Customer Search and Product Returns

By

Marat Ibragimov

B.S, Applied Physics and Mathematics
Moscow Institute of Physics and Technology, 2015
M.S, Applied Physics and Mathematics
Moscow Institute of Physics and Technology, 2017
M.A, Economics
New Economics School, 2018

SUBMITTED TO THE DEPARTMENT OF MANAGEMENT IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2022

Signature of Author: ____
Department of Management
May 3, 2022

Certified by: ____
John R. Hauser
Kirin Professor of Marketing
Professor, Marketing
Thesis Supervisor

Accepted by: ____
Catherine Tucker
Sloan Distinguished Professor of Management
Professor, Marketing
Faculty Chair, MIT Sloan PhD Program

Customer Search and Product Returns

by
Marat Ibragimov

Submitted to the MIT Sloan School of Management on May 3, 2022
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Management Research

Abstract

Online retailers are challenged by frequent product returns. High return rates significantly decrease companies' profit which makes the issue of managing product returns very important from the practical standpoint. Typically, practitioners study returns in connection with purchase decisions or as a part of customer behavior/type. In this paper, we show that the events which precede the purchase decision are related to the return decision. Generally, this information is readily available to online retailers and thus provides a low-cost opportunity to better understand and predict the product returns.

Based on the data provided by a large apparel retailer, we demonstrate that the way customers search for a product is indicative of product returns. We find correlational evidence that using search filters, spending more time, and purchasing the last item searched are negatively associated with the probability of return. We propose a joint model of search, purchase and return which is based on an analytic model of search, purchase, and returns. Our model is consistent with the findings in the data and provides insight into how search and returns are related. Finally, using a machine learning framework, we demonstrate that adding search data improves the prediction accuracy of individual-level return rate above and beyond prior models.

Thesis Supervisor: John R. Hauser
Title: Kirin Professor of Marketing, MIT Sloan School of Management

1. Introduction

Nelly discovered that she needed a new pair of her favorite jeans. She visited the website of a familiar woman's fashion retailer and searched the new pair. Knowing what she wanted, she used pre-search filters, carefully reviewed the product page, and purchased a pair of jeans. When she receives her jeans, she is happy with the purchase and sees no reason to return them. Wendy wants to purchase something for the coming fashion season. She visits the same fashion website and starts searching because the website is new to her. She views many items without spending much time on one item. At the end of her search, she purchases the same pair of fashion jeans that Nelly purchased. The website has a favorable return policy, so even though she is not sure whether the jeans will be a good addition to her wardrobe, the purchase carries little risk. Upon receiving the jeans, she decides they are not for her. She returns them for a full refund.

This simple example illustrates that the way the customer search could be informative of the product return decision. Nelly knows what she wants and uses the website to narrow her search and find the best alternative. Wendy, on the other hand, searches many items, takes a chance, and uses the return policy as a way to fully evaluate whether the jeans finds in the portfolio that is her wardrobe. Nelly's and Wendy's search patterns are but examples, and neither will predict fully their return patterns, but it is possible (and a testable proposition) that items bought after a search similar to Nelly's are less likely to be returned than items sought after a search similar to Wendy's. This is the key concept which we hope to investigate.

If search provides information about product returns, then, from a retailer's perspective, even though Nelly and Wendy are indistinguishable based on purchase outcomes, their return rates might differ. Because returns are extremely costly to the firm, such information enables the firm to enact strategies and tactics to tilt purchasing toward customers such as Nelly and away from customers such as Wendy.

Because the return rate can easily reach 15-35% in online retail, the increase in profit could be dramatic (Ofek, Katona, & Sarvary, 2011). The Economist (2013) estimates the costs range between $6 and $20 per returned item suggesting that returns could substantially decrease the profitability of products. The founder of UK's largest fashion and cosmetic retailer, ASOS, stated that 1% decrease in return rate would lead to 30% increase in company's net income. In some cases, retailers ban some customers from the using the website because of frequent returns (Wall Street Journal 2018).

An accurate model of search and product returns could open new opportunities to the firm. Improved prediction accuracy on individual level return probability could help the firm better plan the backward logistics when the returns are unavoidable. Alternatively, the firm could adapt its deals and special offers based on observed search patterns, at least until customers learn that certain search patterns lead to better offers.

A formal empirically-grounded model could help to better understand the phenomenon of product returns and provide new insights. The firm may be interested in mechanism of how search and returns are related. The greater insight could lead to creative policies affecting the customer search journey. The firm may wish to plan A/B experiments to enhance its understanding and/or test creative policies.

Search-to-returns insight might improve recommendation systems used by retailers or third parties. Typically, recommendation system to maximize the individual-level probability of purchase. However, the item with the highest purchase probability may not be most profitable if it is more likely to be returned. A retailer's recommendation system might recommend a different product. Even a third parties recommendation system may seek to avoid the negative affect due to product returns.

Lastly, the model could provide insights on how to morph the website to better manage product returns. These changes could be relatively low-cost in comparison to changes in the return policy. Moreover, in countries with strong customer protection legislation, it is impossible to change the return policy but, generally, companies are not limited in the way they design the website.

In the paper, we explore whether search information is related to product returns. We begin by identifying model-free stylized facts from data that link a customer's search, purchase, and return. We propose a model that provides one explanation that is consistent with the empirical findings. Our theoretical framework is based on a rational customer and demonstrates why search and product returns are related. Finally, we estimate a simple model based on the insights from model and evaluate the potential to use search information to predict product returns.

2. Literature

This paper contributes to two streams of marketing research: managing product returns and customer search.

2.1. Product returns

Product returns are studied in both theoretical and empirical frameworks. Theoretically, authors demonstrate that option to return products could serve as risk-reducing instrument to experience the product (Che, 1996) or as a signal of item quality (Moorthy & Srinivasan, 1995).

The applied stream of research focuses on optimization of return policies by firms. Authors recognize the tradeoff between higher demand and higher return rates when firms use lenient policies and suggest that optimal return policy must be balanced (e.g., Davis et al. 1998; Bower and Maxham-III 2012). Anderson et al. (2009) propose a structural model where the option to return is embedded in customer purchase decision.

When product returns are unavoidable, authors focus on optimization of backward logistics (Guide, Daniel, & Luk, 2009) and predictive models of product returns. For example, Hess and Mayhew (1997) develop a hazard model to predict the probability and expected time of product return. Dzyabura et al. (2020) demonstrate that product images could be informative of product level return rate and propose a machine learning model which helps retailers decide which products to launch online.

Another stream of research studies the return behavior of individual customers. For example, Bechwati and Siegal (2005) propose a cognitive mechanism underlying product returns. Narang and Shankar (2019) demonstrate that app adopters tend to both purchase and return more. Other empirical studies demonstrate that variety of factors affects the probability of product returns including price, being on sale, shipping fees, truthfulness of product reviews, or even the weather. (e.g., Conlin et al. 2007; Petersen and Kumar 2009, 2010; Sahoo et al. 2018; Shehu et al. 2020)

We contribute to the returns literature by introducing a new source of information which could be relevant for practitioners. Specifically, we demonstrate that customer search is indicative of product returns and data on consumer search directly improves prediction accuracy. Moreover, search data helps firms to better understand why customers return products, thus opening new strategies for managers.

2.2. Customer search

Customer search is a mature field of research both empirically and theoretically. The literature typically follows either sequential (Weitzman 1979) or simultaneous (Stigler 1961) approaches. In both approaches, the customer knows the distribution of the rewards and searches to fully reveal the uncertainty. More recent papers extend these models by adding learning (e.g., Ke et al. 2016; Branco et al. 2012); multiple attributes (Kim et al. 2010); intermediaries (Dukes and Liu 2015), or account for duration of search (Ursu et al. 2020).

Our research differs from existing papers as we look beyond the purchase decision and treat the return decision as an important part of the customer's journey.

3. Stylized facts

We obtained data from a large apparel retailer in Western Europe. The retailer has both online and offline channels and distributes medium-price fashion products for women, men, and children. The share of children's apparel is approximately 5% suggesting that the company specializes in adults' products. The retailer has a very generous return policy, where each item could be returned within 60 days after the purchase without monetary losses to the customer.[1]

We base our analysis on the online channel because it is the only feasible way to gather detailed search data. Our data include complete search and transaction history from October 2019 to mid-May 2020. For each browsing (search) session we observe the unique identifier which could be matched to the transaction if the customer purchased something during this session. For each product in the transaction, we observe whether the item was returned by the customer and date of the return. In total our data contain over 4.5 million browsing sessions with 330 thousand ended with purchase of at least one product. For each product in the assortment of searched items, our data include the characteristics of the product such as category, material, size, brand, and color.

The search data record a sequence of actions made by the customer; the data are highly multi-dimensional. To identify model-free stylized facts we use these data in a simple regression framework. Our variables are aggregated statistics of the customer search. All variables are many-to-one mappings from the search to numerical data. Aggregation leads to a significant loss of information, but aggregation enables us to generate interpretable results and build intuition.

[1] The firm operates in country with strong customer protection legislation requiring the firm to have generous return policies.

As an initial pass on the data, we consider the subset of the data with the orders having single product purchased. Our reason is to isolate purchase and search relations with return propensity. By focusing on simple-product purchases, we abstract away from instances where the customer is strategic and purchases several colors to retain but one. Although customer search might also indicate such behavior, such considerations are left to future research and might confound to the goals of the current research. However, for robustness, we check whether the identified stylized facts apply to multiple-product orders.

We focus on four variables related to the search and provide descriptive data in Table 1:

1. *Last item searched* – dummy variable indicating whether this product was the last customer looked at
2. *Number product searched* – number of unique products searched[2]
3. *Applied filter* – dummy variable of whether the pre-search filter was applied
4. *Time spent on product* – time spent on looking at product page during the browsing session[3]

Table 1. Summary statistics for search variables

Variable	Mean	Std. dev.	Median	1% quantile	99% quantile
Last item searched	0.1196	0.3245	0	0	1
Number product searched	*19.4738*	*18.2678*	*14*	*1*	*86*
Applied filter	*0.5145*	*0.4998*	*1*	*0*	*1*
Time spent on product	*136.4158*	*1944.3450*	*44*	*6*	*813*

From Table 1 we see that 49% of customers do not use pre-search filters. Only 10% of orders have more than one filter, hence we use the binarized variable to improve the statistical power of the analysis.

Notice that variables *Number product searched,* and *Time spent on product* can take only positive values and could have a fat right tail. This implies that there could be high-leverage outliers, hence it is natural to consider the logarithm of these variables in subsequent analysis.

The dependent variable in our analysis is binary. It takes one if the product was returned and zero otherwise[4]. This nature of the outcome suggests using logit or probit framework in estimation. We present the results for logit model in this section and leave probit results to the appendix. We also consider linear probabilistic model for robustness in Appendix A.

Mathematically, we consider the following specification in our analysis:

$$\mathbb{P}[Y_i = 1|Z_i, X_i] = \frac{e^{Z_i' \cdot \alpha + X_i' \cdot \beta}}{1 + e^{Z_i' \cdot \alpha + X_i' \cdot \beta}} \quad (1)$$

[2] We assume that one unique product includes all the size and color varieties, in other words, if customer looked at two colors of the same product it would be counted as one unique product.
[3] We assume that the time the customer looked at particular product is equal to the difference in time between closing and opening the product page.
[4] Formally, there could be the case when the customer purchased two identical products and returned one of them. However, the proportion of these cases was negligibly small, and we treat such cases as one item purchases and returned.

Where $Y_i \in \{0,1\}$ is binary variable taking one if the product was returned and zero otherwise, Z_i - matrix of control variables and X_i matrix of variables related to the customer search.

The set of control variables summarized in Table 2. Notice that not all variables could be used in case of single-product orders. For example, product price and order sum would be identical in this case which leads to perfect multicollinearity.

Table 2. Set of control variables

Variable	#	Description
Same Product Quantity	1	Number of different sizes of the same product purchased
Product Colors Quantity	1	Number of different colors of the same product purchased
Same Product Category Quantity	1	Number of different items in the same product category
Category Quantity	1	Number of different categories of products purchased
Price	1	Base price of product before the discount
Discount	1	Discount on product
Order Sum	1	Sum of the order before the discount
Order Discount	1	Sum of the order after all discounts
Order Quantity	1	Number of products in the order
Device FE	2	Computer, tablet
Time FE	14	Month and weekday
Category FE	21	Dress, skirt, bluse
Color FE	23	Red, black, green
Size FE	10	XL, M
Material FE	13	Cotton, leather
Brand FE	5	Casual, Menplus

Most of our control variables are categorical variables which we use as one-hot encodings[5]. We omit the exact coefficients and use them as fixed effects.

We document the results in Table 3. Model 1 is based on the subsample of the order which contained only one purchased product while Model 2 uses the whole sample (one or more purchases). In Appendix A we present the average marginal effects of variables of interest, that is the measure of the effect on probability.

Table 3. Estimation results for Equation (1). Standard error in parenthesis.

Variable	Model 1	Model 2
Last item searched	*-0.1083*	*-0.0638*
	(0.018)	*(0.007)*
Number product searched (log)	*0.0593*	*0.0529*
	(0.032)	*(0.010)*
Applied filter	*-0.0803*	*-0.0385*
	(0.017)	*(0.005)*

[5] For example, if variable x takes three possible values $\{S, M, L\}$ then one-hot encoding would map one-dimensional variable x to 3-dimensional variable $[\mathbb{I}[x = S], \mathbb{I}[x = M], \mathbb{I}[x = L]]$. Note that including one-hot encoding is identical to use fixed effects.

Time spent on product (log)	*-0.0475*	*-0.0731*
	(0.015)	*(0.006)*
Same Product Quantity (log)		*3.6102*
		(0.043)
Product Colors Quantity (log)		*0.4735*
		(0.030)
Same Product Category Quantity (log)		*0.3338*
		(0.020)
Category Quantity (log)		*-0.5838*
		(0.027)
Price (log)	*1.3072*	*0.7812*
	(0.063)	*(0.021)*
Discount (log)	*-0.2705*	*-0.1367*
	(0.015)	*(0.008)*
Order Sum (log)		*1.1579*
		(0.019)
Order Discount (log)		*-0.1532*
		(0.005)
Order Quantity (log)		*0.3915*
		(0.030)
Device FE	*Yes*	*Yes*
Time FE	*Yes*	*Yes*
Category FE	*Yes*	*Yes*
Color FE	*Yes*	*Yes*
Size FE	*Yes*	*Yes*
Material FE	*Yes*	*Yes*
Number points	*94,477*	*993,429*
Pseudo R^2	*0.0529*	*0.1081*

The directions of the effects are identical in both of these subsamples of the data. This suggests that the way search affects the probability of return does not depend on whether the customer purchased one item or several. Thus, considering only single-product orders is not restrictive for identifying the stylized facts in this paper. Although examining the additional information provided by multiple-product orders could provide further insights, we leave these analyses for further research.

Looking carefully at the results identifies potential stylized facts.

1. **Customers who purchased the last item searched are less likely to return the product.** The fact that it was the last item searched indicates that customer stopped the search because she found something which she really likes. As the results she is less likely to return the item.
2. **Customers who consider many products are more likely to return the product.** The increasing part of the graph is explained by the type of customers. Specifically, a customer who has a lower variance of unobserved fit would search for fewer items as she or she doesn't expect to find something really good. However, when she

accidentally finds a product which she likes and makes a purchase - the expectations about this product would be very close to the realization. Thus, the customer is less likely to return the product.

3. **Customers who apply pre-search filters are less likely to return the product**. Intuitively, this implies that customers who know their preferences or have a very strong preference for a particular attribute would be willing to change the sampling distribution by paying some mental costs of applying filters. These customers know better what they want and thus less likely to be negatively surprised by the item they purchase. Thus, they would return less frequently.
4. **Customers who spend more time reviewing the product page are less likely to return the product.** Intuitively, by reviewing the product page, the customers extract information from the page. For example, she reads the description, compares with already searched items, looks at the picture. Thus, she gets a better signal of unobserved fit and finds a better product that fits her preferences. As a result, the customer is less likely to return the product which she reviewed carefully.

4. Core model

4.1. Overview

Our model assumes that the customer journey consists of three main parts.

1. **Search**. The customer browses the website and looks at different product pages. By inspecting the product page, the customer not only learns the objective characteristics of the product like price or materials but also learns how well the product would fit him. While she is browsing, she forms her consideration or choice set. When she feels that she won't find anything better to improve the consideration set or just tired she finishes her search and moves to the next step.
2. **Purchase**. After the previous step, the customer formed the choice set of products from which she is choosing the best alternative. She takes into account both the objective characteristics and the information which she learned about the product fit. However, if she is not satisfied with the alternatives, she can leave the website without the purchase.
3. **Return**. If the customer made a purchase, she receives her order and immediately tries it on. By doing this she perfectly learns how the product matches her preferences and understands precisely how much she likes it. If she thinks that the product does not fit she has an option to return it to the retailer, however, she has to pay some return costs which could be either monetary or mental.

We start the development of the model from the terminal stage and at the end describe the dynamic problem which is solved by the customer. The basic notation for the model is summarized in Table 4.

Table 4. Notation for the core model

Variable	Description
x_i	k-dimensional random vector of product's i characteristics like price, material, size, etc
$F_x(\cdot)$	cumulative distribution function (c.d.f) of x_i
s_{nj}	marginal search cost of customer j after looking at n products
c_j	return costs of the customer j
β_j	k-dimensional preference vector of customer j
θ_j	type of the customer j. This variable contains all the variables which are fixed for one customer. For example, in this section $\theta_j = (c_j, \beta_j, s_{nj})$
ψ_{ij}	preference shock of customer j for product i (or individual fit)
$\hat{\psi}_{ij}$	signal of individual fit which she can extract from the product page
η_{ij}	pair $(x_i, \hat{\psi}_{ij})$ the information about the product which customer has before the purchase
$\mathcal{I}_n$	set indices of product which customer already searched after she looked at n products
H_n	$\{(x_i, \hat{\psi}_{ij}) : i \in \mathcal{I}_n\}$ the set of products which customer already searched after she looked at n products

4.2. Return decision

Consider customer j who already received the product and is deciding whether to return or keep the product. We assume that:

$$\begin{aligned} &\mathcal{U}(keep)_{ij} = x_i'\beta_j + \psi_{ij} = \mu_{ij} + \psi_{ij} \\ &\mathcal{U}(return)_{ij} = -c_j \end{aligned} \tag{2}$$

The utility of the return is simply the negative return costs and by assumption, it is always negative. We assume that this variable includes all possible costs associated with product return and could include both monetary like shipping or fee and mental costs such as opportunity costs. The utility of keeping an item consists of two components. $\mu_{ij} = x_i'\beta_j$ is part of the utility observable both to the customer and to the researcher. This specification allows the heterogeneity in preferences and customer-specific intercept. The second part ψ_{ij} is the individual fit of the customer which is not observed by the researched and observed by the customer only when she receives the product.

The customer maximizes her utility and thus chooses the option which is higher in expected utility. Thus, we can write the final utility of the customer in a compact form as[6][7]:

$$\begin{aligned} \mathcal{U}(final)_{ij} = \mathbb{I}\left[\mathcal{U}(keep)_{ij} \geq \mathcal{U}(return)_{ij}\right] \cdot \mathcal{U}(keep)_{ij} + \mathbb{I}\left[\mathcal{U}(keep)_{ij} < \mathcal{U}(return)_{ij}\right] \\ \cdot \mathcal{U}(return)_{ij} = \\ = \left(x_i'\beta_j + \psi_{ij} + c_j\right)^+ - c_j \end{aligned} \tag{3}$$

[6] We use the following notation:

$$\mathbb{I}[\mathcal{A}] = \begin{cases} 1, & \text{if } \mathcal{A} \text{ is True} \\ 0, & \text{if } \mathcal{A} \text{ is False} \end{cases}$$

[7] $(x)^+ = x \cdot \mathbb{I}[x \geq 0] = \max\{0, x\}$

4.3. Purchase decision

Consider the customer who have I items in her consideration set after the search. By assumption for each element of the choice set the customer observes the vector of product characteristics x_i and signal of product fit $\hat{\psi}_{ij}$. Intuitively, it implies that the purchase online is associated with uncertainty, for example, the fashionable dress could look very good online but when the customer tries it's on at home, she realizes that it does not fit. The fact that the customer does not observe ψ_{ij} partly explains the reason for high return rates in online retail as offline the customer could try the item at the store and get the realization of ψ_{ij}. However, at the moment of purchase, the website allows him to extract some information about the product, for example, by looking at the pictures or reading the description. In other words, the customer observes some estimates $\hat{\psi}_{ij}$ of the unobserved preference shock ψ_{ij} and based her purchase decision on this estimate and distribution $\psi_{ij}|\hat{\psi}_{ij}$. We assume that the customer knows the distribution of all the variables.

The customer wants to maximize the expected utility by using all the information she has at the moment of purchase. Thus, for each product in her consideration set she calculates the expected utility of purchasing this particular product:

$$\begin{aligned} \mathcal{U}(purchase\ i)_{ij} &= \mathbb{E}\left[\mathcal{U}(final)_{ij}|x_i, \hat{\psi}_{ij}\right] = \\ &= \mathbb{E}_{\psi_{ij}}\left[\left(x_i'\beta_j + \psi_{ij} + c_j\right)^{+}|\eta_{ij}\right] - c_j \equiv g\left(x_i, \hat{\psi}_{ij}, \beta_j, c_j\right) \equiv g_{ij} \end{aligned} \tag{4}$$

Where function $g(\cdot)$ is some deterministic function which depends on the distribution of $\psi_{ij}|\hat{\psi}_{ij}$. For some distributions, this has a closed-form which we consider later in the chapter. We would omit the parameters of the function when they are not important for the understanding and assume that this is some random variable g_{ij}.

The customer chooses the product which yields the highest expected utility. Thus, the expected utility of purchase given the choice set $\mathcal{S}$:

$$\begin{aligned} \mathcal{U}(purchase|\mathcal{S})_j &= \max_{i\in\mathcal{S}} \mathcal{U}(purchase\ i)_{ij} = \\ &= \max_{i\in\mathcal{J}} \mathbb{E}_{\psi_{ij}}\left[\left(x_i'\beta_j + \psi_{ij} + c_j\right)^{+}|\eta_{ij}\right] - c_j \end{aligned} \tag{5}$$

4.4. Search decision

After the customer searched n products she needs to make a decision either to search for a new item i' and add it to her choice set $\mathcal{S}_n$ or go to purchase decision and select the best alternative from the set H_n.

$$\begin{aligned} &\mathcal{U}(search|H_n)_j = \mathbb{E}_{\eta_{i'j}}\left[\mathcal{U}\left(search|H_n \cup \{\eta_{i'j}\}\right)_j\right] - s_{nj} \\ &\mathcal{U}(not\ search|H_n)_j = \mathcal{U}(purchase|H_n)_j \end{aligned} \tag{6}$$

Note that if the customer decides to sample a new product, she would be in the same state except with new $\mathrm{H}_{n+1} = \mathrm{H}_n \cup \{\eta_{i'j}\}$, however, she does not know the realization of $\eta_{i'j}$ before the decision and thus takes the expectation about it.

4.5. Dynamic problem

Finally, using the Equation (6) above we can write down the dynamic program solved by the customer

(7)
$$\begin{aligned}
&\mathcal{U}(search|H_n)_j = \\
&= max\left\{\mathbb{E}_{\eta_{i'j}}\left[\mathcal{U}\left(search|H_n \cup \{\eta_{i'j}\}\right)_j\right] - s_{nj},\ \ \mathcal{U}(purchase|\mathcal{S})_j\right\} \\
&= max\left\{\mathbb{E}_{\eta_{i'j}}\left[\mathcal{U}\left(search|H_n \cup \{\eta_{i'j}\}\right)_j\right] - s_{nj},\ \ \max_{i\in\mathcal{J}} \mathbb{E}_{\psi_{ij}}\left[\left(x_i'\beta_j + \psi_{ij} + c_j\right)^+ |\eta_{ij}\right] - c_j\right\}
\end{aligned}$$

By denoting the value function $\mathcal{V}(\mathrm{H}, n) = \mathcal{U}(search|\mathrm{H}_n)_j$ we finalize the dynamic program for customer j in Equation (8):

(8)
$$\begin{aligned}
&\mathcal{V}(H, n) = \\
&= max\left\{\mathbb{E}_{\eta'}[\mathcal{V}(H \cup \{\eta'\}, n+1)] - s_n,\ \ \max_{i\in\mathcal{J}} \mathbb{E}_{\psi_{ij}}\left[\left(x_i'\beta_j + \psi_{ij} + c_j\right)^+ |\eta_{ij}\right] - c_j\right\}
\end{aligned}$$

The dynamic program in this form is not very convenient for the solution as the value function depends on the set of random variables H. In Appendix B we show that this problem could be formulated in a simpler form where the value function $\mathcal{V}(\mathrm{z}, n)$ depends only on the number of products searched n and the maximal expected purchase utility z:

(9)
$$\begin{aligned}
&\mathcal{V}(z, n) = \\
&= max\left\{F_g(z)\cdot\mathcal{V}(z, n+1) + \int_z^\infty \mathcal{V}(t, n+1)dF_g(t) - s_{nj},\ z\right\}
\end{aligned}$$

Where $F_g(\cdot)$ is the known distribution of the reward discussed in the Appendix B.

In Appendix C we show that the dynamic program in Equation (9) has a solution in a form of a threshold policy. In other words, for each number of searched products n there exist a threshold z_n^* such that

(10)
$$\mathcal{V}(z, n) = \begin{cases} z, & z \geq z_n^* \\ F_g(z)\cdot\mathcal{V}(z, n+1) + \int_z^\infty \mathcal{V}(t, n+1)dF_g(t) - s_{nj}, & z < z_n^* \end{cases}$$

Moreover, thresholds are decreasing in n and could be found from the equation:

(11)
$$F(z)\cdot z + \int_z^\infty t\cdot dF(t) - z = s_n$$

The threshold policy results make an intuitive sense. While searching customer always keeps in mind her best alternative from the product she already searched (this is exactly the variable z). At each state (number of products searched n) she compares her best alternative with the expected reward of the search z_n^* and makes a purchase decision if she does not expect to find something better than she already has.

The decreasing thresholds z_n^* reflect the fact that customer getting tired of the search. As the results each new product requires more effort from the customer and her reservation utility of the search decreasing with each product searched.

It is helpful to contrast the model to the case when $s_n = s = const$. In this case the thresholds would be identical, and customer would search until she finds the product which yields the utility greater than the threshold. This implies that all search sessions must end with product purchase and that it would be always the last product searched. Both of these implications are unrealistic and are not supported by our data. Increasing search costs ensures that the customer could purchase an item which she searched at the beginning or not purchasing at all.

An extension, to be explored in future research, complicates the model by allowing the customer to update her beliefs with each product search. We currently do not have data to distinguish between these models, so we select a more parsimonious specification that is consistent with the stylized facts.

The careful reader may notice that the solution of the model does not explicitly respond for product returns. This observation is correct in a sense that the solution only requires the distribution of the reward $F(\cdot)$ and agnostic to the return itself. Nonetheless, returns are implicit because the option to return changes the distribution of the rewards which, in turn, changes the thresholds z_n^* and thus the search behavior.

4.6. Distributional assumptions.

The dynamic problem was formulated for general assumptions on the distribution and the threshold solution holds for a wide class of distributions. To analyze further the behavior of the model we impose distributional assumption. This allows us to find a closed-form solution for the $g(\cdot)$ function and helps illuminate the properties of the model.

We impose the normality assumption on the distribution of unobserved fit. In particular,

Table 5. Distributional assumptions of the core model

$\hat{\psi}_{ij} = \psi_{ij} + \nu_{ij}$	signal of product fit is unbiased
$\psi_{ij} \perp \nu_{ij} \perp x_{ij}$	the noise of the signal; true product fit and product characteristics are independent
$\psi_{ij} \sim \mathcal{N}(0, \sigma_\psi)$ $\nu_{ij} \sim \mathcal{N}(0, \sigma_\nu)$	noise and product fit come form a joint normal distribution

This set of assumptions assume that products are homogenous in unobserved fit. However, the original problem allows the model to assume the category, dependent unobserved fit variance. In Appendix D we show that under such assumptions the $g(\cdot)$ function could be written in the following form:

$$g(x_j', \hat{\psi}_{ij}', \beta_j, c_j) = \sqrt{\frac{\sigma_\nu^2 \sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}} \cdot \left(\varphi(\alpha) + \alpha \cdot \Phi(\alpha)\right) - c_j$$

(12)

$$\alpha = \frac{x_i' \beta_j + c_j + \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2} \hat{\psi}_{ij}}{\sqrt{\frac{\sigma_\nu^2 \sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}}}$$

Remember that $g(\cdot)$ is expected utility of purchase given product characteristics x_i and signal of unobserved fit $\hat{\psi}_{ij}$. Under assumption that the reservation utility of the customer equals zero the customer would make a purchase only if her final choice set contains at least one product with positive expected purchase utility. The Equation (12) takes into account the return option which in essence implies that the negative utility is constrained from the below by $-c_j$ and return options allow customers to hedge from the negative utility.

Expectedly, the higher the return costs c_j the smaller the expected utility from the purchase, for example, if returns are prohibited then the expected utility is just $\mathbb{E}\left[\mathcal{U}(keep)_{ij} | x_i, \hat{\psi}_{ij}\right] = x_i' \beta_j + c_j + \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2} \hat{\psi}_{ij}$. This is exactly the estimate of the final utility given the product characteristics x_i and signal of product fit $\hat{\psi}_{ij}$.

The more interesting variables are σ_ψ^2 and σ_ν^2. σ_ψ^2 represents the variance of customer's preferences. The larger the value of σ_ψ^2, the more likely the customer extremely likes or dislikes the product. However, because the customer is hedged by the return option, she likes this uncertainty and the expected purchase utility is an increasing function of σ_ψ^2. Specifically, by purchasing the product the customer has a chance to get something which would be wonderful while the chance of getting an awful product is mitigated by the return option.

σ_ν^2 represents the quality of the signal of unobserved fit. Intuitively, the customer with small σ_ν^2 knows well her unobserved preferences or is familiar with the website. Customers with $\sigma_\nu^2 = 0$ know perfectly the final utility of the product from the website alone, because $\hat{\psi}_{ij} = \psi_{ij}$. Such customers would purchase only the product with observed positive utility and never return the purchase. Customers with $\sigma_\nu^2 = \infty$ would use only objective product characteristics in her decision making. Such customers frequently make purchase products that do not match the customer's preferences, hence returns would be common.

Can the return option affect the search behavior?

Search as modeled is fully defined by the distribution of the expected purchase utility $F_g(\cdot)$ in Equation (10). At the same time notice that under normality assumptions we find the distribution of the reward as:

$$F_g(t) = \mathbb{P}[g_i \le t] = \mathbb{E}_{x_j, \widehat{\psi_{ij}}}\left[\mathbb{I}\left[g\left(x_j', \hat{\psi}_{ij}', \beta_j, c_j\right) \le t\right]\right] \tag{13}$$

Where the expectation is taken over random variables x_j and $\hat{\psi}_{ij}$ with known distributions and $g(\cdot)$ is function defined in Equation (12).

In the previous subsection we discussed how c_j, $\sigma_\psi^2, \sigma_\eta^2$ are all related to the return behavior of the customer. But at the same time the Equation (12) clearly suggests that change in these variables would lead to the change of search behavior as they change the distribution $F_g(\cdot)$.

This could have significant consequence for the retailer. For example, by making return policy more generous the retailer could expect the customer would spend more time on the website. This follows from the fact that by decreasing c_j the mean of the distribution $F_g(\cdot)$ goes up and customer would have higher expected utility of search.

Can the search affect the return decision?

For clarity in addressing this question, we simplify the model and assume that we observe customer's choice set (set of searched products) but not the underlying process. Also, assume that customer does not have an outside option. We still assume that the customer chooses the best alternative from the set of searched products and that the distributional assumption holds.

In Appendix E we show that probability of return is:

$$\int_{-\infty}^{\infty} F_{\psi_n|\hat{\psi}_n,x_n}(-x_n'\beta - c) \cdot \prod_{k=1,k\neq n}^{N} F_{\hat{\psi}_k|x_k}\left(t_n - \frac{\sigma_\psi^2 + \sigma_v^2}{\sigma_\psi^2}(x_k' - x_n')\beta\right) \cdot dF_{\hat{\psi}_n|x_n}(t_n) \tag{14}$$

We seek only an example where search affects the return decision. We illustrate this case with the following parameter values:

Table 6. Simplifying assumptions for Equation (14)

$\beta = 0$	products are identical in observable characteristics. Notice that this implies that each product has equal probability of being purchased from a set $\frac{1}{N}$
$\sigma_\psi^2 = \sigma_\eta^2 = 1$	noise and unobserved fit variance both equal to 1

The Equation (14) further simplifies and by dividing by probability of purchase $\frac{1}{N}$ we get the conditional probability of return as:

$$\int_{-\infty}^{\infty} \Phi\left(\frac{-c - \frac{1}{2}t_n}{\sqrt{1/2}}\right) \cdot N \cdot \Phi\left(\frac{t_n}{\sqrt{2}}\right)^{N-1} \cdot d\Phi\left(\frac{t_n}{\sqrt{2}}\right) \tag{15}$$

In this example, which demonstrates existence, the customer is less likely to return the product if she has more products in her search set. This has an intuitive sense as the customer who have huge choice set is more likely to get a good realization of unobserved fit (for example, the mean of the distribution of the maximum over identical random variable is increasing in number of random variables).

Finally, the Equation (15) clearly indicates that the probability of returns depends not only on the characteristics of the purchased product but also characteristics of all other products which were searched by the customer. To see this, consider simple case when product has only binary price characteristic p_H or p_L ($p_H \gg p_L$) and unobserved fit. Two customers search two products each: one searched product with $\{p_H, p_L\}$ and another $\{p_L, p_L\}$. Both of them purchased product with p_L. In this case the first customer would have a higher probability of return as her product with p_H has a lower utility and thus it does not constraint the unobserved utility. Notice that the firm which disregards the search information would not be able to distinguish these two types of customers.

5. Model, stylized facts and predictions

We now demonstrate that our model is consistent with the stylized facts discussed in Section 3. Our goal is to establish existence. In some cases, we select particular parameter values and in other cases we elaborate the model, but without modifying the basic structure.

5.1. Customers who purchased the last item searched are less likely to return the product

Consider the simplest case when all the products are identical in product characteristics (for example, $x_i = 0\ \forall i$) and customer is sampling the unobserved fit (and thus receives only the signal at the moment of purchase). Our model assumes that the customer ends search when her best alternative is better than the current threshold z_n^*. Because z_n^* is decreasing in number of searched products n this could happen either if customer tired of search or she finds something which she really likes. Consider the example where there are two identical sets customers with session length n. The first set contains customers who purchased the item searched next to last. The second set contains customers who purchased the item searched last.

The customer from the first set found her product earlier in her search session. Moreover, she didn't purchase it at $n-1$ step thus it implies that her chosen alternative yields expected utility lower than the threshold z_{n-1}^*. Therefore, her expected purchase utility would lie within z_{n-1}^* and z_n^*. In contrast, customer who ended her search by finding the product she really likes would imply that the chosen alternative yields the expected utility above z_n^* but there isn't an upper bound. Clearly, the second customer is less likely to return the product as her expected purchase utility on average is higher than the first customer's one.

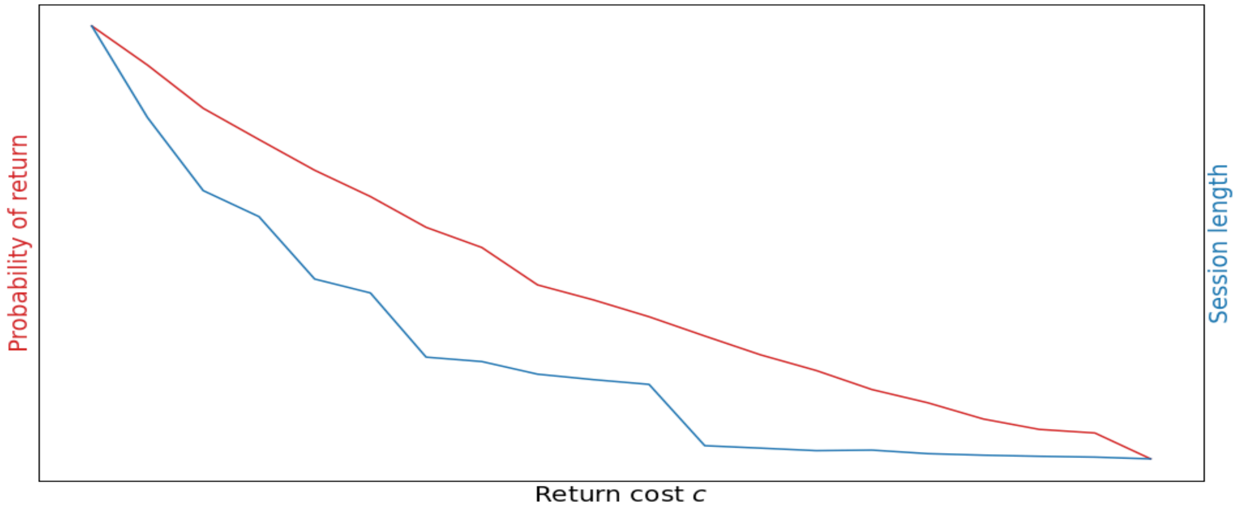
Therefore, the "purchased the last item searched" indicators segments customers based on the reason she stopped the search process which in turn is indicative of product returns. Notice that this variable is observable to the researcher (or firm) while the signals on unobserved fit are observable only by the customer.

To validate the result of the model we empirically investigate related prediction. From the discussion above it is directly follows that customer who purchased the last item searched should have on average shorter sessions. We run a regression where set of control was the same as in Equation (1) but the outcome was changed to the *Number of products searched.* Our empirical exercise revealed that customers who purchased the last searched items on average search 26% fewer products. This supports the prediction of our model.

5.2. Customers who consider many products are more likely to return the product

Consider a set of customers who are identical except for the return costs c. We implement Monte-Carlo simulations approach to plot the dependence of session length and probability of return as function of c. We leave the details of simulation to Appendix F and plot the results in Figure 1:

Figure 1. Results of Monte Carlo Simulations



Customer with higher c would have higher penalty for the return and thus she is clearly less likely to return the product. At the same time, her distribution of the expected purchase utility has a lower mean. The return option makes the distribution truncated from below by $-c$ and thus by increasing return costs the mean of the truncated variable goes down. Intuitively, the customer knowing the that return policy is not very generous would sample couple of items on the website. If she doesn't find something really good, she would stop her search and, for example, would go to the competitor with lenient return policy. Therefore, customers with smaller search sessions would return items less frequently.

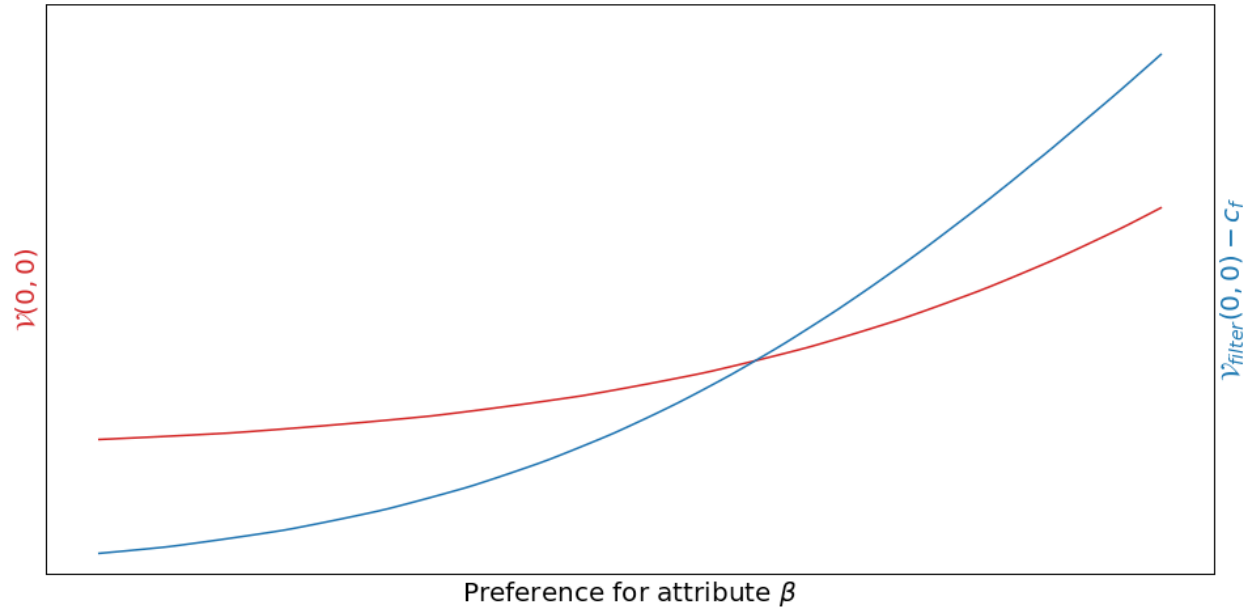
5.3. Customers who apply pre-search filters are less likely to return products

Consider similar simulation setting as in the previous subsections despite now let the product have one binary attribute (for example, size M). Because customers could be heterogenous in their preferences let the customer have different preference β for this attribute and she has an option to sample from the distribution of the products which all have (or not) this attribute. However, this is not the costless option and customer needs to pay some fixed cost c_f to change the distribution. Intuitively, the application of pre-search filters requires customer to pay some costs, for example, she has to spend time on getting used with the interface and pay mental costs on thinking which filter to apply.

From the model perspective it implies that the customer has an option to change the distribution of the reward $F_g(\cdot)$ in Equation (11) at the beginning of her search journey. This would change the optimal thresholds and expected search utility of search. Therefore, she compares the net gains in expected search utility with costs c_f and makes a decision either to change the distribution (or apply filter) or sample from the baseline one.

In **Figure 2** we plot as the function of β the base search expected utility at the beginning of session $\mathcal{V}(0,0)$ and net gains of applying filter $\mathcal{V}_{filter}(0,0) - c_f$. We assume that $\beta \geq 0$ and thus customer would not benefit to sampling from distribution where all products don't have an attribute.

Figure 2. Results of estimation



The expected search utility is an increasing function of preference for attribute and after some threshold customer would prefer to apply the filters and sample from the better distribution.

This implies that the use of pre-search filters provides the information about the preferences of the customer. There two reasons why the return probability would be smaller for those who use filter. Firstly, customer would sample from better distribution as the result she would not face "bad" product and more likely to enjoy the purchase. Notice that this is true for all customers who have positive preference for attribute and generally the retailer would want to make the pre-search interface more intuitive to reduce c_f. Secondly, customer who apply filter have on average higher preference for the attribute than the rest of the population. For these customers the expected purchase utility is higher which results in lower return rates.

Notice that customer who decides to use pre-search filter would sample from a better distribution (distribution with higher mean of expected reward). Therefore, on average such customers would sample more product because their reservation utility is higher. We test this prediction on our data. Specifically, after controlling for the same set of variables as in Section 3 we observe that customers who use pre-search on average sample 9% more products.

5.4. Customer who spends more time on reviewing the product page is less likely to return the product

Consider the scenario when customer open the product page and started to consume the information about the product. She reads the description, looks at pictures, look at the patters, etc. However, each new line of the description makes her to pay some mental or processing costs, thus at some point she would stop the investigation and would have to make a decision. Therefore, with more time the customer extracts more information about the product, but the time spent is costly.

Mathematically, by consuming each unit of information customer gets a signal $\eta_{\Delta t}$ of quality $\sigma_{\Delta t}$ for the cost of $c_{\Delta t}$. The customer could choose how many of these signals to get before she makes a purchase decision. For example, we may treat each signal $\eta_{\Delta t}$ as additional picture which customer looks at. If customer got T signals and her final estimate for the product unobserved fit would be $\hat{\psi} = \frac{\sum_t \hat{\psi}^t}{T} = \frac{\sum_t \psi + \eta_{\Delta t}^t}{T} = \psi + \frac{\sum_t \eta_{\Delta t}^t}{T} = \psi + \eta$. Notice that $\eta \sim \mathcal{N}\left(0, \frac{\sigma_{\Delta t}^2}{T}\right)$ and thus we can refer to the previous notation as $\sigma_\eta^2 = \frac{\sigma_{\Delta t}^2}{T}$

The discussion above suggests us that time spend on the product is negatively correlated with return probability. Our core model accounts for this through the quality of the signal σ_η and could be estimated with σ_η being a function of time spent on product page. More sophisticated model would endogenize the time spent on reviewing the product page, however, the extension is straightforward, and we leave it to further research.

6. Demonstration that the Model Can be Estimated

As of this writing, estimation is not complete. For the purposes of this paper, we seek to demonstrate that it is possible to establish a likelihood function from which to estimate the model. Full analysis of identification and a feasible method (other than time-consuming brute force) is left to future research. We use the following notation in this section. We continue to use all notation from Section 5.

Table 7. Notation for Section 6

Variable	Description
$\mathcal{G}_j$	$\{g_i : i = 1 \dots N\}$ the sequence of expected purchase utilities of searched products by the customer j
z_n^*	search decision threshold after n were searcehd
g_0	outside option of the customer
g_i	search expected utility of purchasing product i. Note that it is actually a function of observable to customer characteristics of the product $g(x_i, \psi_{ij})$ we omit arguments to simplify notation.
z_n	$\max_{i \in \{0..n\}} g_i$ or maximal expected purchase utility by the time customer searched n items.
$\mathcal{X}_j$	$\{x_i : i = 1 \dots N_j\}$ the sequence of product features of searched products by the customer j. We assume it is ordered sequence.
θ_j	$(g_0, \beta_j, \sigma_{\psi j}, \sigma_{\eta j}, c_j, s_{nj})$ type of the customer j or set of parameters of the model

We start by considering one customer with the type θ. Therefore, we can use the Equation (11) to compute the sequence of thresholds z_n^* given θ and treat them as fixed throughout our analysis. We also use the distributional assumption introduced in Section 4 for the derivation of likelihood function.

The researcher observes only the sequence of searched product by customer, the purchased product (if any) and whether it was returned or not. Thus, our goal is to compute the probability of each event given the $\mathcal{X}_j$.

In Appendix G we show that under model assumptions the probabilities could be computed as:

$$\mathbb{P}[purchase\ N\ and\ return\ it|\mathcal{X}] =$$

$$= \int_{h_{x_N}(max\{z_N^*, g_0\})}^{\infty} \Phi\left(\frac{-x_N'\beta - c - \frac{\sigma_\psi^2}{\sigma_\nu^2 + \sigma_\psi^2} t_N}{\sqrt{\frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}}}\right)$$

$$\cdot \prod_{k=1}^{N-1} \Phi\left(\frac{min\left\{h_{x_k}(z_{N-1}^*), \frac{\sigma_\psi^2 + \sigma_\nu^2}{\sigma_\psi^2} \cdot (x_N'\beta - x_k'\beta) + t_N\right\}}{\sigma_\psi^2 + \sigma_\nu^2}\right) \cdot d\Phi\left(\frac{t_N}{\sigma_\psi^2 + \sigma_\nu^2}\right)$$

$$\mathbb{P}[purchase\ n\ and\ return\ it|\mathcal{X}] =$$

(16)

$$= \int_{h_{x_n}(max\{z_N^*, g_0\})}^{h_{x_n}(z_{N-1}^*)} \Phi\left(\frac{-x_n'\beta - c - \frac{\sigma_\psi^2}{\sigma_\nu^2 + \sigma_\psi^2} t_n}{\sqrt{\frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}}}\right)$$

$$\cdot \prod_{k=1, k\neq n}^{N-1} \Phi\left(\frac{min\left\{h_{x_k}(z_{N-1}^*), \frac{\sigma_\psi^2 + \sigma_\nu^2}{\sigma_\psi^2} \cdot (x_n'\beta - x_k'\beta) + t_n\right\}}{\sigma_\psi^2 + \sigma_\nu^2}\right)$$

$$\cdot d\Phi\left(\frac{t_n}{\sigma_\psi^2 + \sigma_\nu^2}\right)$$

$$\mathbb{P}[no\ purchase|\mathcal{X}] = \prod_{k=1}^{N} \Phi\left(\frac{h_{x_k}(g_0)}{\sigma_\psi^2 + \sigma_\nu^2}\right)$$

where $h_{x_i}(\cdot)$ is known function defined in Appendix G and $\Phi(\cdot)$ is c.d.f of standard normal distribution.

Notice that the probability for $n = N$ and $n \neq N$ have different forms. This is not surprising as the fact that customer purchased not the last product searched gives additional information about the product's utility. Specifically, when the last searched product was purchased the researcher only knows that it was greater than the last threshold z_N^* but is not constrained from the above. The probability of not purchasing anything is simply the probability of the event when all products are worse than outside option. Computation of the probability of event "purchase item n and keep it" is trivial.

In Appendix G we also discussed feasibility constrains on the parameters of the model:

(17)

$$\mathcal{C}(\theta) = \begin{cases} \mathbb{I}[g_0 < z_{N-1}^*], & purchase \\ \mathbb{I}[z_N^* \leq g_0 < z_{N-1}^*], & no\ purchase \end{cases}$$

Intuitively, violation of these constraints imply that it is impossible to observe the given sequence under such model. For example, customer can't have too huge search costs and good outside option as in this case she would not use the website for search. Formally, we could multiply the probabilities by corresponding feasibility constraints.

The constraints in Equation (17) imply that our model relies on the heterogeneity in customer parameters. Specifically, given parameters of the model fixed our model that all sessions without a purchase would have identical length. This implies that having the data without purchases provides additional information. Notice that for the case when we ignore the data without a purchase, we only have first part of constraints in Equation (17) and $g_0 \to -\infty$. Therefore, all constraints in Equation (17) could be ignored.

In general case customer type comes from some distribution $\theta_j \sim F_\theta\left(\cdot \mid \tilde{\theta}\right)$. In the data we observe each customer's sequence of searched products $\mathcal{X}_j$ and whether any of the product was purchased or returned $\mathcal{Y}_j$.

We can establish the likelihood of observing $\mathcal{Y}_j$ given $\mathcal{X}_j$ by combining Equations (16) and (17) and denote it as $\mathcal{L}(\theta)$. Because θ random we integrate over distribution $F_\theta\left(\cdot \mid \tilde{\theta}\right)$[8]

$$\hat{\theta} = \underset{\tilde{\theta}}{\operatorname{argmax}} \int \mathcal{L}(t) dF_\theta\left(t \middle| \tilde{\theta}\right) \tag{18}$$

Equation (18) is sufficient to estimate model parameters. The main objective is computational complexity of the optimization problem. All probabilities in Equation (16) and Equation (18) involve several layers of integration which requires substantial computational resources. We leave feasible estimation for further work and could achieve it by adding simplifying assumptions.

7. Prediction results

While work proceeds on estimating the theoretical model, we explore a simple reduced-form prediction model to demonstrate that search data contain information in the sense that, if they are included in a model, they predict significantly better than a simple model without search data.

As an initial proof of concept, we estimate and evaluate our models on the subset of the data with single-product order. In this case we can ignore within-order interactions between products. For example, a customer could purchase two shirts with intent to keep one of them. Formally, this implies that the return of one product could be correlated with the return of other product in the order. This imposes complications for the construction of a model which does not use search information.

For the purposes of this paper, we estimate two reduced-form prediction models:

$$\begin{aligned} &\mathbb{P}\left[Y_i = 1 \middle| X_i^{purch}, X_i^{search}\right] = f\left(X_i^{purch}, X_i^{search}\right) \\ &\mathbb{P}\left[Y_i = 1 \middle| X_i^{purch}\right] = g\left(X_i^{purch}\right) \end{aligned} \tag{19}$$

[8] Note that it is straightforward to constraint some parameters of θ to be homogeneous.

Where X_i^{purch} and X_i^{search} are the sets of variables characterizing the order purchase and search information respectively; $f(\cdot,\cdot)$ and $g(\cdot)$ are some nonlinear function estimated on the data; $Y_i \in \{0,1\}$ binary variable indicating whether the product was returned or not

Because we have an extremely large number of variables and because we want to account for nonlinearity, we use a gradient-boosted-trees framework. The main advantage of this approach is the ability to capture higher-order interactions among variables. Moreover, by using regularization, we avoid overfitting to obtain the best predictive accuracy. We use the LightGBM package based on its proven performance in the Kaggle machine-learning competitions. We describe both sets of variables, X_i^{purch} and X_i^{search}, in Appendix H.

Table 8 reports the predictive ability of different models. We evaluate the predictions with 10-fold cross-validation approach. Specifically, for each of ten draws we select 60% of the data to train the model, validate on 30% of sample (optimize over set of hyper-parameters) and compare the performance on the remaining 10%. The final performance (error) is average (standard error) over all ten draws.

In Table 8 we report out-of-sample U^2 measure (Hauser, 1978):

(20) $$U^2 = \frac{I(Y;X)}{H(Y)} = \frac{\sum_{i=1}^{N} \frac{1}{N} \sum_{a\in\{0,1\}} \mathbb{I}[Y_i = a] \cdot \ln \frac{\mathbb{P}[Y_i = a|X_i]}{\mathbb{P}[Y_i = a]}}{-\sum_{a\in\{0,1\}} \mathbb{P}[Y_i = a] \cdot \ln \mathbb{P}[Y_i = a]}$$

Where $\mathbb{P}[Y_i = a|X_i]$ is out-of-sample predictions of the model and $\mathbb{P}[Y_i = 1]$ is an estimate of unconditional return probability.

Intuitively, U^2 measures the proportion of information in Y_i explained by variable X_i. For example, if Y_i and X_i are independent then $\mathbb{P}[Y_i = a|X_i] = \mathbb{P}[Y_i = a]$ and thus $U^2 = 0$ which implies that there is no information about Y_i in X_i. Consider, on opposite, the case when Y_i is deterministic function of X_i and X_i perfectly explains Y_i, thus $\mathbb{P}[Y_i = a|X_i] = \mathbb{I}[Y_i = a]$ and $U^2 = 1$.[9] Therefore, U^2 perfectly fits our goal of demonstrating that search data contain incremental information about product returns.

Table 8. $100 \cdot U^2$ for estimated models

Model	Predictive out-of-sample accuracy (std.error)	Improvement over the benchmark
Purchase data	5.50 (0.0578)	–
Purchase + Search data	5.85 (0.0884)	6.41%

The results support our hypothesis. Search information improves prediction accuracy of the model by 6.41% ($p < 0.01$). The result provides a proof of concept of the value of search information. Further research could significantly improve the performance of the prediction model. For example, recurrent neural network could take into account the sequential nature of customer browsing session and capture higher order dependencies between search and

[9] Here we used a common convention that $0 \cdot \ln 0 = 0$ which is formally undefined and only $\lim_{x\to 0} x \cdot \ln x = 0$ is correct.

product returns. Development of sophisticated prediction model goes beyond the scope of this paper and we leave it for future research.

8. Summary

Returned items generate substantial costs to the retailer and significantly decrease the profitability in online shopping. This makes better understanding and managing product returns an important topic both for retailers and researchers. The existing literature on product returns focuses on return as a consequence of purchase while the search literature treats purchase as a terminal stage of customer journey. Our paper seeks a first step in uniting search, purchase, and return in the customer journey.

First, we find correlational evidence that search information is related to individual level return probability. Second, we develop a theoretically grounded model which combines search, purchase and return decisions of the customer. We demonstrate that the model is consistent with the model-free evidence and discuss the mechanism how search and returns are related. We show that with sufficient computational power the model can be estimated. Lastly, we validate the findings in applied scenario by showing that search data improves the prediction accuracy of returns on the real company data.

9. References

Anderson, E., Hansen, K., & Simester, D. (2009). The Option Value of Returns: Theory and Empirical Evidence. *Marketing Science, 28*(3), 405-423.

Bechwati, N., & Siegal, W. (2005). The Impact of the Prechoice Process on Product Returns. *Journal of Marketing Research, 42*, 358-367.

Bower, A., & Maxham-III, J. (2012). Return shipping policies of online retailers: Normative assumptions and the long-term consequences of fee and free returns. *Journal of Marketing, 76*(5), 110-124.

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

Che, Y.-K. (1996). Customer Return Policies for Experience Goods. *The Journal of Industrial Economics, 44*(2), 17-24.

Conlin, M., O'Donoghue, T., & Vogelsang, T. (2007). Projection Bias in Catalog Orders. *American Economic Review, 97*(4), 1217-1249.

Davis, S., Hagerty, M., & Gerstner, E. (1998). Return policies and the optimal level of "hassle". *Journal of Economics and Business, 50*(5), 445-460.

Dukes, A., & Liu, L. (2015). Online shopping intermediaries: the strategic design of search environments. *Management Science, 62*(4), 1064-1077.

Dzyabura, D., El Kihal, S., Hauser, J., & Ibragimov, M. (2020). Leveraging the Power of Images in Predicting Product Return Rates. *Working Paper*.

Guide, V., Daniel, R., & Luk, N. (2009). The Evolution of Closed-Loop Supply Chain Research. *Operations Research, 57*(1), 10-18.

Hauser, J. (1978). Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach. *Operations REsearch, 26*(3), 406-421.

Hess, J., & Mayhew, G. (1997). Modeling merchandise returns in direct marketing. *Journal of Direct Marketing, 11*(2), 20-35.

Janakiraman, N., Syrdal, H., & Freling, R. (2016). The Effect of Return Policy Leniency on Consumer Purchase and Return Decisions: A Meta-analytic Review. *Journal of Retailing, 92*(2), 226-235.

Ke, R., Shen, M., & Villas-Boas, M. (2016). Search for information on multiple products. *Management Science, 62*(12), 3576-3603.

Kim, J., Albuquerque, P., & Bronnenberg, B. (2010). Online Demand Under Limited Consumer Search. *Marketing Science, 29*(6), 963-1169.

Moorthy, S., & Srinivasan, K. (1995). Signaling quality with a money-back guarantee: The role of transaction costs. *Marketing Science, 14*(4), 442-466.

Narang, U., & Shankar, V. (2019). Mobile App Introduction and Online and Offline Purchases and Product Returns. *Marketing Science, 38*(5), 756-772.

Ofek, E., Katona, Z., & Sarvary, M. (2011). Bricks and clicks: The impact of product returns on the strategies of multichannel retailers. *Marketing Science, 30*(1), 42-60.

Petersen, J., & Kumar V. (2009). Are Product Returns a Necessary Evil? Antecedents and Consequences. *Journal of Marketing, 73*(3), 35-51.

Petersen, J., & Kumar, V. (2010). Can product returns make you money? *MIT Sloan Management Review, 51*(3), 85.

Sahoo, N., Dellarocas, C., & Srinivasan, S. (2018). The Impact of Online Product Reviews on Product Returns. *INFORMATION SYSTEMS RESEARCH, 29*(3), 723-738.

Shehu, E., Papies, D., & Neslin, S. (2020). Free Shipping Promotions and Product Returns. *Journal of Marketing Research, 57*(4), 640-658.

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

The Economist. (2013, December 21). *E-commerce firms have a hard core of costly, impossible-to-please customers*. Retrieved from The Economist: www.economist.com/news/business/21591874-e-commerce-firms-have-hard-core-costly-impossible-please-customers-return-santa

Thomasson, E. (2013, October 1). *Online retailers go hi-tech to size up shoppers and cut returns*. Retrieved from Reuters: http://www.reuters.com/article/net-us-retail-online-returns-idUSBRE98Q0GS20131002

Toktay, B., van der Laan, A., & de Brito, P. (2004). Managing Product Returns: The Role of Forecasting. *Reverse Logistics*, 45-64.

Ursu, R., Wang, Q., & Chntagunta, P. (2020). Search Duration. *Marketing Science, 39*(5), 849-871.

Wall Street Journal. (2018, June 11). *Banned from Amazon: The shoppers who make too many returns*. Retrieved from WSJ: https://www.wsj.com/articles/banned-from-amazon-the-shoppers-who-make-too-many-returns-1526981401

Weitzman, M. (1979). Optimal search for the best alternative. *Econometrica, 47*(3), 641-654.

10. Appendix

Appendix A

In Table A1 we present the robustness results for Model 1 in Section 3. Specifically, we estimate Probit and Linear Probabilistic models (OLS).

Table A1. Robustness tables for analysis in Section 3. Standard errors in parenthesis.

Variable	Probit	OLS	Model 1 (Logit)
Last item searched	*-0.0647*	*-0.0222*	*-0.1083*
	(0.011)	*(0.004)*	*(0.018)*
Number product searched	*0.0346*	*0.0118*	*0.0593*
	(1.780)	*(0.007)*	*(0.032)*
Applied filter	*-0.0485*	*-0.0165*	*-0.0803*
	(0.010)	*(0.003)*	*(0.017)*
Time spent on product	*-0.0279*	*-0.0099*	*-0.0475*
	(0.009)	*(0.003)*	*(0.015)*
Price (log)	*0.7965*	*0.2673*	*1.3072*
	(0.038)	*(0.012)*	*(0.063)*
Discount (log)	*-0.1658*	*-0.0578*	*-0.2705*
	(0.009)	*(0.003)*	*(0.015)*
Device FE	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*
Category FE	*Yes*	*Yes*	*Yes*
Color FE	*Yes*	*Yes*	*Yes*
Size FE	*Yes*	*Yes*	*Yes*
Material FE	*Yes*	*Yes*	*Yes*
Number points	*94,477*	*94,477*	*94,477*
Pseudo R^2	*0.0529*	*0.064*	*0.0529*

The results are consistent between the specifications in terms of the significance level and direction of the effect. The discussed findings do not depend on particular estimation methods.

In Table A2 we present the average marginal effects for variable k in set X_i defined as:

$$\text{(A1)} \qquad \frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mathbb{P}[Y_i = 1|Z_i, X_i]}{\partial X_{ki}} = \frac{1}{N}\sum_{i=1}^{N}\beta_k \cdot \mathbb{P}[Y_i = 1|Z_i, X_i] \cdot (1 - \mathbb{P}[Y_i = 1|Z_i, X_i])$$

These values measure the average change in probability that the customer would return the product if variable X_{ki} would be changed by 1.

Table A2. Average Marginal effects

Variable	Model 1	Model 2
Last item searched	*-0.0223*	*-0.0137*
Number product searched	*0.0122*	*0.0114*
Applied filter	*-0.0165*	*-0.0083*
Time spent on product	*-0.0098*	*-0.0157*

The results for marginal effects are similar to OLS model in Table A1 which suggest us that linear probabilistic model is a reasonable approximation for our setting.

Appendix B

The dynamic problem formulated in Equation (8) is not very computationally convenient as the value function depends on the set of variables H. Each variable in a set is random and talking expectation over set of random variables is computationally difficult as the number of items in a set is also random.

Notice that the expected purchase utility has the following property (we omit the customer index j for compactness):

$$\begin{aligned} \mathcal{U}(purchase|H \cup \{\eta\}) &= \max_{i \in \mathcal{I} \cup \{i'\}} \mathcal{U}(purchase\ i) = \\ &= max\left\{\max_{i \in \mathcal{I}} \mathcal{U}(purchase\ i), \quad \mathcal{U}(purchase\ i')\right\} = \\ &= max\{\mathcal{U}(purchase|H), \quad \mathcal{U}(purchase\ i')\} \end{aligned} \tag{A2}$$

Thus, we only need to compare the additional item in a set with the maximal element of the rest. This allows us to formulate the dynamic problem in more convenient form.

Consider the customer j who is in the middle of her search process and she already searched n items and thus generated a choice set H_n. Let's denote $\mathcal{V}(z,n)$ as a value function of search. That is the expected utility of the customer who already searched n items and her best alternative among this n items yield the expected purchase utility equal to z.

Therefore, the dynamic problem could be rewritten as:

$$\begin{aligned} &\mathcal{V}(z,n) = \\ &= max\{\mathbb{E}_{\eta'}[\mathcal{V}(max\{z,\eta'\},n)] - s_n, \quad \mathcal{U}(purchase|\mathcal{S})_j\} \\ &= max\{\mathbb{E}_{\eta'}[\mathcal{V}(max\{z, g(x',\hat{\psi}',\beta,c)\},n)] - s_n,\ z\} \end{aligned} \tag{A3}$$

Where $\eta' = (x',\hat{\psi}')$ is new product's characteristics and signal of fit; $g(x',\hat{\psi}',\beta,c)$ deterministic function of expected purchase utility of new item. The dynamic program in this form is significantly simpler as the value function is a function of only two variables both of which are numerical ones. Notice that η' is the only source of randomness and $g(\cdot)$ is deterministic function, therefore, under assumption that distribution of η' is known we can compute the distribution of $g' = g(x',\hat{\psi}',\beta,c)$ and denote it as $F_g(\cdot)$.

In this case the Equation (A3) could be rewritten as

$$\mathcal{V}(z,n) = $$
(A4)
$$= max\{\mathbb{E}_{g'}[\mathcal{V}(max\{z,g'\},n)] - s_n,\ z\}$$
$$= max\left\{F_g(z)\cdot\mathcal{V}(z,n+1) + \int_z^\infty \mathcal{V}(t,n+1)dF_g(t) - s_{nj},\ z\right\}$$

Appendix C

Remember we formulated a dynamic program in a form in Equation (A4). In this section we omit the customer specific index j and random variable g to ease notation. Thus, we solve:

$$\mathcal{V}(z,n) = $$
(A5)
$$= max\left\{F(z)\cdot\mathcal{V}(z,n+1) + \int_z^\infty \mathcal{V}(t,n+1)dF(t) - s_n,\ z\right\}$$

Where $F(\cdot)$ is cumulative distribution function of the reward g and s_n is strictly increasing function of n and $\lim_{n\to\infty} s_n = \infty$ (for example, $s_n = s\cdot n$)

Claim 1. *$\forall z,n\ \mathcal{V}(z,n)$ is non-increasing function of n.*

Let's consider some fixed z and n. Then the value function for the +1 step could be written as:

(A6)
$$\mathcal{V}(z,n+1) = $$
$$= max\left\{F(z)\cdot\mathcal{V}(z,n+2) + \int_z^\infty \mathcal{V}(t,n+2)dF(t) - s_{n+1},\ z\right\} \rightarrow$$
$$\tilde{\mathcal{V}}(z,n) = $$
$$= max\left\{F(z)\cdot\tilde{\mathcal{V}}(z,n+1) + \int_z^\infty \tilde{\mathcal{V}}(t,n+1)dF(t) - \tilde{s}_n,\ z\right\}$$

where $\tilde{s}_n = s_{n+1} = s_n + (s_{n+1} - s_n) = s_n + \Delta_n$.

Because we assume that s_n is strictly increasing function of n we have that for any n $\Delta_n > 0$. This implies that the problem in Equation (A6) is equivalent to the original dynamic problem but where each search cost is higher as $\Delta_n > 0$. Therefore, $\mathcal{V}(z,n) \geq \tilde{\mathcal{V}}(z,n) = \mathcal{V}(z,n+1)$. Because we chose z and n arbitrary we may conclude that for any pair n,z $\mathcal{V}(z,n) \geq \mathcal{V}(z,n+1)$.

This result is intuitive. Having fixed z implies that the customer is unlucky and samples products which are no better than the best from her choice set. Thus, after searching a new product the customer faces exactly the same problem, but she is getting tired of the search which makes a search decision more effort costly.

Claim 1.1. *If $\overline{\mathcal{V}}(z)$ is a value function for the model when $s_n = const = s_0$ then for any n and z we have $\overline{\mathcal{V}}(z) \geq \mathcal{V}(z,n+1)$.*

When costs are fixed the customer always faces the same problem and thus her search expected utility is constant. Equation (A5) simplifies to $\overline{\mathcal{V}}(z) = \max\{C,z\}$ where constant C could be found from $\overline{\mathcal{V}}(C) = C = F(C)\cdot \mathrm{C} + \int_C^\infty tdF(t) - \mathrm{s}_0$

Claim 2. *If z bounded from below by Z then there exists N such that for any $n \geq N$ and $z \geq Z$ we have $\mathcal{V}(z,n) = z$*

Consider the difference between search and purchase expected utilities for some n

$$F(z) \cdot \mathcal{V}(z,n+1) + \int_z^\infty \mathcal{V}(t,n+1)dF(t) - s_n - z \tag{A7}$$

From Claim 1.1 if follows that it is bounded from above

$$\begin{aligned} F(z) \cdot \mathcal{V}(z,n+1) + \int_z^\infty \mathcal{V}(t,n+1)dF(t) - s_n - z &\leq F(z) \cdot \overline{\mathcal{V}}(z) + \int_z^\infty \overline{\mathcal{V}}(t)dF(t) - z - s_n \\ &= F(z) \cdot max\{z, C\} + \int_z^\infty max\{t, C\}\, dF(t) - z - s_n \end{aligned} \tag{A8}$$

The final expression has two regimes for $z \geq C$ and $z < C$. We consider both of them and start with $Z \leq z < C$:

$$F(z) \cdot C + \int_z^C CdF(t) + \int_C^\infty tdF(t) - s_n - z = F(C) \cdot C + \int_C^\infty tdF(t) - z - s_n \leq C_1 - s_n \tag{A9}$$

Because C is known constant and z is bounded from below there exists some constant C_1 such that the Equation (A9) is bounded from above by $C_1 - s_n$. Next consider the case $z \geq C$:

$$F(z) \cdot z + \int_z^\infty tdF(t) - z - s_n \equiv Q(z) - s_n \tag{A10}$$

Where we introduced a new function $Q(z)$ which would be used in other sections of the proof. We prove that $Q(z)$ is monotonic by considering some $z_2 > z_1$ and $Q(z_1) - Q(z_2)$:

$$\begin{aligned} Q(z_1) - Q(z_2) &= F(z_1) \cdot z_1 - F(z_2) \cdot z_2 + \int_{z_1}^{z_2} t \cdot dF(t) - (z_1 - z_2) \\ &\geq F(z_1) \cdot z_1 - F(z_2) \cdot z_2 + z_1\big(F(z_2) - F(z_1)\big) - (z_1 - z_2) \\ &= F(z_2) \cdot (z_1 - z_2) - (z_1 - z_2) = \big(1 - F(z_2)\big)(z_2 - z_1) \geq 0 \end{aligned} \tag{A11}$$

Therefore, function $Q(z)$ is non-increasing and thus Equation (A10) could be bounded from above by $Q(C) - s_n$:

$$F(z) \cdot z + \int_z^\infty tdF(t) - z - s_n \leq Q(C) - s_n \tag{A12}$$

Combining Equations (A9) and (A12) we have that for any $z \geq Z$ and n:

$$F(z) \cdot \mathcal{V}(z,n+1) + \int_z^\infty \mathcal{V}(t,n+1)dF(t) - s_n - z \leq max\{C_1, Q(C)\} - s_n \tag{A13}$$

Finally, because $\max\{C_1, Q(C)\}$ is constant and $\lim_{n\to\infty} s_n = \infty$ we have that there exists N such that for any $n \geq N$

(A14) $$F(z)\cdot\mathcal{V}(z,n+1)+\int_z^\infty \mathcal{V}(t,n+1)dF(t)-s_n-z\le max\{C_1,Q(C)\}-s_n<0$$

The Equation (A14) implies that for any $n\ge N$ we have $\mathcal{V}(z,n)=z$.

The results imply that regardless of customer's current best option z she will eventually finish her search even she won't find anything better. Notice that we can treat Z as and outside option without loss of generality we assume that it is equal to 0.

Claim 3. *The solution of dynamic program in Equation (A5) is in a form of a threshold policy. For each number of searched products* n *there exist a threshold* z_n^* *such that*

(A15) $$\mathcal{V}(z,n)=\begin{cases} z, & z\ge z_n^* \\ F_g(z)\cdot\mathcal{V}(z,n+1)+\int_z^\infty \mathcal{V}(t,n+1)dF_g(t)-s_{nj}, & z<z_n^*\end{cases}$$

Moreover, thresholds are decreasing in n *and could be found from the equation:*

(A16) $$F(z)\cdot z+\int_z^\infty t\cdot dF(t)-z=s_n$$

Let's denote N such that for any $n\ge N$ and any z we have $\mathcal{V}(z,n)=z$. Thus, the decision at state $n-1$ is based on

(A17) $$\begin{aligned}\mathcal{V}(z,n-1)&=max\left\{F(z)\cdot\mathcal{V}(z,n)+\int_z^\infty \mathcal{V}(t,n)dF(t)-s_{n-1},z\right\}\\ &=max\left\{F(z)\cdot z+\int_z^\infty z\cdot dF(t)-s_{n-1},z\right\}\end{aligned}$$

The change of decision points could be found from the equation:

(A18) $$Q(z)\equiv F(z)\cdot z+\int_z^\infty z\cdot dF(t)-z=s_{n-1}$$

Assume $z_2>z_1$ and consider $Q(z_1)-Q(z_2)$:

(A19) $$\begin{aligned}Q(z_1)-Q(z_2)&=F(z_1)\cdot z_1-F(z_2)\cdot z_2+\int_{z_1}^{z_2} t\cdot dF(t)-(z_1-z_2)\\ &\ge F(z_1)\cdot z_1-F(z_2)\cdot z_2+z_1\big(F(z_2)-F(z_1)\big)-(z_1-z_2)\\ &=F(z_2)\cdot(z_1-z_2)-(z_1-z_2)=\big(1-F(z_2)\big)(z_2-z_1)\ge 0\end{aligned}$$

Which implies that the function $Q(z)$ is non-increasing. Notice that $\lim_{z\to\infty} Q(z)=0$ and $s_{n-1}>0$.

Therefore, the Equation (A18) either has a unique solution or $Q(z)<s_{n-1}$ for any z. Thus, if the (A18) does not have a solution we have $\mathcal{V}(z,n-1)=z$ for any z and the analysis for $\mathcal{V}(z,n-2)$ is identical. However, if $Q(z_{n-1}^*)=s_{n-1}$ we can write $\mathcal{V}(z,n-1)$ as:

(A20) $$\mathcal{V}(z,n-1)=\begin{cases} z, & z\ge z_{n-1}^* \\ F(z)\cdot z+\int_z^\infty z\cdot dF(t)-s_{n-1}, & z<z_{n-1}^*\end{cases}$$

The Equation (A20) implies that if after the next product the customer would choose to finish search then the decision at current state would have a threshold form. Now we show that at any state n this would have a threshold policy.

Let's assume that $\mathcal{V}(z,n)$ is in a threshold form where threshold $z_n^* = Q^{-1}(s_n)$:

$$\mathcal{V}(z,n) = \begin{cases} z, & z \geq z_n^* \\ F(z) \cdot \mathcal{V}(z,n+1) + \int_z^\infty \mathcal{V}(z,n+1) \cdot dF(t) - s_n, & z < z_n^* \end{cases} \quad \text{(A21)}$$

We want to find $\mathcal{V}(z,n-1)$ and start by considering $z \geq z_n^*$ where the form of $\mathcal{V}(z,n)$ is known:

$$\begin{aligned} \mathcal{V}(z,n-1) &= max\left\{F(z) \cdot \mathcal{V}(z,n) + \int_z^\infty \mathcal{V}(t,n) dF(t) - s_{n-1}, z\right\} \\ &= max\left\{F(z) \cdot z + \int_z^\infty z \cdot dF(t) - s_{n-1}, z\right\} \end{aligned} \quad \text{(A22)}$$

Notice that this form is exactly the same as Equation (A17) and because $s_{n-1} < s_n$ we have that $z_{n-1}^* = Q^{-1}(s_{n-1}) \geq z_n^*$. This implies that for $z \geq z_n^*$ the $\mathcal{V}(z,n-1)$ is threshold function:

$$\mathcal{V}(z,n-1) = \begin{cases} z, & z \geq z_{n-1}^* \\ F(z) \cdot z + \int_z^\infty z \cdot dF(t) - s_n, & z_n^* \leq z < z_{n-1}^* \end{cases} \quad \text{(A23)}$$

Next, we consider the interval $z < z_n^*$ and the difference between terms in $\max$ function:

$$\begin{aligned} F(z) \cdot \mathcal{V}(z,n) + \int_z^\infty \mathcal{V}(t,n) dF(t) - s_{n-1} - z &\geq F(z) \cdot z + \int_z^\infty t \cdot dF(t) - s_{n-1} - z \\ &= Q(z) - s_{n-1} = Q(z) - s_n + (s_n - s_{n-1}) > 0 \end{aligned} \quad \text{(A24)}$$

The last inequality follows from the fact that when $z < z_n^*$ we have $Q(z) - s_n \geq 0$ and $s_n - s_{n-1} > 0$. This implies that for $z < z_n^*$ we always have customer choosing to search over the purchase.

Combining Equation (A23) and (A24) we conclude that $\mathcal{V}(z,n-1)$ would have a threshold form:

$$\mathcal{V}(z,n-1) = \begin{cases} z, & z \geq z_{n-1}^* \\ F(z) \cdot \mathcal{V}(z,n) + \int_z^\infty \mathcal{V}(z,n) \cdot dF(t) - s_{n-1}, & z < z_{n-1}^* \end{cases} \quad \text{(A25)}$$

We can continue the chain until $n = 0$ and always have that $\mathcal{V}(z,n)$ is a threshold function. Moreover, all thresholds could be found from equation:

$$Q(z) \equiv F(z) \cdot z + \int_z^\infty z \cdot dF(t) - z = s_{n-1} \quad \text{(A26)}$$

Appendix D

Let's denote the random variable $\xi \equiv x_i'\beta_j + \psi_{ij} + c_j$ and thus we want to find $\mathbb{E}\left[\max\{\xi, 0\} \,| x_i, \hat{\psi}_{ij}\right] - c_j$. We denote the c.d.f of distribution $\psi_{ij}|\hat{\psi}_{ij}$ as $F_{\psi_{ij}|\hat{\psi}_{ij}}(\psi_{ij})$ and assume that there also exists the p.d.f of this distribution. By assumption ψ_{ij} and $\hat{\psi}_{ij}$ are independent of x_i, thus distribution $\psi_{ij}|x_i, \hat{\psi}_{ij}$ is equal to distribution $\psi_{ij}|\hat{\psi}_{ij}$.

Because the customer knows β_j, c_j and x_i then conditional on x_i the expression $x_i'\beta_j + c_j$ is some constant. Therefore, we can find the distribution $\xi|x_i, \hat{\psi}_{ij}$ as

$$\begin{aligned}\mathbb{P}\left[\xi \le t \middle| x_i, \hat{\psi}_{ij}\right] &= \mathbb{P}\left[x_i'\beta_j + \psi_{ij} + c_j \le t \middle| x_i, \hat{\psi}_{ij}\right] = \mathbb{P}\left[\psi_{ij} \le t - x_i'\beta_j - c_j \middle| x_i, \hat{\psi}_{ij}\right] \\ &= F_{\psi_{ij}|\hat{\psi}_{ij}}\left(t - x_i'\beta_j - c_j\right) = F_\xi(t)\end{aligned} \tag{A27}$$

We omit the conditioning for clarity and remember that F_ξ is some function of $x_i, \hat{\psi}_{ij}, \beta_j, c_j$. Thus, we can omit the conditioning in expectation and find $\mathbb{E}[\xi^+]$ according to given distribution $F_\xi(t)$

$$\mathbb{E}[max\{\xi, 0\}] = \int_{-\infty}^{+\infty} max\{0, t\} \cdot dF_\xi(t) = \int_0^{+\infty} t \cdot dF_\xi(t) \tag{A28}$$

In Section 4 we introduced the distributional assumptions and here we develop the Equation (A28) under such assumptions.

Table A3. Distributional model assumptions.

$\hat{\psi}_{ij} = \psi_{ij} + \nu_{ij}$	signal of product fit is unbiased
$\psi_{ij} \perp \nu_{ij} \perp x_{ij}$	the noise of the signal; true product fit and product characteristics are independent
$\psi_{ij} \sim \mathcal{N}(0, \sigma_\psi)$ $\nu_{ij} \sim \mathcal{N}(0, \sigma_\nu)$	noise and product fit come form a joint normal distribution

Because ψ_{ij} and ν_{ij} are independent the distribution of vector (ψ_{ij}, ν_{ij}) is joint normal with diagonal variance matrix Σ.

Under properties of normal distribution, we can find the distribution:

$$\begin{bmatrix}\psi_{ij} \\ \hat{\psi}_{ij}\end{bmatrix} = \begin{bmatrix}1 & 0 \\ 1 & 1\end{bmatrix}\begin{bmatrix}\psi_{ij} \\ \nu_{ij}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}1 & 0 \\ 1 & 1\end{bmatrix}\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}1 & 0 \\ 1 & 1\end{bmatrix}^T \Sigma \begin{bmatrix}1 & 0 \\ 1 & 1\end{bmatrix}\right) \sim \mathcal{N}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_\psi^2 & \sigma_\psi^2 \\ \sigma_\psi^2 & \sigma_\psi^2 + \sigma_\nu^2\end{bmatrix}\right) \tag{A29}$$

Thus, we can calculate the conditional distribution $\psi_{ij}|\hat{\psi}_{ij}$:

$$\psi_{ij}|\hat{\psi}_{ij} \sim \mathcal{N}\left(\frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}\hat{\psi}_{ij}, \frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}\right) = \mathcal{N}\left(\mu_{\psi|\hat{\psi}}, \sigma^2_{\psi|\hat{\psi}}\right) \tag{A30}$$

Finally, we can find the distribution of random variable ξ[10]

[10] where $\varphi(\cdot)$ and $\Phi(\cdot)$ are standard normal p.d.f and c.d.f respectively.

(A31) $$F_{\psi_{ij}|\hat{\psi}_{ij}}(t) = \Phi\left(\frac{t-\mu_{\psi|\hat{\psi}}}{\sigma_{\psi|\hat{\psi}}}\right) \Rightarrow F_{\xi}(t) = \Phi\left(\frac{t - x_i'\beta_j - c_j - \mu_{\psi|\hat{\psi}}}{\sigma_{\psi|\hat{\psi}}}\right)$$

Which implies that given $x_i, \hat{\psi}_{ij}$ the distribution ξ is normal with mean $\mu_\xi = x_i'\beta_j + c_j + \mu_{\psi|\hat{\psi}}$ and variance $\sigma_\xi^2 = \sigma_{\psi|\hat{\psi}}^2$. We can find the $\mathbb{E}[\xi^+]$ now:

(A32)
$$\mathbb{E}[\xi^+] = \frac{1}{\sqrt{2\pi}\cdot\sigma_\xi}\int_0^{+\infty} t\cdot e^{-\frac{(t-\mu_\xi)^2}{2\cdot\sigma_\xi^2}}dt$$
$$= \frac{1}{\sqrt{2\pi}\cdot\sigma_\xi}\int_0^{+\infty}(t-\mu_\xi)\cdot e^{-\frac{(t-\mu_\xi)^2}{2\cdot\sigma_\xi^2}}dt + \frac{\mu_\xi}{\sqrt{2\pi}\cdot\sigma_\xi}\int_0^{+\infty}\cdot e^{-\frac{(t-\mu_\xi)^2}{2\cdot\sigma_\xi^2}}dt$$
$$= \sigma_\xi\cdot\varphi\left(\frac{\mu_\xi}{\sigma_\xi}\right) + \mu_\xi\cdot\Phi\left(\frac{\mu_\xi}{\sigma_\xi}\right)$$

Finally, under normality assumptions we find the expected purchase utility in Equation (A28) as:

(A33)
$$\mathbb{E}\left[(\xi)^+|x_i,\hat{\psi}_{ij}\right] - c_j = \sigma_\xi\cdot\left(\varphi(\alpha) + \alpha\cdot\Phi(\alpha)\right) - c_j$$
$$\alpha = \frac{x_i'\beta_j + c_j + \frac{\sigma_\psi^2}{\sigma_\psi^2+\sigma_\nu^2}\hat{\psi}_{ij}}{\sqrt{\frac{\sigma_\psi^2\cdot\sigma_\nu^2}{\sigma_\psi^2+\sigma_\nu^2}}}$$
$$\sigma_\xi^2 = \frac{\sigma_\psi^2\cdot\sigma_\nu^2}{\sigma_\psi^2+\sigma_\nu^2}$$

Notice that the (A33) could be written in a form $\sigma_\xi \cdot Q(\alpha) - c_j$, where $Q(\alpha) = \varphi(\alpha) + \alpha\cdot\Phi(\alpha)$ is know function. By taking the derivative of this function we can see that the function $Q(\alpha)$ is strictly increasing function of α.

Appendix E

As we mentioned at the final subsection of Section 4 we assume that the search model is not available. The customer just chooses the best alternative from the choice set. We use the standard notation:

Table A4. Notation for the Appendix E

Variable	Description
$\mathcal{G}_j$	$\{g_i : i = 1 \ldots N\}$ the sequence of expected purchase utilities of searched products by the customer j
z_n^*	search decision threshold after n were searcehd
g_0	outside option of the customer

g_i	search expected utility of purchasing product i. Note that it is actually a function of observable to customer characteristics of the product $g(x_i, \hat{\psi}_{ij})$ we omit arguments to simplify notation.
z_n	$\max_{i \in \{0..n\}} g_i$ or maximal expected purchase utility by the time customer searched n items.
$\mathcal{X}_j$	$\{x_i : i = 1 \dots N_j\}$ the sequence of product features of searched products by the customer j. We assume it is ordered sequence.
θ_j	$(g_0, \beta_j, \sigma_{\psi j}, \sigma_{\eta j}, c_j, s_{nj})$ type of the customer j or set of parameters of the model

In this case the purchase decision could be summarized in Equation (A34) which takes 1 if customer purchased item n and 0 otherwise:

$$\mathbb{I}\left[g_n \geq \max_{k=1\dots N}\{g_k\}\right] = \prod_{k=1}^{N} \mathbb{I}[g_n \geq g_k] \tag{A34}$$

Remember that the return decision is based on the realization of true unobserved fit ψ_{ij}. Specifically, the customer returns the item if her utility of possessing the product is lower than the negative return costs.

$$\mathbb{I}\left[x_i'\beta_j + \psi_{ij} \leq -c_j\right] \tag{A35}$$

Under the normality assumption which we imposed earlier in section 4 the search expected utility of purchasing product g_i has a form of:

$$g_i = g(x_i, \hat{\psi}_i) = \sigma_\xi \cdot (\varphi(\alpha) + \alpha \cdot \Phi(\alpha)) - c$$

$$\alpha = \frac{x_i'\beta + c + \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}\hat{\psi}_i}{\sqrt{\frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}}} \tag{A36}$$

$$\sigma_\xi^2 = \frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}$$

Because $Q(\alpha) = \varphi(\alpha) + \alpha \cdot \Phi(\alpha)$ is a strictly increasing function of α there exists and inverse function $Q^{-1}(\cdot)$. Therefore, when x_i is fixed the function $g(x_i, \hat{\psi}_i)$ is inversible for $\hat{\psi}_i$ and we denote it as $h_{x_i}(\cdot)$. This implies that the condition in Equation (A34) could be rewritten in more convenient form:

$$\text{(A37)}\quad \begin{aligned}\mathbb{I}\left[g(x_i,\hat{\psi}_i)\le g(x_j,\hat{\psi}_j)\right] &\Leftrightarrow \mathbb{I}\left[x_i'\beta+\frac{\sigma_\psi^2}{\sigma_\psi^2+\sigma_v^2}\hat{\psi}_i\le x_j'\beta+\frac{\sigma_\psi^2}{\sigma_\psi^2+\sigma_v^2}\hat{\psi}_j\right]\\ &\Leftrightarrow \mathbb{I}\left[\frac{\sigma_\psi^2+\sigma_v^2}{\sigma_\psi^2}(x_i'-x_j')\beta+\hat{\psi}_i\le\hat{\psi}_j\right]\end{aligned}$$

Therefore, we can compute the probability of an event that customer purchases product n and returns it as:

$$\text{(A38)}\quad \begin{aligned}&\mathbb{P}[purchase\ n\ and\ return\ it|\mathcal{X}]=\\ &\iiint \mathbb{I}\left[x_i'\beta_j+\tilde{t}\le -c_j\right]\cdot\prod_{k=1}^{N}\mathbb{I}[g(x_k,t_k)\le g(x_n,t_n)]\cdot dF_{\psi_n|\hat{\psi}_n,x_n}(\tilde{t})\prod_{k=1}^{N}dF_{\hat{\psi}_k|x_k}(t_k)=\\ &\iiint F_{\psi_n|\hat{\psi}_n,x_n}(-x_n'\beta-c)\cdot\prod_{k=1}^{N}\mathbb{I}\left[\frac{\sigma_\psi^2+\sigma_v^2}{\sigma_\psi^2}(x_k'-x_n')\beta+t_k\le t_n\right]\prod_{k=1}^{N}dF_{\hat{\psi}_k|x_k}(t_k)=\\ &\int_{-\infty}^{\infty}F_{\psi_n|\hat{\psi}_n,x_n}(-x_n'\beta-c)\cdot\prod_{k=1,\mathrm{k\neq n}}^{N}F_{\hat{\psi}_k|x_k}\left(t_n-\frac{\sigma_\psi^2+\sigma_v^2}{\sigma_\psi^2}(x_k'-x_n')\beta\right)\cdot dF_{\hat{\psi}_n|x_n}(t_n)\end{aligned}$$

Appendix F

Simulation parameters for Figure 1 where we varied the return cost c. For each point on the graph, we simulated 100,000 identical customers where each of them received unique choice set.

Table A5. Simulation parameters for Figure 1

Parameter	Value	Comment
c	$[0.1,\dots,2]$	20 values on the $[0.1,2]$ interval
s_n	$0.01\cdot n$	linear search search costs
β	$[-1]$	1-dimensional vector of customer preferences
x	$[8]$	1-dimensional vector of product characteristics
σ_η	5.0	quality of the signal
σ_ψ	5.0	variance of the unobserved fit

For **Figure 2** to model customer choice of pre-search filters, we consider simplified setting where each product has only one attribute and unobserved fit. Customer likes this attribute and can choose to sample from distribution where all products have this attribute. Simulation parameters for **Figure 2** where we varied the preference of the customer for an attribute.

Table A6. Simulation parameters for **Figure 2**

Parameter	Value	Comment
c	1	return costs
s_n	$0.01\cdot n$	linear search search costs
β	$[-1,3]$	2-dimensional vector of customer preferences

x	$[8, x_a]$	2-dimensional vector of product characteristics. x_a is binary random variable with $\mathbb{P}[x_a = 1] = 0.5$
σ_η	5.0	quality of the signal
σ_ψ	5.0	variance of the unobserved fit

This implies that customer who applied filter would have $\mathbb{P}[x_a = 1] = 1$

Appendix G

In Section 6 we introduced the notation

Table A7. Notation for Appendix G

Variable	Description
$\mathcal{G}_j$	$\{g_i: i = 1 \dots N\}$ the sequence of expected purchase utilities of searched products by the customer j
z_n^*	search decision threshold after n were searcehd
g_0	outside option of the customer
g_i	search expected utility of purchasing product i. Note that it is actually a function of observable to customer characteristics of the product $g(x_i, \hat{\psi}_{ij})$ we omit arguments to simplify notation.
z_n	$\max\limits_{i \in \{0..n\}} g_i$ or maximal expected purchase utility by the time customer searched n items.
$\mathcal{X}_j$	$\{x_i: i = 1 \dots N_j\}$ the sequence of product features of searched products by the customer j. We assume it is ordered sequence.
θ_j	$(g_0, \beta_j, \sigma_{\psi j}, \sigma_{\eta j}, c_j, s_{nj})$ type of the customer j or set of parameters of the model

Throughout this section we omit the customer index j and treat $\theta_j = \theta$ as fixed to ease notation. We start with the probability of purchase of item i given the set of products searched by the customer $\mathcal{X}_j = \mathcal{X}$. Note that the researcher does not observe $\mathcal{G}_j = \mathcal{G}$ as the product fit signals $\hat{\psi}_i$ are not observed by the researcher.

Assume the customer purchased the item n and her session length was $N = |\mathcal{X}|$ products. The condition of purchasing item i could be described by three constraints summarized in Table A8.

Table A8. Constraints defining purchasing decision

Constraint	Comment
$z_N \geq z_N^*$	maximal expected purchase utility at last item was greater than the threshold
$\forall n = \{1 \dots N-1\}$ $z_n < z_n^*$	the search process was not stopped before the item N
$g_n \geq z_N$	the chosen product yielded the highest expected utility

We combine all these constraints in Equation (A39).

(A39) $$\mathbb{I}[z_N \geq z_N^*] \cdot \left(\prod_{i=1}^{N-1} \mathbb{I}[z_i < z_i^*]\right) \cdot \mathbb{I}[g_n \geq z_N]$$

This variable summarizes the deterministic purchase decision for the customer because she observes all the components. Next, we simplify the equation above.

Firstly, notice that $\forall n = \{1 \dots N-1\}$ we have[11]

(A40) $$\mathbb{I}[z_n < z_n^*] = \mathbb{I}\left[\max_{k=0\dots n} g_k < z_n^*\right] = \mathbb{I}[g_0 < z_n^*] \cdot \prod_{k=1}^{n} \mathbb{I}[g_k < z_n^*]$$

Thus, we can rewrite the product in the middle as

(A41) $$\begin{aligned}\prod_{n=1}^{N-1}\left(\mathbb{I}[g_0 < z_n^*]\prod_{k=1}^{n}\mathbb{I}[g_k < z_n^*]\right) &= \left(\prod_{n=1}^{N-1}\mathbb{I}[g_0 < z_n^*]\right)\cdot\left(\prod_{k=1}^{N-1}\prod_{n=k}^{N-1}\mathbb{I}[g_k < z_n^*]\right)\\ &= \left(\mathbb{I}\left[g_0 < \min_{n=1\dots N-1} z_n^*\right]\right)\cdot\left(\prod_{k=1}^{N-1}\mathbb{I}\left[g_k < \min_{n=k\dots N-1} z_n^*\right]\right)\\ &= \mathbb{I}[g_0 < z_{N-1}^*]\cdot\prod_{k=1}^{N-1}\mathbb{I}[g_k < z_{N-1}^*]\end{aligned}$$

Where the last equation follows from the fact that the threshold z_n^* is a decreasing function of n. Next, we look at the other two conditions:

(A42) $$\begin{aligned}\mathbb{I}[z_N \geq z_N^*]\cdot\mathbb{I}[g_n \geq z_N] &= \mathbb{I}\left[\max_{i\in\{0..N\}} g_i \geq z_N^*\right]\cdot\mathbb{I}\left[g_n \geq \max_{i\in\{0..N\}} g_i\right] =\\ &= \mathbb{I}\left[\max_{i\in\{0..N\}} g_i \geq z_N^*\right]\cdot\prod_{k=0}^{N}\mathbb{I}[g_n \geq g_k] = \mathbb{I}[g_n \geq z_N^*]\cdot\prod_{k=0}^{N}\mathbb{I}[g_n \geq g_k]\end{aligned}$$

Finally, after substituting Equation (A42) into (A39) we get the purchase condition

(A43) $$\mathbb{I}[g_0 < z_{N-1}^*]\cdot\left(\prod_{k=1}^{N-1}\mathbb{I}[g_k < z_{N-1}^*]\right)\cdot\mathbb{I}[g_n \geq z_N^*]\cdot\left(\prod_{k=0}^{N}\mathbb{I}[g_n \geq g_k]\right)$$

We simplify this equation by considering different cases. Specifically, if customer did not purchase anything or $n = 0$:

(A44) $$\begin{aligned}&\mathbb{I}[z_N^* \leq g_0 < z_{N-1}^*]\cdot\left(\prod_{k=1}^{N-1}\mathbb{I}[g_k < z_{N-1}^*]\right)\cdot\left(\prod_{k=0}^{N}\mathbb{I}[g_0 \geq g_k]\right)\\ &= \mathbb{I}[z_N^* \leq g_0 < z_{N-1}^*]\cdot\prod_{k=1}^{N}\mathbb{I}[g_0 \geq g_k]\end{aligned}$$

If customer purchased the last item or $n = N$:

[11] We use $\mathbb{I}[g_0 < z_n^*]$ as separate term to emphasize that g_0 is not random variable and known to customer.

(A45) $$\mathbb{I}[g_0 < z^*_{N-1}] \cdot \left(\prod_{k=1}^{N-1} \mathbb{I}[g_k < min\{z^*_{N-1}, g_N\}]\right) \cdot \mathbb{I}[g_N \geq max\{z^*_N, g_0\}]$$

And for $n \neq N$:

(A46) $$\mathbb{I}[g_0 < z^*_{N-1}] \cdot \left(\prod_{k=1,k\neq n}^{N-1} \mathbb{I}[g_k < min\{z^*_{N-1}, g_n\}]\right) \cdot \mathbb{I}[max\{z^*_N, g_0\} \leq g_n < z^*_{N-1}] \cdot \mathbb{I}[g_N \leq g_n]$$

Notice that condition is slightly different in the case when the customer purchased the last product she looked at. This reflects the difference between customer who stopped searching because she is tired and the one who found an item which gave her very huge utility. Mathematically, for latter customer we can't infer the upper bound on utility of purchased product.

Our model takes into account not only the purchase decision but also a return one. We model it as joint random variable. That is, it could take $2 \cdot N + 1$ possible values – each product in a choice could be purchased and returned or the customer just buys the outside option. Next, we formalize the return decision. Remember that the return decision is based on the realization of true unobserved fit ψ_{ij}. Specifically, the customer returns the item if her utility of possessing the product is lower than the negative return costs.

(A47) $$\mathbb{I}[\mathrm{x}_\mathrm{i}'\beta + \psi_\mathrm{i} \leq -\mathrm{c}]$$

Remember that given the customer type $g_i = g(x_i, \hat{\psi}_i)$ is some function of product characteristics and a signal of unobserved fit. Both the signal $\hat{\psi}_i$ and unobserved fit variable ψ_i are not observable to the researcher. Therefore, to find the probability of purchase given the set of observable variables $\mathcal{X}_i$ we need to integrate over the random variable $\hat{\psi}_i$ and ψ_i.

In order to construct the likelihood function, we impose the same distributional assumptions as in Section 4.

Table A9. Distributional assumption of the model

$\hat{\psi}_{ij} = \psi_{ij} + \nu_{ij}$	signal of product fit is unbiased
$\psi_{ij} \perp \nu_{ij} \perp x_{ij}$	the noise of the signal; true product fit and product characteristics are independent
$\psi_{ij} \sim \mathcal{N}(0, \sigma_\psi)$ $\nu_{ij} \sim \mathcal{N}(0, \sigma_\nu)$	noise and product fit come form a joint normal distribution

Under these assumptions search expected utility of purchasing product g_i has a form of:

(A48) $$g_i = g(x_i, \hat{\psi}_i) = \sigma_\xi \cdot (\varphi(\alpha) + \alpha \cdot \Phi(\alpha)) - c$$

$$\alpha = \frac{x_i'\beta + c + \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}\hat{\psi}_i}{\sqrt{\frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}}}$$

$$\sigma_\xi^2 = \frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}$$

Because $A(\alpha) = \varphi(\alpha) + \alpha \cdot \Phi(\alpha)$ is a strictly increasing function of α there exists and inverse function $A^{-1}(\cdot)$. Therefore, when x_i is fixed the function $g(x_i, \hat{\psi}_i)$ is inversible for $\hat{\psi}_i$ and we denote it as $h_{x_i}(\cdot)$. Because $A(\alpha)$ is monotonic than $A^{-1}(\cdot)$ is also monotonic.

(A49)
$$\begin{gathered} g(x_i, \hat{\psi}_i) = \sigma_\xi \cdot \left(\varphi(\alpha) + \alpha \cdot \Phi(\alpha)\right) - c = t \Leftrightarrow A(\alpha) = \frac{t + c}{\sigma_\xi} \Leftrightarrow \\ \Leftrightarrow \frac{x_i'\beta + c + \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}\hat{\psi}_i}{\sigma_\xi} = A^{-1}\left(\frac{t + c}{\sigma_\xi}\right) \Leftrightarrow \\ \Leftrightarrow x_i'\beta + c + \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}\hat{\psi}_i = \sigma_\xi \cdot A^{-1}\left(\frac{t + c}{\sigma_\xi}\right) \Leftrightarrow \\ \Leftrightarrow \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_\nu^2}\hat{\psi}_i = \sigma_\xi \cdot A^{-1}\left(\frac{t + c}{\sigma_\xi}\right) - c - x_i'\beta \Leftrightarrow \\ \Leftrightarrow \hat{\psi}_i = \frac{\sigma_\psi^2 + \sigma_\nu^2}{\sigma_\psi^2} \cdot \left(\sigma_\xi \cdot A^{-1}\left(\frac{t + c}{\sigma_\xi}\right) - c - x_i'\beta\right) = h_{x_i}(t). \end{gathered}$$

This implies that the condition in Equation (A46) could be rewritten in more convenient form:

(A50)
$$\begin{gathered} \mathbb{I}\left[g(x_i, \hat{\psi}_i) \le min\{z_{N-1}^*, g(x_j, \hat{\psi}_j)\}\right] \Leftrightarrow \mathbb{I}\left[\hat{\psi}_i \le h_{x_i}\left(min\{z_{N-1}^*, g(x_j, \hat{\psi}_j)\}\right)\right] \\ \Leftrightarrow \mathbb{I}\left[\hat{\psi}_i \le min\left\{h_{x_i}(z_{N-1}^*), h_{x_i}\left(g(x_j, \hat{\psi}_j)\right)\right\}\right] \\ \Leftrightarrow \mathbb{I}\left[\hat{\psi}_i \le min\left\{h_{x_i}(z_{N-1}^*), \frac{\sigma_\psi^2 + \sigma_\nu^2}{\sigma_\psi^2} \cdot (x_j'\beta - x_i'\beta) + \hat{\psi}_j\right\}\right] \end{gathered}$$

Notice that is possible under assumption that all signal are independent of product characteristics. Potentially, it could be the case that the variance of the signal is different in different categories the extension is quite straightforward and we leave it to future research.

We can calculate the probability of possible events by integrating over $\psi, \hat{\psi}$. The probability of the event the customer purchase item $n = N$ and returns it (we omit feasibility constraint for compactness $\mathbb{I}[g_0 < z_{N-1}^*]$):

$$\mathbb{P}[purchase\ N\ and\ return\ it|\mathcal{X}]$$

$$\text{(A51)}\quad \begin{aligned} &= \iiint \mathbb{I}[x_N'\beta + \tilde{t} \leq -c] \cdot \left(\prod_{k=1}^{N-1} \mathbb{I}[g(x_k,t_k) < min\{z_{N-1}^*, g(x_N,t_N)\}]\right) \\ &\cdot \mathbb{I}[g(x_k,t_k) \geq max\{z_N^*, g_0\}] \cdot dF_{\psi_N|\widehat{\psi}_N,x_N}(\tilde{t}) \prod_{k=1}^{N} dF_{\widehat{\psi}_k|x_k}(t_k) \\ &= \iiint F_{\psi_N|\widehat{\psi}_N,x_N}(-x_N'\beta - c) \cdot \prod_{k=1}^{N-1} \mathbb{I}\left[t_k < h_{x_k}(min\{z_{N-1}^*, g(x_N,t_N)\})\right] \\ &\cdot \mathbb{I}\left[t_k \geq h_{x_k}(max\{z_N^*, g_0\})\right] \prod_{k=1}^{N} dF_{\widehat{\psi}_k|x_k}(t_k) \\ &= \int_{h_{x_N}(max\{z_N^*,g_0\})}^{\infty} F_{\psi_N|\widehat{\psi}_N,x_N}(-x_N'\beta - c) \\ &\cdot \prod_{k=1}^{N-1} F_{\widehat{\psi}_k|x_k}\left(h_{x_k}(min\{z_{N-1}^*, g(x_N,t_N)\})\right) \cdot dF_{\widehat{\psi}_N|x_N}(t_N) \end{aligned}$$

The probability of the event the customer purchase item $n \neq N$ and returns it (we omit feasibility constraint for compactness $\mathbb{I}[g_0 < z_{N-1}^*]$):

$$\mathbb{P}[purchase\ n\ and\ return\ it|\mathcal{X}]$$

$$\text{(A52)}\quad \begin{aligned} &= \iiint \mathbb{I}[x_n'\beta + \tilde{t} \leq -c] \cdot \prod_{k=1,k\neq n}^{N-1} \mathbb{I}[g_k < min\{z_{N-1}^*, g_n\}] \\ &\cdot \mathbb{I}[max\{z_N^*, g_0\} \leq g_n < z_{N-1}^*] \cdot \mathbb{I}[g_N \leq g_n] \\ &\cdot dF_{\psi_N|\widehat{\psi}_N,x_N}(\tilde{t}) \prod_{k=1}^{N} dF_{\widehat{\psi}_k|x_k}(t_k) \\ &= \iiint F_{\psi_n|\widehat{\psi}_n,x_n}(-x_n'\beta - c) \cdot \prod_{k=1,k\neq n}^{N-1} \mathbb{I}\left[t_k < h_{x_k}(min\{z_{N-1}^*, g(x_n,t_n)\})\right] \\ &\cdot \mathbb{I}\left[h_{x_n}(z_{N-1}^*) > t_k \geq h_{x_n}(max\{z_N^*, g_0\})\right] \cdot \prod_{k=1}^{N} dF_{\widehat{\psi}_k|x_k}(t_k) = \\ &= \int_{h_{x_n}(max\{z_N^*,g_0\})}^{h_{x_n}(z_{N-1}^*)} F_{\psi_n|\widehat{\psi}_n,x_n}(-x_n'\beta - c) \\ &\cdot \prod_{k=1,k\neq n}^{N-1} F_{\widehat{\psi}_k|x_k}\left(h_{x_k}(min\{z_{N-1}^*, g(x_n,t_n)\})\right) \cdot dF_{\widehat{\psi}_n|x_n}(t_n) \end{aligned}$$

The probability of the event the customer did not purchase anything (we omit feasibility constraint for compactness $\mathbb{I}[z_N^* \leq g_0 < z_{N-1}^*]$):

$$\text{(A53)}\quad \begin{aligned}\mathbb{P}[no\ purchase|\mathcal{X}] &= \iiint \prod_{k=1}^{N} \mathbb{I}[g_0 \geq g_k] \cdot \prod_{k=1}^{N} dF_{g|x_k}(t_k) \\ &= \iiint \prod_{k=1}^{N} \mathbb{I}\left[t_k < h_{x_k}(g_0)\right] \cdot \prod_{k=1}^{N} dF_{\widehat{\psi}_k|x_k}(t_k) = \prod_{k=1}^{N} F_{\widehat{\psi}_k|x_k}\left(h_{x_k}(g_0)\right)\end{aligned}$$

Notice that these equations fully determine the likelihood function. We may think of purchase/return decision as one random variable which given $\mathcal{X}$ taken $2 \cdot N + 1$ possible value (each item could be purchased and consequently be returned or kept which gives $2 \cdot N$ and additionally customer could purchase nothing).

Finally, given:

- Equation (A50)
- Equation (A29) for distribution $F_{\widehat{\psi}_k|x_k}(\cdot) = F_{\widehat{\psi}_k}(\cdot)$
- Equation (A31) for distribution $F_{\psi_n|\widehat{\psi}_n,x_n}(\cdot) = F_{\psi_n|\widehat{\psi}_n}(\cdot)$

$$\begin{aligned}&\mathbb{P}[purchase\ N\ and\ return\ it|\mathcal{X}] = \\ &= \int_{h_{x_N}(max\{z_N^*,g_0\})}^{\infty} \Phi\left(\frac{-x_N'\beta - c - \frac{\sigma_\psi^2}{\sigma_\nu^2 + \sigma_\psi^2} t_N}{\sqrt{\frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}}}\right) \\ &\cdot \prod_{k=1}^{N-1} \Phi\left(\frac{min\left\{h_{x_k}(z_{N-1}^*), \frac{\sigma_\psi^2 + \sigma_\nu^2}{\sigma_\psi^2} \cdot (x_N'\beta - x_k'\beta) + t_N\right\}}{\sigma_\psi^2 + \sigma_\nu^2}\right) \cdot d\Phi\left(\frac{t_N}{\sigma_\psi^2 + \sigma_\nu^2}\right)\end{aligned}$$

$$\text{(A54)}\quad \begin{aligned}&\mathbb{P}[purchase\ n\ and\ return\ it|\mathcal{X}] = \\ &= \int_{h_{x_n}(max\{z_N^*,g_0\})}^{h_{x_n}(z_{N-1}^*)} \Phi\left(\frac{-x_n'\beta - c - \frac{\sigma_\psi^2}{\sigma_\nu^2 + \sigma_\psi^2} t_n}{\sqrt{\frac{\sigma_\psi^2 \cdot \sigma_\nu^2}{\sigma_\psi^2 + \sigma_\nu^2}}}\right) \\ &\cdot \prod_{k=1,k\neq n}^{N-1} \Phi\left(\frac{min\left\{h_{x_k}(z_{N-1}^*), \frac{\sigma_\psi^2 + \sigma_\nu^2}{\sigma_\psi^2} \cdot (x_n'\beta - x_k'\beta) + t_n\right\}}{\sigma_\psi^2 + \sigma_\nu^2}\right) \\ &\cdot d\Phi\left(\frac{t_n}{\sigma_\psi^2 + \sigma_\nu^2}\right)\end{aligned}$$

$$\mathbb{P}[no\ purchase|\mathcal{X}] = \prod_{k=1}^{N} \Phi\left(\frac{h_{x_k}(g_0)}{\sigma_\psi^2 + \sigma_\nu^2}\right)$$

Appendix H

We estimated gradient boosted trees model from LightGBM package for Python. We optimized hyper-parameters on the validation sample by using grid search in Table A10[12]. In all the cases we used an early stopping rule – that is the method stopped adding additional regression trees if the score on validation sample did not improve for 200 consequent iterations.

Table A10. Grid for parameter search

Parameter	**Grid**
learning_rate	$[0.01, 0.005, 0.001]$
max_depth	$[5, 7, 9, -1]$
num_leaves	$[31, 63, 127]$
feature_fraction	$[0.33, 0.66, 0.99]$

As we mentioned in Section 3 the search data are highly multi-dimensional. There our goal was to generate interpretable variables where each of them comes from different domain. For prediction exercise we generated additional features and rely on the ability of the machine learning algorithm to identify the relevant information.

Table A11. Search variables used in prediction exercise

Variable	**#**	**Description**
Filter dummies	6	Indicator whether the customer used the filter on brand, size, color, price, sale, new
Sort dummies	4	Indicator whether the customer sorted products by asc/desc price, new, top sellers
Time from first to last product visit in seconds	2	-
Session length in seconds	1	-
Time spent on product	1	-
Average time spent on product during the session	2	Treating different colors as same/different product
Number unique products viewed (and within same category)	4	Treating different colors as same/different product
Number product categories/ colors viewed	2	-
Number colors of purchased products viewed	1	-
Average number colors of purchased products viewed during the session	1	-
Number products viewed before the purchased one (and within same category)	2	-

[12] We use parameter names from the LightGBM python package