in the ranking of search results. The search model also allows the author to quantify the magnitude of position effects in monetary terms. Interestingly, she finds the position effect estimate to be substantially lower than other studies that rely on non-random variation in rankings. This finding is consistent with the idea that position effects are overestimated with observational data because more popular products will tend to be ranked higher.

Given that rankings have an economically important impact on search and purchase behavior, it is natural to ask how changes in ranking algorithms affect firms' profits. One paper in this vein is Ghose et al. (2014) who show that a utility-based ranking improves revenues by 5% over the default ranking and produces higher revenues than any other considered ranking, such as ranking products by their purchase probabilities, prices, number of reviews, and other features. Other

papers have suggested ranking algorithms that optimize specific objectives. For instance, De los Santos and Koulayev (2017) propose a ranking method that optimizes click-through rates: it takes into account that, even though the intermediary typically knows very little about a consumer, the intermediary observes search refinement actions as well as other search actions, which allows it to learn consumer preferences. The authors find that their proposed ranking method almost doubles click-through rates for a hotel booking website in comparison to the website’s default ranking. These findings show that, because rankings affect consumers’ search and purchase behavior, optimizing rankings can have a large impact on firms’ revenue and profits.

Apart from affecting firms’ profits, the impact of ranking algorithms on consumers has been of interest to policy-makers. One potential worry of regulators is that firms might rank high-margin products very highly because they generate profits for the firm, but such a ranking could decrease consumer welfare. Several recent papers come to different conclusions about the alignment of firm incentives and consumer welfare. Greminger (2022) estimates a model of discovery and search based on the framework in Greminger (2021). He then uses the estimates of the model to derive the revenue-maximizing ranking and shows that such a ranking also leads to an increase in consumer welfare relative to the current status quo ranking. Compiani et al. (2023) address the same research questions using a different modeling approach that allows preferences and search costs to vary flexibly across products. Their empirical results show that ranking products that have a high purchase value but low search value (products that are so-called “diamonds in the rough”) can maximize consumer welfare, whereas a profit-maximizing strategy would require the platform to rank high-margin products higher. The paper argues that rankings that maximize welfare and profits are therefore in tension with each other. While Greminger (2022) and Compiani et al. (2023) investigate the potential tension between firms’ profit maximization and consumer welfare for uniform ranking algorithms, Donnelly et al. (2023) study personalized vs. uniform best-seller ranking algorithms. They find that the personalized ranking implemented by the platform they study increases both consumer welfare and firm’s profits compared to a uniform ranking, suggesting that firm incentives are aligned with consumer welfare in their setting.

To summarize, the three papers end up with conflicting findings with regards to the alignment or tension of welfare and profit/revenue-maximizing rankings. Given the policy relevance of this question, it will be important to reconcile these findings in future work as well as to characterize the conditions under which consumer welfare will decrease or increase when firms choose (personalized) rankings to maximize profits.

4.2 Store & Webpage Design

Rankings are only one decision that online retailers make to increase the saliency of some products. Retailers have many other tools at their disposal that affect the ease with which consumers can find certain products. In the online setting, retailers can influence how consumers navigate a webpage through tools such as rankings, recommendations, and product endorsements/badges. In brick-and-mortar stores, retailers decide where to place products and whether to highlight them with

on-shelf displays. Most of these marketing tools are best thought of as altering consumer search costs because they affect the saliency of products, but, arguably, do not directly affect consumers’ preferences. Therefore, many of the insights from our discussion of rankings carry over to the analysis of other marketing tools that shift search costs. We singled out rankings because they have been studied more extensively, likely due to better data availability. The literature on other marketing tools and their impact on search behavior is more sparse. Below, we highlight a series of papers that analyze these firm decisions through the lens of a search model.

Product recommendations, e.g., displayed at the bottom of a product detail page, are a common feature in online retail. The literature on the effects of product recommendations on consumer searches and purchases is sparse partly because researchers typically do not observe which products a retailer recommended and/or which recommendations a consumer saw. Moreover, retailers usually recommend products that have a high probability of being relevant to the consumer. This correlation between recommendations and preferences generates an endogeneity problem. Kim et al. (2011) document that product recommendations are associated with products that are more frequently searched together. However, their data do not contain random variation in recommendations. Korganbekova and Zuber (2023) observe recommendations and run a randomized experiment that removes some recommendations from product pages. Although the primary focus of the paper is on personalized rankings rather than recommendations, it is the only paper containing experimental variation in recommendations that we are aware of.

Badges are another marketing tool commonly used by online retailers. For example, Amazon highlights specific products on a search results page using badges such as “Amazon’s Choice” or “Best Seller.” Bairathi et al. (2023) study how badges affect consumer search and purchase decisions and find that the highlighted product benefits from the badge. Bairathi et al. (2023) also analyze spill-over effects (a topic we cover in more detail in the context of advertising) and find that products that are spatially close on the screen to the highlighted product experience a decrease in searches, whereas products located further away benefit from the introduction of badges. Apart from Bairathi et al. (2023), there is a broader literature on badges and quality certification (e.g., Elfenbein et al. 2015, Hui et al. 2016) that does not use search data.

A small number of papers investigates the impact of different webpage design choices. Gardete and Hunter (2020) estimate a search model in which consumers search over alternatives with multiple characteristics. A unique aspect of their click-stream data from the website of a used car seller is that information on different characteristics is shown on different pages enabling them to observe which specific characteristics a consumer searched. In counterfactual exercises, Gardete and Hunter (2020) predict the effects of different website designs. The authors find different information design policies to have moderate effects on consumers, but a sizeable impact on seller profits. Zhang et al. (2023) estimate a model of discovery and search based on Greminger (2021) using data from an apparel retailer’s mobile app store. They then use their estimates to study the effects of counterfactual store changes. For example, the authors find that having the option of only looking at items which are on sale is valued by consumers: it provides incremental sales representing about

4% of revenue. Gu and Wang (2022) investigate the question of how much information to reveal at different layers of a website, i.e., whether to display price in search results or only after clicking through to the product detail page. The authors find that revealing prices in the outer layer leads to higher welfare, while not revealing prices increases revenue.

In the context of brick-and-mortar retail, we are aware of only one study that analyzes the impact of store design decisions on consumer search and purchase behavior. Using a unique data set containing nine layout changes in a brick-and-mortar grocery store and sales from two control stores, Ranjan et al. (2017) document that layout changes significantly affect sales. The authors formulate a model of attention/consideration, purchase, and learning and find that the location of a category significantly affects whether a consumer considers it. Furthermore, Ranjan et al. (2017) also note the existence of attention spillover effects: being located adjacent to a popular category increases the consideration probability for the focal category.

In summary, while the literature on the impact of other marketing tools on search is conceptually similar to studying the impact of rankings via a search model, it is relatively scarce. To a large extent, this is due to issues of data availability and a lack of random variation in the marketing tools. For example, recommendations are rarely observed in standard clickstream data and, to the best of our knowledge, only Korganbekova and Zuber (2023) base their analysis on experimental variation in recommendations. Similarly, observing product placement as well as variation in placement in a physical store is rare. To make progress on the analysis of the various store / webpage design tools described above, better data, ideally containing random variation in the specific marketing variable, needs to be collected. There is no particular conceptual obstacle to studying these questions and the analysis of rankings using a search framework can be directly transposed onto the study of other variables.

4.3 Advertising

Advertising is one of the most commonly employed marketing tools and a large literature studies the impact of advertising on consumers' purchases. In this section, we focus on the impact of advertising along the purchase funnel, which becomes observable with pre-purchase data. For example, in the context of online retail, search data permits the researcher to observe whether advertising affects a consumer's clicking decision and/or a consumer's purchase decision conditional on clicking. Our discussion of advertising differs somewhat from the above discussion of other marketing variables. Marketing tools such as rankings or recommendations are usually modeled as variables that alter search costs. However, in the case of advertising, it is not a priori clear how it affects consumer behavior. It is conceivable that advertising lowers search costs by making the advertised product more salient or it might directly increase the utility from purchasing the advertised product. Several studies measuring the impact of advertising also include an awareness stage in which advertising might affect whether consumers know of the existence of a product.

Similar to the other marketing variables described above, endogeneity concerns also arise in the case of advertising. These concerns have been addressed by using random variation in advertising

(e.g., Morozov and Tuchman 2023), by using quasi-experimental approaches in a reduced-way approach (e.g., Tsai and Honka 2021), or by using a control function approach inside a structural search model that exploits variation in arguably exogenous variables that impact advertising (e.g., Honka et al. 2017).

Honka et al. (2017) quantify the effects of advertising along the purchase funnel for savings accounts. They use variation in advertising costs to address the endogeneity of advertising exposure and find that advertising has a large indirect effect on choice via increasing awareness and only a small direct effect on choices conditional on awareness. Tsai and Honka (2021) study how advertising affects consumers along the auto insurance purchase funnel. The authors account for advertising endogeneity using variation in advertising across media market borders (Shapiro 2018).¹⁴ They find that advertising increases awareness, but has no significant effects on searches conditional on awareness, suggesting that advertising primarily affects the awareness stage of the purchase funnel. Morozov and Tuchman (2023) set up an online e-book store to run field experiments and find that advertising affects both search and purchase decisions with the effect on searches being larger. In the context of feature advertising for a supermarket, Seiler and Yao (2017) show that advertising leads to higher sales, but does not increase traffic to the advertised category. Contrary to the various studies of online retail, it therefore appears that advertising in brick-and-mortar stores only affects consumers at the purchase stage of the conversion funnel.¹⁵

Advertising might lead to spillover effects depending on where in the purchase funnel advertising affects behavior. Advertising that increases awareness for a product could, for instance, increase awareness for other similar products. Sahni (2016) shows advertising spillovers on an online restaurant-search website between restaurants that belong to the same category/cuisine. While he does not estimate a structural search model, his analysis is motivated by a model in which advertising affects awareness of products with particular characteristics.¹⁶ In the context of a Weitzman (1979)-style model, it would also be possible to model spillovers in advertising to lower search costs not only for the focal product but also for other similar products.¹⁷ To the best of our knowledge, no structural search paper has yet modeled spillovers in this fashion.

Several papers have examined spillovers (or their absence) using a reduced-form approach. Seiler and Yao (2017) do not find evidence of advertising spillovers to other categories in close physical proximity in a brick-and-mortar grocery store environment. Within a product category, the authors only find spillovers between products that belong to the same brand. They hypothesize

¹⁴Shapiro (2018) first proposed the use of the border strategy to address advertising endogeneity concerns. It exploits the discontinuity in TV advertising along the borders of television markets (DMAs). The border strategy rests on the assumption that consumers on both sides of the border of a DMA are similar, but are exposed to different TV advertising. The advertising effects are then identified by the differences in how consumers on both sides of a border react to differences in advertising.

¹⁵In a related paper, Ursu et al. (2023b) model how online display ads can expand consumers' awareness sets before and during the search process.

¹⁶Similarly, Liang et al. (2019) document advertising spillover between products with similar characteristics in the market for mobile apps.

¹⁷If advertising leads to some form of learning across products, the independence assumption regarding the error revealed after search would be violated and the Weitzman (1979) framework can no longer be used.

that spillovers in a physical store might be limited because consumers' paths through a store are difficult to influence through advertising. In the online context, the findings are mixed with Sahni (2016) and Liang et al. (2019) documenting spillover effects between similar products, while Fong (2017) shows that targeted ads lead to more searches of the advertised product at the expense of searches for non-advertised similar products in the case of an online wine retailer. Morozov and Tuchman (2023) also find no evidence of advertising spillovers to similar products in an online bookstore.

A small set of recent papers have started to investigate the effects of different advertising content on consumer search decisions. Tsai and Honka (2021) show that informational advertising increases aided awareness, while non-informational advertising increases unaided awareness in the car insurance market. Morozov and Tuchman (2023) analyze three types of ad content for e-books: plain ads with no information, ads with information about a horizontal characteristic (book genre) and ads with information about a vertical characteristic (price). The authors find that price advertising increases search among all consumers. However, genre advertising induces some consumers to search the advertised book and others to reject it without search. These patterns suggest that advertising affects consumer search behavior differently depending on the content of the ad.

In summary, most evidence points towards advertising having a large impact on consumers' awareness of products and less of an influence on choices conditional on awareness. In terms of behavioral mechanism, this pattern suggests an informative rather than a persuasive role of advertising in many settings. The one notable exception is Seiler and Yao (2017) who document an impact of advertising only at the purchase stage of the conversion funnel, which might point to a difference in how advertising affects consumers in an online versus offline setting. The evidence on spillovers effects is more mixed with only a subset of papers findings evidence for spillover effects to similar products. The existing studies do not point to a clear pattern with regard to the circumstances under which spillovers arise.

5 How Search Frictions Affect Equilibrium Outcomes

In the preceding section, we discussed several advances made by prior work in trying to understand how firms' marketing actions impact consumer search and purchase decisions in a partial equilibrium fashion, i.e., without an explicit model of how firms optimally choose these marketing variables. To understand how markets function in the presence of search frictions, we also need to consider how firms optimally behave when facing consumers who engage in costly search. In this section, we therefore turn to outcomes of firm-consumer interactions when both firms' and consumers' decisions are jointly determined in equilibrium.

In what follows, we describe the theoretical results pertaining to equilibrium prices, store design, and advertising and discuss relevant empirical applications. In general, most empirical work on consumer search has focused on the demand side and the empirical literature that studies equilibrium

outcomes is sparse.

5.1 Equilibrium Pricing

Equilibrium price setting in the presence of search costs can be characterized by extreme outcomes, as illustrated by the well-known “Diamond Paradox” (due to Diamond 1971). In the Diamond (1971) model, consumers have unit demand, a large number of firms sells a homogeneous product, and consumers search to learn the prices firms charge. Each consumer starts at a firm and must pay a search cost to visit another firm to learn the price it charges. In this framework, the market completely collapses in equilibrium: the equilibrium price in the market equals the monopoly price, and there is therefore no price dispersion. Without price dispersion, there is no need for consumers to search.

In contrast to the stark predictions from the Diamond (1971) model, an abundance of data has shown that prices, even for homogeneous products, differ across firms, i.e., that there is price dispersion (Stigler 1961; Sorensen 2000; Hortaçsu and Syverson 2004; Hong and Shum 2006; Hitsch et al. 2021), and that consumers frequently search more than one firm. Many theoretical search models therefore explore ways in which the Diamond Paradox can be avoided and price dispersion is generated in equilibrium.

Subsequent work has proposed two main mechanisms that lead to price dispersion as an equilibrium outcome: consumer and firm heterogeneity. Stahl (1989) and Stahl (1996) show that consumer search cost heterogeneity can lead to price dispersion. In these models, two groups of consumers exist: “shoppers” who have negative search costs, i.e., enjoy searching, and “non-shoppers” with positive search costs. In this framework, price dispersion arises as firms employ mixed strategies in prices to cater to both groups of consumers. Similarly, in Salop and Stiglitz (1977), there are two types of consumers: “informed” consumers, who know the distribution of prices and can always identify the lowest-priced firm, and “uninformed” consumers who know nothing. Similar to the Stahl (1989) and Stahl (1996) models, in equilibrium, different firms sell to different types of consumers, leading to price dispersion in the market. In comparison, Reinganum (1979) and MacMinn (1980) investigate how firm production cost heterogeneity can lead to price dispersion. For instance, Reinganum (1979) proves the existence of price dispersion in equilibrium when firms have different marginal costs and consumers have elastic demand.

An important counterfactual, that is of interest to researchers, policy-makers, and managers, is whether a decrease in search costs leads to an increase or a decrease in equilibrium prices. For example, the emergence of online price comparison websites can be conceived as lowering consumers’ search costs. The literature has derived ambiguous results regarding the relation between equilibrium price levels and search costs. On the one hand, in Diamond (1971), Stahl (1989), and Stahl (1996), prices can increase in search costs. In Diamond (1971), this happens because the introduction of even an infinitesimal search cost results in prices jumping from perfectly competitive levels to the monopoly level. In Stahl (1989) and Stahl (1996), prices can increase in search costs because of the existence of two types of consumers (“shoppers” and “non-shoppers”), who differ in

their search costs: as the search costs of non-shoppers decreases, prices decrease and converge to marginal cost.

On the other hand, more recent work shows that equilibrium prices can also decrease in search costs (Armstrong and Zhou 2011; Choi et al. 2018; Ding and Zhang 2018; Garcia and Shelegia 2018; Haan et al. 2018). This type of result generally arises when products are differentiated, consumers search to learn about an attribute other than price, and consumers search options sequentially. For example, in Haan et al. (2018), two firms sell a horizontally differentiated product with two attributes: one that is observable without search and another one for which the consumer pays a search cost to learn about. In addition, products differ in their prices and firms can make these prices observable without search, thereby directing the search process. The authors show that, when prices are observable, a higher search cost implies more competition between firms to be searched first by a consumer and therefore lower prices. In other words, firms are willing to lower prices more when search costs are higher because they know that consumers will then be less likely to visit the other firm.

Overall, the theoretical literature suggests that settings in which consumers search over prices lead to equilibrium prices that increase in search costs. However, when prices are observable prior to search, a decrease in search costs can lead to higher prices in equilibrium. In many online settings, the latter case appears to be more relevant. For example, on the typical online retail webpage, prices are observable on the search results page, but a consumer needs to visit a product detail page (i.e., search) to gather information about other product attributes. This result is important for managers and policy-makers because it shows that consumers may be hurt by a decrease in search costs.

Turning to empirical work, a small set of papers has studied the equilibrium relationship between price levels and search costs. In line with the ambiguous predictions from the theory literature, empirical papers have found evidence for both a positive and a negative relation between price levels and search costs. For example, Brown and Goolsbee (2002) look at the impact of the Internet on the price of life insurance in the 1990s. They posit that the Internet has reduced search costs by enabling consumers to more easily compare products. The authors find that the Internet has led to a reduction in prices by 8–15%. Studying the new car market in The Netherlands, Moraga-González et al. (2023) show that reducing search costs, for example by allowing at-home test drives, results in lower prices. In contrast, in settings in which prices can be more easily observed prior to searching, empirical work generally finds that lower search costs lead to higher prices. For instance, Wildenbeest (2011) shows that higher search intensity can lead to higher prices when studying supermarkets in the UK. The empirical literature therefore supports the key distinction between prices being observable before or only after search as the main determinant for the direction of the relationship between search costs and equilibrium prices.¹⁸

¹⁸Only Brown and Goolsbee (2002) directly estimate the impact of an exogenous decrease in search costs on equilibrium prices. The other papers cited in the paragraph estimate models of consumer search and supply-side price setting and then assess the impact of search cost changes in counterfactuals.

5.2 Other Marketing Tools

To the best of our knowledge, empirical papers studying equilibrium outcomes solely focus on how search frictions affect equilibrium prices. In principle, all marketing tools that we discussed in Section 4 could also be studied in a general equilibrium framework. However, this has not happened because of a lack of good data and the difficulty of modeling supply-side behavior. For example, the theoretical literature on equilibrium advertising outcomes often considers specific forms of advertising (such as price advertising or advertising that solely raises awareness) as well as advertising content decisions by firms. As we outlined in Section 4.3, the empirical work using advertising content is very limited. Moreover, modeling advertising decisions on the supply side is more complex than modeling price setting behavior. Similarly, empirical work on how store and webpage design decisions affect consumer behavior is relatively sparse and there are few theoretical papers that tackle this issue and could be used to as a modeling framework for empirical work. Due to the lack of empirical work, we confine ourselves to a brief overview of theoretical work in the areas of equilibrium store design and advertising decisions some of which could provide a template for future empirical work.

In the context of store design, Petrikaitė (2018) studies how a multi-product firm should affect consumer search cost for different products to maximize profits. The author shows that the firm has an incentive to raise the price of the earlier searched product (the one with lower search costs) since consumers are more likely to search higher-valued products first. In other words, Petrikaitė (2018)’s findings suggest that it is optimal for the retailer to increase prices of products easily found in a store. Ursu and Dzyabura (2020) study a similar setting but ask a different question: where should a retailer locate products to maximize revenue? The authors find that the retailer will make lower-valuation products more prominent and easier to find, since products with higher valuations will be searched even at higher search costs.

Turning to advertising, theoretical work either examines settings in which advertising is the only source of information for consumers (Iyer et al. 2005) or settings in which consumers can decide whether to search after receiving ads from a firm in a first stage (Butters 1977; Robert and Stahl 1993; Anderson and Renault 2006, 2013; Haan and Moraga-González 2011; Mayzlin and Shin 2011; Shin and Yu 2021). This work can be further divided into papers that only consider price advertising (which subsumes advertising about the existence of a firm) and papers that include advertising containing both price and other content. For example, in Butters (1977) and Robert and Stahl (1993), firms can send ads revealing the product price to consumers in a first stage, and consumers can search in a second stage (including those consumers who did not receive any ads). In Butters (1977), search occurs only if consumers receive no ads. In Robert and Stahl (1993), firms charge either high or low prices, and those who charge high prices only sell to consumers who are uninformed.

Anderson and Renault (2006) is one of a few papers that study equilibrium advertising content decisions in search markets. In the paper, a monopolist has to decide whether to advertise price and/or match value information to consumers before these decide whether to search for the

remaining information or to purchase the item when all information has been revealed through the ad. The authors show that it is optimal for the firm to only minimally advertise the match value or, alternatively, to advertise prices and partial match information. When prices are advertised, they decrease in search costs, i.e., advertised prices are lower for higher search cost levels. Mayzlin and Shin (2011) derive similar results but show a new mechanism: firms prefer to withhold information in order to encourage consumers to search. In a follow-up paper, Anderson and Renault (2013) investigate the optimal mix of advertising content (quality, price, and horizontal match) for a search good. The authors find that lower-quality firms need to provide more information than higher-quality firms. Furthermore, for a given quality level, quality information is revealed first, followed by price information and then by horizontal match information. Taken together, the findings from these papers indicate that it might be optimal for firms operating in search markets to employ advertising that only partially reveals information about products.

6 Conclusion

Our overview takes two stylized facts about search behavior as a starting point: consumers typically search relatively little and, in many settings, there is substantially more search than purchase data. Based on these empirical patterns, we highlight three research areas for which search data can be particularly valuable. We argue that the abundance of search data relative to purchase data can help researchers better estimate consumer preferences, which, in turn, can improve optimal marketing strategies such as targeted marketing campaigns. Second, we show that a search framework coupled with data on consumer search behavior can be used to study the impact of a host of variables that shift search costs such as product rankings, recommendations, product placement and advertising. Because consumers search few products, the impact of some marketing variables on search and, by extension, on purchase outcomes can be large. There has been a recent surge in empirical work on some marketing variables such as rankings; however, many other variables that can shift search behavior are still relatively understudied due to a lack of good data. Finally, we show how search data can be used to understand how information frictions can affect market outcomes. While there is a small set of papers studying the impact of search costs on prices, there is little to no empirical work on equilibrium outcomes for other marketing variables so far.

References

- ALLENBY, G. M. AND P. E. ROSSI (1999): “Marketing Models of Consumer Heterogeneity,” *Journal of Econometrics*, 89, 57–78.
- AMANO, T., A. RHODES, AND S. SEILER (2022): “Flexible Demand Estimation with Search Data,” *CEPR Discussion Paper No. DP16933*.
- ANDERSON, S. P. AND R. RENAULT (2006): “Advertising Content,” *American Economic Review*, 96, 93–113.
- (2013): “The Advertising Mix for a Search Good,” *Management Science*, 59, 69–83.
- ARMONA, L., G. LEWIS, AND G. ZERVAS (2021): “Learning Product Characteristics and Consumer Preferences from Search Data,” Working Paper.
- ARMSTRONG, M. AND J. ZHOU (2011): “Paying for Prominence,” *The Economic Journal*, 121, F368–F395.
- BAIRATHI, M., X. ZHANG, AND A. LAMBRECHT (2023): “The Value of Platform Endorsement,” Working Paper.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (2004): “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market,” *Journal of Political Economy*, 112, 68–105.
- BRONNENBERG, B. J., J. B. KIM, AND C. F. MELA (2016): “Zooming In on Choice: How Do Consumers Search for Cameras Online?” *Marketing Science*, 35, 693–712.
- BROWN, J. R. AND A. GOOLSBEE (2002): “Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry,” *Journal of Political Economy*, 110, 481–507.
- BUTTERS, G. (1977): “Equilibrium Distributions of Prices and Advertising,” *Review of Economic Studies*, 44, 465–491.
- CHEN, Y. AND S. YAO (2017): “Sequential Search with Refinement: Model and Application with Click-stream Data,” *Management Science*, 63, 4345–4365.
- CHOI, M., A. Y. DAI, AND K. KIM (2018): “Consumer Search and Price Competition,” *Econometrica*, 86, 1257–1281.
- COMPIANI, G., G. LEWIS, S. PENG, AND W. WANG (2023): “Online Search and Optimal Product Rankings: A Empirical Framework,” *Marketing Science*, conditionally accepted.
- DE LOS SANTOS, B. AND S. KOULAYEV (2017): “Optimizing Click-Through in Online Rankings with Endogenous Search Refinement,” *Marketing Science*, 36, 542–564.
- DIAMOND, P. A. (1971): “A Model of Price Adjustment,” *Journal of Economic Theory*, 3, 156–168.
- DING, Y. AND T. ZHANG (2018): “Price-directed Consumer Search,” *International Journal of Industrial Organization*, 58, 106–135.
- DONNELLY, R., A. KANODIA, AND I. MOROZOV (2023): “Welfare Effects of Personalized Rankings,” *Marketing Science*, forthcoming.

- DUBÉ, J.-P. (2019): “Microeconometric Models of Demand,” in *Handbook on the Economics of Marketing*, ed. by J.-P. Dubé and P. E. Rossi, Elsevier.
- ELFENBEIN, D. W., R. FISMAN, AND B. MCMANUS (2015): “Market Structure, Reputation, and the Value of Quality Certification,” *American Economic Journal: Microeconomics*, 7, 83–108.
- FONG, N. M. (2017): “How Targeting Affects Customer Search: A Field Experiment,” *Management Science*, 63, 2353–2364.
- GARCIA, D. AND S. SHELEGIA (2018): “Consumer Search with Observational Learning,” *The RAND Journal of Economics*, 49, 224–253.
- GARDETE, P. AND M. HUNTER (2020): “Guiding Consumers through Lemons and Peaches: An Analysis of the Effects of Search Design Activities,” Working Paper.
- GENSCH, D. (1987): “A Two-Stage Disaggregate Attribute Choice Model,” *Marketing Science*, 6, 223–231.
- GHOSE, A., P. G. IPEIROTIS, AND B. LI (2014): “Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue,” *Management Science*, 60, 1632–1654.
- GREMINGER, R. (2021): “Optimal Search and Discovery,” *Management Science*, 68, 3904–3924.
- (2022): “Heterogeneous Position Effects and the Power of Rankings,” Working Paper.
- GU, C. AND Y. WANG (2022): “Consumer Online Search with Partially Revealed Information,” *Management Science*, 68, 4215–4235.
- HAAN, M. AND J. L. MORAGA-GONZÁLEZ (2011): “Consumer Search with Observable and Hidden Characteristics,” Working Paper.
- HAAN, M. A., J. L. MORAGA-GONZÁLEZ, AND V. PETRIKAITĖ (2018): “A Model of Directed Consumer Search,” *International Journal of Industrial Organization*, 61, 223–255.
- HAUSER, J. AND B. WERNERFELT (1990): “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research*, 16, 393–408.
- HITSCH, G., A. HORTACSU, AND X. LIN (2021): “Prices and Promotions in U.S. Retail Markets,” *Quantitative Marketing and Economics*, 19, 289–368.
- HONG, H. AND M. SHUM (2006): “Using Price Distributions to Estimate Search Costs,” *The RAND Journal of Economics*, 37, 257–275.
- HONKA, E. (2014): “Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry,” *The RAND Journal of Economics*, 45, 847–884.
- HONKA, E., A. HORTACSU, AND M. A. VITORINO (2017): “Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry,” *The RAND Journal of Economics*, 48, 611–646.
- HONKA, E., A. HORTACSU, AND M. WILDENBEEST (2019): “Empirical Search and Consideration Sets,” in *Handbook on the Economics of Marketing*, ed. by J. Dube and P. Rossi, North Holland.

- HORTAÇSU, A. AND C. SYVERSON (2004): “Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds,” *The Quarterly Journal of Economics*, 119, 403–456.
- HOWARD, J. AND J. SHETH (1969): *The Theory of Buyer Behavior*, John Wiley, NY.
- HUI, S., E. BRADLOW, AND P. FADER (2009): “The Traveling Salesman Goes Shopping: The Systematic Deviations of Grocery Paths from TSP-Optimality,” *Marketing Science*, 28, 566–572.
- HUI, X., M. SAEEDI, Z. SHEN, AND N. SUNDARESAN (2016): “Reputation and Regulations: Evidence from eBay,” *Management Science*, 62, 3604–3616.
- IYER, G., D. SOBERMAN, AND J. M. VILLAS-BOAS (2005): “The Targeting of Advertising,” *Marketing Science*, 24, 461–476.
- JAIN, A., S. MISRA, AND N. RUDI (2020): “Sales Assistance and Purchase Decisions: An Analysis Using Retail Video Data,” *Quantitative Marketing and Economics*, 18, 273–303.
- JIANG, Z., T. CHEN, H. CHE, AND Y. WANG (2021): “Consumer Search and Purchase: An Empirical Investigation of Retargeting Based on Consumer Online Behaviors,” *Marketing Science*, 40, 219–240.
- KIM, J. B., P. ALBUQUERQUE, AND B. J. BRONNENBERG (2010): “Online Demand under Limited Consumer Search,” *Marketing Science*, 29, 1001–1023.
- (2011): “Mapping Online Consumer Search,” *Journal of Marketing Research*, 48, 13–27.
- KORGANBEKOVA, M. AND C. ZUBER (2023): “Balancing User Privacy and Personalization,” Working Paper.
- KOULAYEV, S. (2014): “Search for Differentiated Products: Identification and Estimation,” *The RAND Journal of Economics*, 45, 553–575.
- LAM, H. T. (2023): “Platform Search Design and Market Power,” Working Paper.
- LIANG, C., Z. M. SHI, AND T. S. RAGHU (2019): “The Spillover of Spotlight: Platform Recommendation in the Mobile App Market,” *Information Systems Research*, 30, 1296–1318.
- LIU, J. AND O. TOUBIA (2018): “A Semantic Approach for Estimating Consumer Content Preferences from Online Search Queries,” *Marketing Science*, 37, 930–952.
- (2020): “Search Query Formation by Strategic Consumers,” *Quantitative Marketing and Economics*, 18, 155–194.
- LIU, J., O. TOUBIA, AND S. HILL (2021): “Content-Based Model of Web Search Behavior: An Application to TV Show Search,” *Management Science*, 67, 6378–6398.
- MACMINN, R. D. (1980): “Search and Market Equilibrium,” *Journal of Political Economy*, 88, 308–327.
- MARTINOVICI, A., F. PIETERS, AND T. ERDEM (2021): “Attention Trajectories Predict Brand Choice,” Working Paper.
- MAYZLIN, D. AND J. SHIN (2011): “Uninformative Advertising as an Invitation to Search,” *Marketing Science*, 30, 666–685.

- MORAGA-GONZÁLEZ, J. L., Z. SÁNDOR, AND M. R. WILDENBEEST (2023): “Consumer Search and Prices in the Automobile Market,” *The Review of Economic Studies*, 90, 1394–1440.
- MOROZOV, I., S. SEILER, X. DONG, AND L. HOU (2021): “Estimation of Preference Heterogeneity in Markets with Costly Search,” *Marketing Science*, 40, 871–899.
- MOROZOV, I. AND A. TUCHMAN (2023): “Where Does Advertising Content Lead You? We Created a Bookstore to Find Out,” Working Paper.
- PADILLA, N., E. ASCARZA, AND O. NETZER (2019): “The Customer Journey as a Source of Information,” Working Paper.
- PETRIKAITE, V. (2018): “Consumer Obfuscation by a Multiproduct Firm,” *The RAND Journal of Economics*, 49, 206–223.
- RANJAN, B., P. ELLICKSON, AND M. LOVETT (2017): “Effect of Location and Assortment on Category Consideration, Learning, and Choice,” Working Paper.
- REINGANUM, J. F. (1979): “A Simple Model of Equilibrium Price Dispersion,” *Journal of Political Economy*, 87, 851–858.
- RINGEL, D. M. AND B. SKIERA (2016): “Visualizing Asymmetric Competition among more than 1,000 Products Using Big Search Data,” *Marketing Science*, 35, 511–534.
- ROBERT, J. AND D. O. STAHL (1993): “Informative Price Advertising in a Sequential Search Model,” *Econometrica*, 657–686.
- ROBERTS, J. AND J. LATTIN (1991): “Development and Testing of a Model of Consideration Set Composition,” *Journal of Marketing Research*, 28, 429–440.
- ROSSI, P. E. AND G. M. ALLENBY (1993): “A Bayesian Approach to Estimating Household Parameters,” *Journal of Marketing Research*, 30, 171–182.
- SAHNI, N. S. (2016): “Advertising Spillovers: Evidence from Online Field Experiments and Implications for Returns on Advertising,” *Journal of Marketing Research*, 53, 459–478.
- SALOP, S. AND J. STIGLITZ (1977): “Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion,” *The Review of Economic Studies*, 44, 493–510.
- SEILER, S. AND F. PINNA (2017): “Estimating Search Benefits from Path-Tracking Data: Measurement and Determinants,” *Marketing Science*, 36, 565–589.
- SEILER, S. AND S. YAO (2017): “The Impact of Advertising along the Conversion Funnel,” *Quantitative Marketing and Economics*, 15, 241–278.
- SHAPIRO, B. (2018): “Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants,” *Journal of Political Economy*, 126, 381–437.
- SHIN, J. AND J. YU (2021): “Targeted Advertising and Consumer Inference,” *Marketing Science*, 40, 900–922.
- SHOCKER, A., M. BEN-AKIVA, B. BOCCARA, AND P. NEDUNGADI (1991): “Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions,” *Marketing Letters*, 181–197.

- SORENSEN, A. T. (2000): “Equilibrium Price Dispersion in Retail Markets for Prescription Drugs,” *Journal of Political Economy*, 108, 833–850.
- STAHL, D. O. (1989): “Oligopolistic Pricing with Sequential Consumer Search,” *American Economic Review*, 700–712.
- (1996): “Oligopolistic Pricing with Heterogeneous Consumer Search,” *International Journal of Industrial Organization*, 14, 243–268.
- STIGLER, G. (1961): “The Economics of Information,” *Journal of Political Economy*, 69, 213–225.
- TSAI, Y.-L. AND E. HONKA (2021): “Informational and Noninformational Advertising Content,” *Marketing Science*, 40, 1030–1058.
- URSU, R., T. ERDEM, Q. WANG, AND Q. P. ZHANG (2022): “Prior Information and Consumer Search: Evidence from Eye-tracking,” *Management Science*, forthcoming.
- URSU, R., S. SEILER, AND E. HONKA (2023a): “The Sequential Search Model: A Framework for Empirical Research,” Working Paper.
- URSU, R., A. SIMONOV, AND E. AN (2023b): “Online Advertising as Passive Search,” *Management Science*, forthcoming.
- URSU, R., Q. WANG, AND P. CHINTAGUNTA (2020): “Search Duration,” *Marketing Science*, 39, 849–871.
- URSU, R., Q. ZHANG, AND E. HONKA (2023c): “Search Gaps and Consumer Fatigue,” *Marketing Science*, 42, 110–136.
- URSU, R. M. (2018): “The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions,” *Marketing Science*, 37, 530–552.
- URSU, R. M. AND D. DZYABURA (2020): “Retailers’ Product Location Problem with Consumer Search,” *Quantitative Marketing and Economics*, 18, 125–154.
- WEITZMAN, M. L. (1979): “Optimal Search for the Best Alternative,” *Econometrica*, 47, 641–654.
- WILDENBEEST, M. R. (2011): “An Empirical Model of Search with Vertically Differentiated Products,” *The RAND Journal of Economics*, 42, 729–757.
- XU, J., B. BOLLINGER, R. URSU, AND G. J. FITZSIMONS (2023): “The Impact of Product Location Changes on Habits, Search, and Purchase Decisions,” Working Paper.
- YAO, S., W. WANG, AND Y. CHEN (2017): “TV Channel Search and Commercial Breaks,” *Journal of Marketing Research*, 54, 671–686.
- 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, 1–52.
- ZHANG, L., R. URSU, E. HONKA, AND O. YAO (2023): “Product Discovery and Consumer Search Routes: Evidence from a Mobile App,” Working Paper.