**From Clicks to Returns: Website Browsing and Product Returns**

Marat Ibragimov
Assistant Professor of Marketing
Emory University Goizueta Business School
marat.ibragimov@emory.edu

Siham El Kihal
Assistant Professor of Marketing
Frankfurt School of Finance & Management
s.elkihal@fs.de

John R. Hauser
Kirin Professor of Marketing
MIT Sloan School of Management
hauser@mit.edu

This paper has benefited from feedback obtained at the 44th INFORMS Marketing Science Conference 2022, 11th Workshop on Consumer Search and Switching Costs (NYU Stern), ISPOC 2022 Series, 45th INFORMS Marketing Science Conference 2023, and the seminars at MIT Sloan, Frankfurt School of Finance & Management, Boston College, UT Dallas, University of Utah, Emory University, University of Kansas, University of Delaware, CU Boulder, Purdue University, Rutgers University. We wish to thank Giovanni Compiani, Daria Dzyabura, Tetyana Kosyakova, Ilya Morozov, Artem Timoshenko, Duncan Simester, Raluca Ursu, and Jeremy Yang for their valuable comments and suggestions.

## From Clicks to Returns: Website Browsing and Product Returns


### Abstract

Online retailers are challenged by frequent product returns, which approach a staggering annual value of nearly $1 trillion in the US alone (The New Yorker 2023). While existing research focused on managing returns using a purchase/return framework, we explore how prepurchase customer activities on retailers' websites can improve product return management. We demonstrate that such information provides important insights and can inform retailer's return management strategies. Using data from a large European apparel retailer, we propose and estimate a joint model of customer search, purchase, and returns. The model-free evidence and our empirically-based customer-journey model consistently show how specific customer browsing patterns are linked to product returns. More specifically, we find that purchasing the last clicked product, browsing fewer products, using filters, and browsing a more focused variety of products are linked to a lower return probability. Using our model, we show how strategic adjustments of product visibility on the website can improve retailers' overall performance.


*Keywords*: product returns, customer journey, retailing, customer search

Product returns pose a substantial challenge for retailers. Product returns significantly decrease the profits by reducing revenue (refunds) and increasing cost (backward logistics, dry cleaning, etc.). For example, L.L. Bean spent $50 million annually on return costs, amounting to about 30% of the retailer's annual profits (Abbey et al. 2018). Return costs are often so high that major online retailers such as Amazon and Walmart have begun to allow customers to keep the product because extracting benefit from the returned product is less than the return costs (Wall Street Journal 2022). Zara started charging online shoppers for returns unless the products are returned to the physical store (BBC 2022).

While managing product returns is critical to retailers' profitability, reducing returns is challenging because discouraging returns might harm future purchases (The New Yorker 2023). Retailers use customers' purchase and post-purchase (returns) information to manage product returns leading most researchers to study product returns in a "purchase/return" framework – a framework that takes the product purchase event as the starting point of the customer journey. Their research suggests that product characteristics jointly affect the probability of purchase and return because the option to return a product has value to the customer and impacts purchase decisions. From a managerial perspective, changes in the return policy (for example, towards a more lenient policy) impact customers' purchase behavior.

We explore whether customers' pre-purchase browsing behavior at the retailer's website is informative of return behavior and whether data on such behavior help retailers improve return management strategies. To achieve our goals, we extend the purchase/return framework to incorporate rational search. Within the framework, online retailers track the customer's journey from the moment the customer starts browsing the website (pre-purchase), through the decision on whether or not to purchase, and to the decision on whether to keep or return the product (post-purchase). In keeping with recent trends influenced by privacy concerns, we focus on first-party data (Padilla, Ascarza, and Netzer 2023). The value of first-party data is likely to increase if major browsers continue to phase out third-party cookies (Kruppa and Haggin 2024). By understanding and modeling the more complete customer journey, we gain insights that inform product return management. We expect pre-purchase search information will be informative of customer return behavior because pre-purchase information has proven valuable for insights into customers' purchase behavior. For example, Moe (2003) shows how clickstream data can help categorize shoppers, and Chen and Yao (2017) show that refinement tools significantly impact customer behavior and market structure in the pre-purchase stage.

To illustrate the potential for including customers' browsing behavior when managing returns, consider the hypothetical case of Nelly and Wendy, who both purchased the same pair of jeans. Nelly kept the jeans while Wendy returned them for a full refund. From purchase data, these customers are indistinguishable, however, pre-purchase browsing data might reveal that Nelly used refinement tools (filtering products by color), browsed many colors of the chosen option, and spent considerable time reviewing the product webpage. Wendy, on the other hand, purchased jeans from the front page without using filters, viewed only one color, and did not spend much time browsing the retailer's website. We will argue and demonstrate empirically that Nelly's and Wendy's observed search patterns reveal, in part, their likelihoods of returning the jeans – Wendy is more likely to return her purchase. Assuming Nelly and Wendy made rational decisions, their decisions provide insights about subsequent purchase behavior. For example, the use of filters reveals Nelly's more careful search, which in turn helps predict her likelihood of returning the jeans. We do not claim the use of filters is causal and we cannot infer encouraging filters would change Wendy's return behavior, but we can make managerial decisions informed by improved knowledge of customers' likelihoods of purchase and returns.

Data on the entire search/purchase/return customer journey are rare, but obtainable from at least one retailer. Using browsing sessions linked to data on purchasing and returning products at a major European apparel online retailer, we provide data-based stylized facts connecting search, purchase, and returns. The stylized facts motivate a rational model of the customer journey which explains the relation between customer browsing activities and customers' purchase and return decisions. Using the rational model, we show that modeling the complete journey is important. Without including the complete customer journey in the model, the retailer could make incorrect decisions. Lastly, we use the model to suggest low-cost low-effort changes to the retailer's website to better manage the entire customer journey.

## Related Literature

This paper contributes mainly to the literature on product returns. Research on product returns has been both theoretical and empirical. Theoretically, researchers have focused on return policies to demonstrate that the option to return products serves as a risk-reducing mechanism that encourages the customer to experience the product (Che 1996); also studied empirically by Petersen and Kumar (2015)) or as a signal of product quality (Moorthy and Srinivasan 1995).

Empirical research has focused on the optimization of return policies by firms. In an attempt to identify the optimal return policy, researchers recognize the trade-off between higher demand and higher return rates when firms use lenient policies and suggest that the optimal return policy must be balanced (Davis et al. 1998; Bower and Maxham III 2012; Abbey et al.

2018) because overly strict return policies lead to a decrease in purchases (Bechwati and Siegal 2005). Janakiraman et al. (2016) provide an extensive review of the effect of return policy leniency on purchases and returns. Anderson et al. (2009) propose a structural model where the option to return is embedded in a customer purchase decision – the customer learns private information only after purchasing the product. Other empirical studies demonstrate that a variety of policy factors affect the probability of product returns including price, discounts, marketing instruments (e.g., free shipping), or the truthfulness of product reviews (Petersen and Kumar 2009 2010; Sahoo et al. 2018; Shehu et al. 2020; El Kihal and Shehu 2022). Other empirical studies suggest prescriptive instruments to decrease return rates including visualization systems and online product forums. These instruments decrease return rates by decreasing uncertainty in the match of the product to the customer (Hong and Pavlou 2014). Other researchers use machine learning to accurately predict returns and identify product-related features that enable the firm to better select and design fashion products for the retailer's website (Cui et al. 2020; Dzyabura et al. 2023).

We contribute to the product returns literature by including consumer search activities at the retailer's website, that precede purchase, to have a more complete understanding of the customer journey. Customer search is an established and mature field of research. The literature typically follows either sequential (Weitzman 1979) or simultaneous (Stigler 1961) approaches. Both approaches assume the customer knows the distribution of the rewards and searches to resolve uncertainty. For example, Weitzman examines a stylized problem of sequentially opening boxes to learn their value and then deciding when to stop searching and collect the value of the best box (but paying the search cost for every box opened). If the value distributions are known for all boxes, Weitzman proves that the optimal (dynamic programming) search strategy is an index strategy – choose next the box with the highest index and stop searching when the value of the best box already opened exceeds the indices of all remaining boxes. Most of the literature focuses on sequential search buttressed by Bronnenberg et al. (2016) who report strong evidence to support sequential search.

Recent papers allow for flexible preference heterogeneity (Morozov et al. 2021), add learning (Ke et al. 2016; Branco et al. 2012; Dzyabura and Hauser 2019), multiple attributes (Kim et al. 2010), intermediaries (Dukes and Liu 2016), search duration (Ursu et al. 2020), and search fatigue (Ursu et al. 2023). The availability of click-stream data has enabled researchers to study empirically customer search behavior (Bronnenberg et al. 2016; Chen and Yao 2017; Ursu et al. 2020) and provide detailed insights on search-to-purchase customer behavior. For example, Bronnenberg et al. (2016) examine customer search behavior for cameras and show that early search is highly predictive of customer purchase and that the first-time discovery of the purchased alternative happens towards the end of the search. Chen and Yao (2017) show that refinement tools significantly impact customer behavior and the market structure. Ursu et al. (2020) study search duration, quantify customer preferences and search costs, and develop insights on how much information to provide on a platform.

To date, researchers have focused primarily on the purchase-to-return sub-journey (returns literature) or the search-to-purchase sub-journey (search literature). Research on the search-to-return is scarce and uses a theoretical lens (Jerath and Ren 2023; Janssen and Williams 2023). We expand these research streams to focus on the entire search-to-purchase-to-return journey in the empirical setting. Our research provides complementary insights to the returns literature (search predicts returns) and to the search literature (the possibility of returning a product changes a customer's optimal sequential search strategy). We demonstrate that by focusing on the entire customer journey, we gain additional insight into customer behavior and propose possible new strategies and tactics to improve retailer performance.

## Data and Model Free evidence

### *Data Used in the Paper*

We sought and obtained online-channel individual-level data from a large apparel retailer in Western Europe. We focus on the online channel because (1) most returns are through the online channel (the retailer has in total 53% of sold products being returned – typical for the European apparel industry) and (2) the online channel is an ideal situation in which to observe browsing, purchases, and returns for each customer. We preprocessed the data by removing noise and outliers (for example, extremely short/long sessions). In this paper, we focus on orders which had at most one product purchased. This focus isolates the impact of search on product returns by excluding situations when the customer purchases several colors or variations of a product with the intention to keep only one. We leave the extension to multiple-product orders for future research. A detailed description of data pre-processing can be found in Web Appendix A.

The retailer sells medium-priced fashion products for women, men, and children. The retailer sells mostly adult apparel, which compromises 95% of the purchases. As is typical for Europe, the retailer has a generous return policy, where products can be returned for free, for a full refund, within 60 days after the purchase with or without providing a reason.

Data include both mobile and desktop usage and consist of three main components:

- Pre-purchase (search): Website browsing records the sequence of actions made by the customer during the browsing session, at the retailer's website. We observe products listed on the website for the customer and the set of products considered (clicked to view the detailed product page), as well as the sequence of these clicks. We also observe all actions (e.g., clicking on a product, sorting by price) and the timing between the different actions, allowing us to observe how much time a customer spends on a specific product page.

- Purchase: Purchases include the product purchased (if any) by the customer during browsing sessions. These data include product characteristics such as price, category, fabric, size, brand, color, and product image.
- Post-purchase (returns): Returns contain information on whether the purchased product was kept or returned by the customer. Data also include the date and stated reason, if any, for the return.

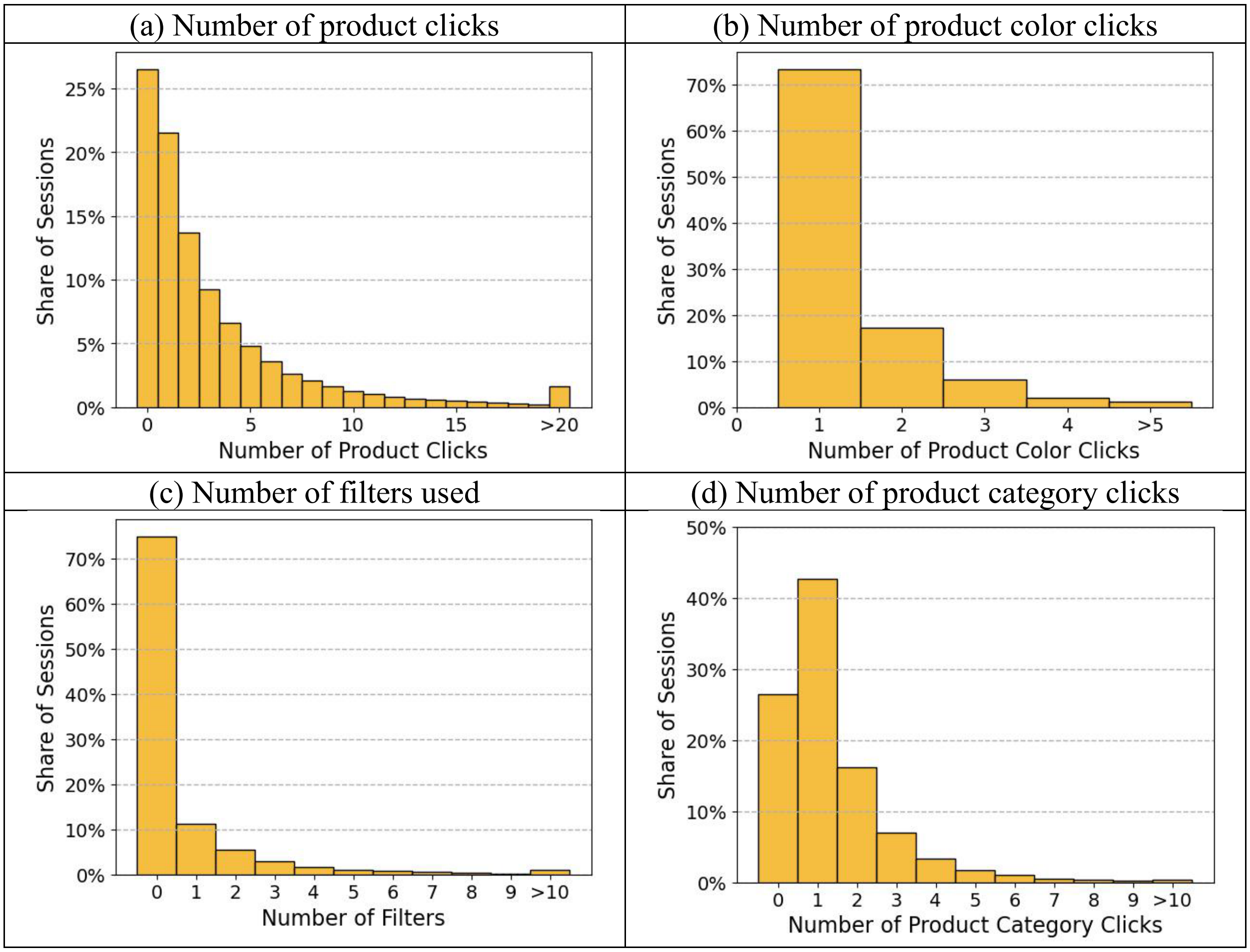
All three components of data are matched by a unique identifier. For each session, we observe the complete customer journey from opening the retailer's website to deciding whether to keep a fashion product. Future research might examine the impact of searching on other websites before visiting the retailer's website. However, we note the increasing concern for privacy in the European Union and the increased emphasis of first-party data (Padilla, Ascarza, and Netzer 2023).

These data allow us to build and estimate a model which combines customer browsing, purchase, and return behaviors. The observation period is between October 1, 2019 and February 28, 2020[1]. After the preprocessing described in Web Appendix A, we observe 919,225 browsing sessions, of which 52,053 (5.7%) result in a purchase. In 40.9% of these single-item cases, customers decide to return the product purchased. As anticipated, the return rate for the single-

[1] We have access to data until May 15, 2020, however, we excluded the months when the Covid related restrictions took place in the country where our retailer primarily operates.

item subsample is lower (40.9%) than the return rate for the multiple-item subsample (53%), but the qualitative implications remain intact. Figure 1 provides summary statistics for customer browsing at the retailer's website.

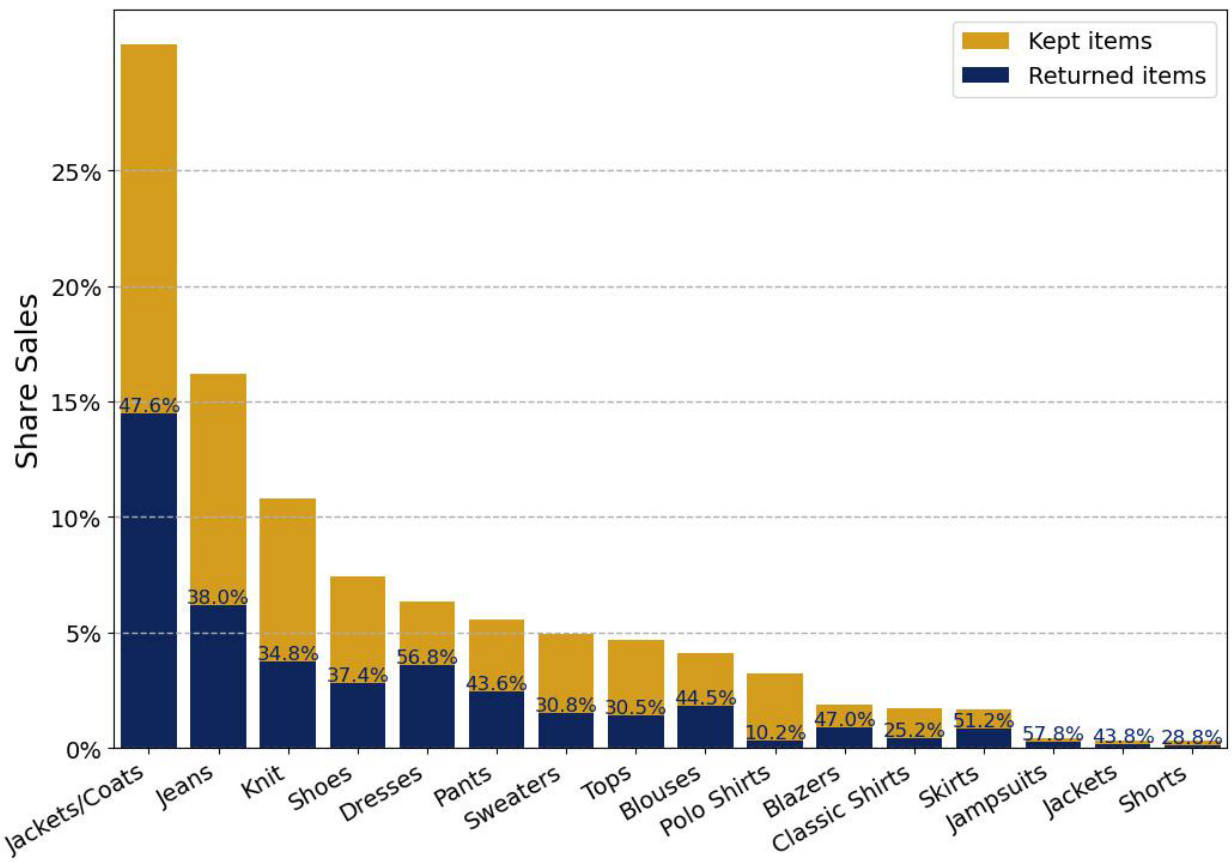
Figure 1: Descriptive statistics of customer browsing behavior at the retailer's website.




Customers can access the retailer's website through a desktop or a mobile device (54.9% access through a mobile device in the data). On the website, the customer observes a product list, which displays an image of the product, its price, category, and a small picture of the product. Customers can use search filtering tools to select a more specific product list. Customers can filter by brand, color, products on sale, new products, or product size. When the customer clicks on a specific product, further information is revealed on the product page, such as more (and higher quality) product images, and detailed product descriptions. During these 919,225 browsing sessions, the customers review on average 3.2 products (median 2). In 25.1% of sessions, the customer used at least one filtering tool (for example, by color) and in 26.6% of cases reviewed more than one color variety of the product.

Customers can choose among 16 high-level product categories, as predefined by the retailer (e.g., jeans, blouses, dresses, coats, shoes). The most popular purchased product categories are "jackets and coats" (29.8%) and "jeans" (15.9%). "Dresses" and "jumpsuits" have the highest return rate (56.8% and 57.8% respectively) and "t-shirts" have the lowest return rate (10.0%). Figure 2 displays return rates by category and the sales share of each category.

Figure 2: Sales share and return rates by product category.



***Model-Free Evidence***

Before we turn to the model of the search/purchase/return customer journey, we provide illustrative model-free evidence. The probability of returning a product is linked to different aspects of customers' browsing sessions (Figure 3). Search and browsing data are high dimensional not only because of the number of options available to the customer but because the order of customer actions matters. To gain intuition, we summarize the search with aggregate statistics of different aspects of customer pre-purchase browsing at the retailer's website and investigate how these statistics relate to product returns. We examine relationships in the data with the caveat that these relationships are not necessarily causal. Our focus is on identifying search-behavior characteristics that (probabilistically) reveal the customer situation which, in turn, is a predictor of returns behavior.

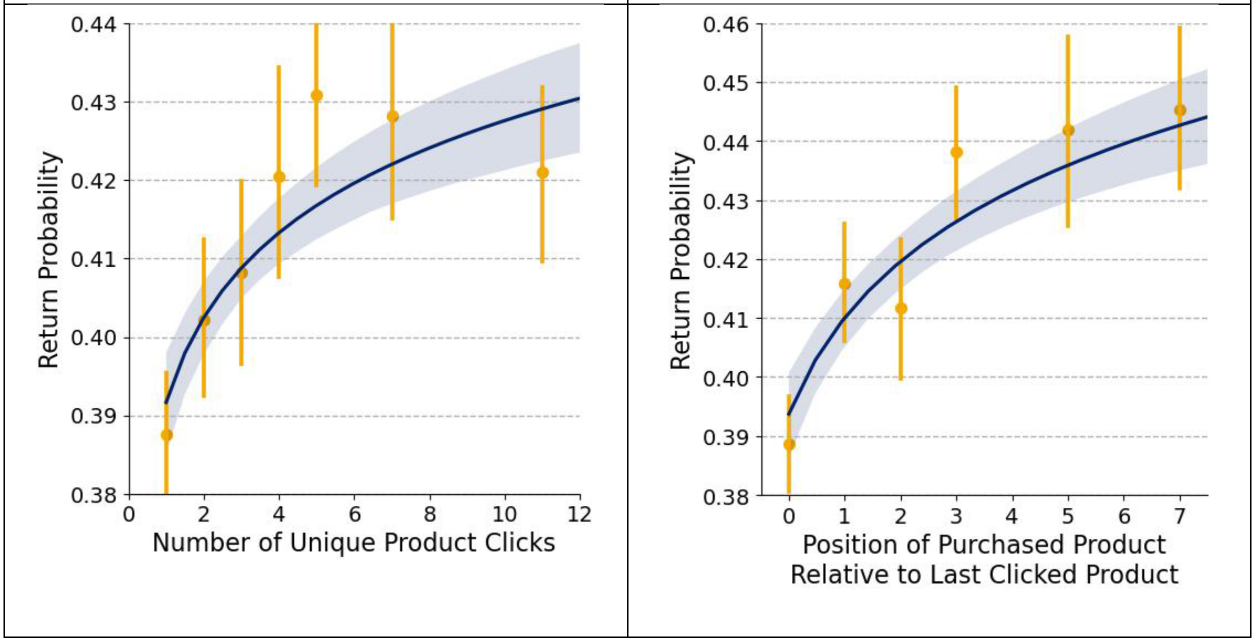
We begin by examining the number of products on which the customer clicked during the browsing session to reveal additional information. The gold lines in Figure 3 indicate the mean and 95% confidence interval, the blue line indicates a log-based regression plus 95% confidence intervals for the estimated curve. Figure 3a suggests a strong positive correlation between the number of product clicks and the probability of return. Customers who click and review more products return on average more frequently. Similarly, we found that customers who purchase products closer to the end of their click sequence are less likely to return the product. Figure 3b illustrates this dependence where the horizontal axis represents the distance (number of products clicked) between the purchased product and the last clicked product (0 implies that the customer purchased the last clicked product). It follows that customers who purchased the last clicked item are less likely to return the product.[2]

[2] For this graph, we focus only on customers who clicked on more than 1 product before purchasing at most one product.

Moreover, we see that actions that precede product views (or clicks) are also related to product returns. Specifically, customers who use tools to refine their browsing experience (for example, filtering products by price, size, color, etc.) are less likely to return the purchased product. Figure 4 relates the number of filters applied by the customer to the return probability and shows that customers who apply one additional filter are four percentage points less likely to return purchased products.

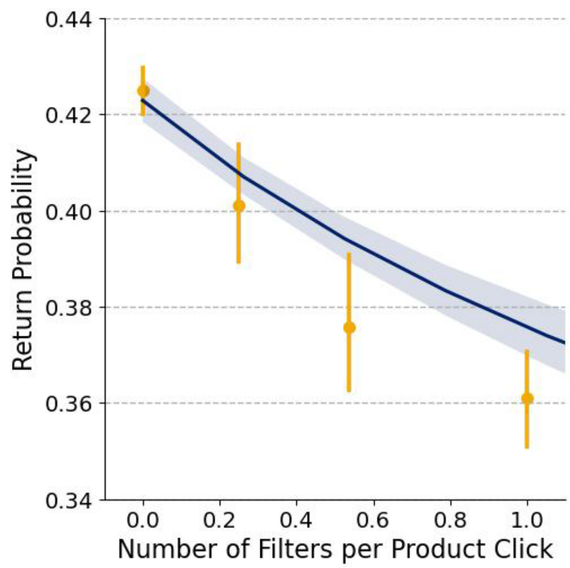
Figure 3: Model Free Evidence – Browsing products, position of purchased product relative to last clicked, and return probability.

| (a) Number of product clicks | (b) Position of purchased product relative to last clicked |
|---|---|




Figures 3 and 4 demonstrate the relation between customer browsing and the probability of returns. However, all these measures ignore an important aspect of customer pre-purchase browsing behavior – upon which products the customer clicked. For example, consider two hypothetical customers who both clicked on five products during their browsing session. The first customer clicked on five similar t-shirts, while the second customer clicked on two t-shirts, two pairs of jeans, and a dress. This variation in click behavior might foreshadow different return behaviors. The first customer seems to be more focused and specifically looking for a t-shirt, while the second customer seems to be considering a variety of wardrobe choices. There are of course other explanations, we are examining correlation not a one-to-one focus-to-behavior relationship. Because Figure 3a would treat both customers the same, we look more deeply at the data.

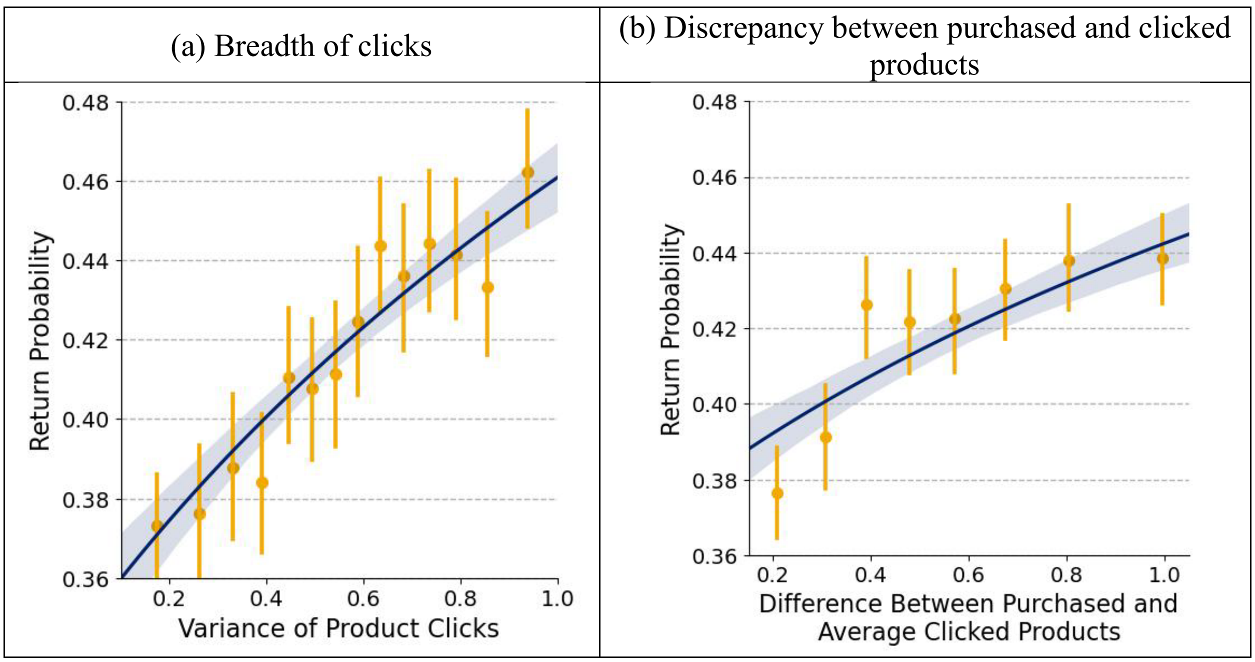
Figure 4: Model Free evidence – Filter usage and return probability.




To see how the type of searched products impacts customer returns behavior, we use deep-learning product embeddings. Intuitively, product embeddings allow us to summarize the information about the product in a seven-dimensional vector with the property that similar products would have similar product embeddings. The details on the construction of such embeddings can be found in Web Appendix B. Once we obtain product vectors, we can evaluate the clicks set of customers – whether customers click on different products (high variance of clicked product embeddings) or similar products (low variance). We plot the result in Figure 5a. Customers who click on very different products are more likely to return the product after the purchase. We support the results by examining the distance of the purchased product and the "average" clicked product in Figure 5b – customers who purchase a product which is very different from their clicked products are more likely to return it. Figure 5b suggests a potential difference between deep search (looking for a specific product) and broad search (casually browsing for a variety of products). Deep-search customers appear to be less likely to return products, likely because they are either more informed, more focused, or less impulsive.

We now seek to capture these insights with a formal model.

Figure 5: Model Free evidence – Breadth of clicks and discrepancy between purchased and clicked products, and return probability.




## Model Development

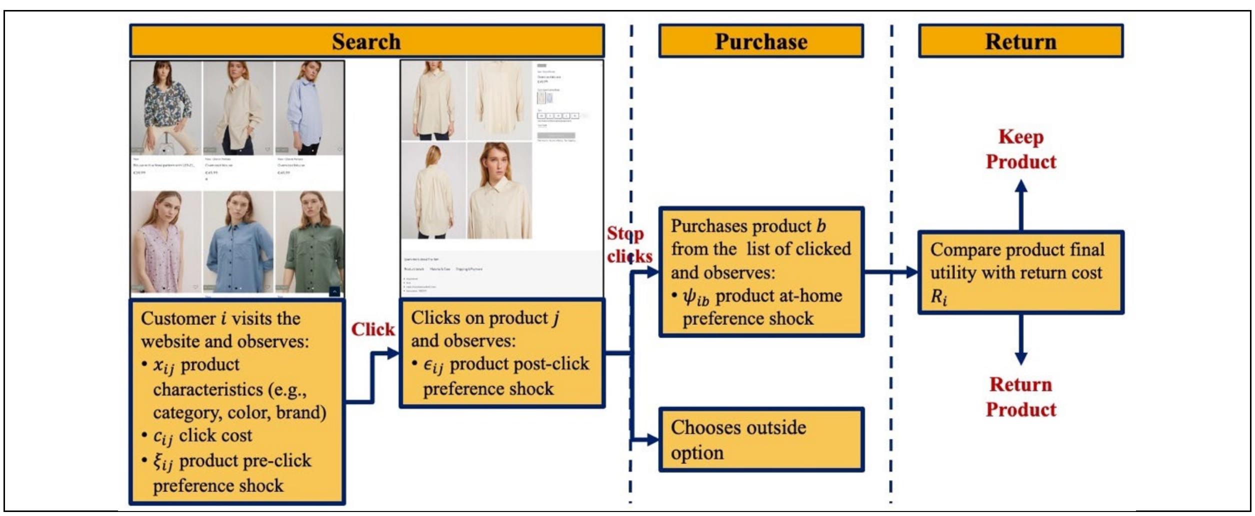
To understand how customer browsing is related to the return decision, we adopt methods from the developed field of customer search. Specifically, we model customer search as sequential and rational – customers review products one by one and make a decision to purchase the product when the expected value of purchasing exceeds the expected value of reviewing more products. In this section, we introduce the formal model. In subsequent sections, we address parameter estimation, examine the stylized facts based on the model, and suggest actions the firm can take to better manage profitability when returns are allowed.

Figure 6 provides an overview of the search/purchase/return model of the customer journey. Consider customer i who visited the retailer's website, the customer observes the list of products $V_i$. Simply by viewing the product list, the customer forms initial impressions: some product-related characteristics $x_{ij}$ (price, category, color, etc.) and an individual pre-click preference shock $\xi_{ij}$. The customer has the option to click on any of these products to reveal additional information $\epsilon_{ij}$. However, clicking to gain additional information requires the customer to incur some (possibly minor) costs $c_{ij}$. (For example, the customer may need to move the mouse, click, and process the information revealed.) After the click, the customer either continues clicking (if they see other interesting options) or stops to decide whether they like any of the products clicked so far. If the customer decides to terminate the search, the customer purchases the best product among those clicked or leaves the website without a purchase. If the customer purchases the product, the customer receives the purchased product b and can inspect it in more detail at home (e.g., try it on, hold it up, feel the material, and compare its fit to the customer's other fashion products). Inspection reveals additional information (for example, fit with the body type) denoted as the $\psi_{ib}$. Based on all information accumulated (online search and offline inspection), the customer decides whether to keep the product or to incur a return cost $R_i$ (e.g., return label, travel time, etc.) by returning the product to the retailer for a full refund.

To model the full customer journey from search to purchase to return, we adopt a model from the literature on customer search – the popular Weitzman model. The choice of the framework was governed by three factors. First, the framework models the customer as a rational agent and provides a baseline with which to understand customer behavior. Second, the framework has been widely used in empirical settings enabling us to build upon the many estimation and identification challenges that have been addressed. This prior research enables us to focus on the new challenges introduced by adding the post-purchase customer decision of whether to return the product. Third, it is natural to extend the framework to model the return decision as rational.

We modify the classical search framework. In particular, our model assumes that the customer does not infer the true utility upon click: part of the utility $\psi_{ij}$ remains unknown until the customer receives the product at home. While on the retailer's website, the customer makes the purchase decision under uncertainty about the true fit. For example, for products with $\sigma_\epsilon \gg \sigma_\psi$ the customer extracts most of the information from online search and thus makes a highly-informed purchase decision. In this case, returns would be low. On the other hand, when most of the customer's learning takes place after the customer receives the product, returns might be higher depending on the detailed parameters of the model.

Figure 6: Overview of the search/purchase/return model of the customer journey.



*Click Costs, Utility, and Returns*

For ease of notation without loss of generality, we number products in a way that $j$ indexes the sequence in which customer $i$ clicks on products (for example, $j = 2$ implies the second clicked product, while $j = 0$ implies the outside option which is always available). The customer's final utility could take one of three possible forms (click costs are paid prior to the realization of this utility and thus not included in the equation):

$$u_{ij} = \begin{cases} \mu_{ij} + \xi_{ij} + \epsilon_{ij} + \psi_{ij} & \text{purchased and kept a product } j \neq 0 \\ -R_i & \text{purchased and returned a product } j \neq 0 \\ 0 & \text{chose outside option } j = 0, \text{normalized to } 0 \end{cases} \quad (1)$$

where $\mu_{ij}$ is the customer's preference for the attributes $x_{ij}$ of product $j$. (Recall that $j$ indexes the click order, thus for every customer the $j^{th}$ product could be different requiring notation that allows the product's attributes to differ.) The customer's preference is linear in product's attributes, $\mu_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta}_i^u$ where $\boldsymbol{\beta}_i^u \sim N(\boldsymbol{\mu}_{\beta^u}, \boldsymbol{\sigma}_{\beta^u})$ is the customer's preference vector. The other terms represent information revealed to the customer: $\xi_{ij} \sim \mathcal{N}\left(0, \sigma_{\xi_{ij}}\right)$ is customer's preference "shock" for product $j$ revealed at the beginning of the browsing session (product list view), $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_{\epsilon_{ij}}\right)$ is customer's preference "shock" revealed after a click on product $j$ (product page view), and $\psi_{ij} \sim \mathcal{N}\left(0, \sigma_{\psi_{ij}}\right)$ is the customer's preference "shock" revealed after the customer receives and examines product $j$ (offline product inspection). We assume that all individual preference shocks are independent.

*Variation by product*. A simple t-shirt might be easier to evaluate than a nuanced evening dress during the purchase decision and during the at-home inspection, thus we allow the distributions of the after-click and after-purchase shocks to depend upon observable characteristics for the product webpage $\mathbf{y}_{ij}$ and observable characteristics of the product $\mathbf{z}_{ij}$. For the purpose of this paper, we follow Anderson et al. (2009) and assume that we can use the product attributes $\mathbf{x}_{ij}$ rather than introducing new $\mathbf{z}_{ij}$, but with different weights, when modeling the standard deviation of the after-purchase shock. Thus, we set $\log \sigma_{\psi_{ij}} = \mathbf{x}_{ij}'\boldsymbol{\beta}^{\psi}$.

*Expected purchase utility***.** Because the customer does not observe the at-home-inspection shock $\psi_{ij}$ until after the purchase decision, the customer must evaluate the product by taking an expectation over the unobserved variable $\omega_{ij} = \mathbb{E}_{\psi_{ij}}[u_{ij}]$. In Web Appendix C, we demonstrate that

$$\omega_{ij} = \sigma_{\psi_{ij}} T\left(\frac{R_i + \mu_{ij} + \xi_{ij} + \epsilon_{ij}}{\sigma_{\psi_{ij}}}\right) - R_i \tag{2}$$

where $T(\kappa) = \kappa\Phi(\kappa) + \varphi(\kappa)$ and $\Phi(\kappa)$ and $\varphi(\kappa)$ are cumulative and probability density functions of standard normal distribution and $\kappa$ is shorthand for the terms in the parentheses.

Equation 2 demonstrates how the return option indirectly impacts the customer search relative to the standard framework. The distribution of utility depends in part on the distribution of the inspection shock <u>and</u> the distribution of the expected reward is bounded from below by $-R_i$. To examine the face validity of Equation 2, we let $R_i \to \infty$ as would be the case if returns were not allowed. In this case, $v_{ij} \to \mathbf{x}_{ij}'\boldsymbol{\beta}_i^u + \xi_{ij} + \epsilon_{ij}$ and the model converges to the standard case. It is straightforward to show that $T(\kappa) \geq \kappa \; \forall \kappa$, which implies that, for any product attributes, the option to return improves the customer's expected search utility. Intuitively, having the option, but not the obligation, to return a product is weakly better than not having the option to return the product.

*Search and return costs*. Let $c_{ij}$ be the search costs incurred by customer i when the customer clicks on product j. Search costs can depend upon the search environment. For example, clicking on a product at the top of the website might require less effort. Because j indexes the search order, we write $\log c_{ij} = \mathbf{d}_{ij}'\boldsymbol{\beta}^c$ where $\mathbf{d}_{ij}$ represents the search environment that the customer experiences for the $j^{th}$ product. In a general model, we might write return costs as a function of the characteristics $\mathbf{f}_{ij}$ of the customer and product, $\log R_{ij} = \mathbf{f}_{ij}'\boldsymbol{\beta}^R$, however, for the purposes of this paper, we assume that return costs do not vary by product or customer and write return costs as R.

***Optimal Click Strategy when Returns are Available***

Because the option to return changes the distribution of the reward, the decision rules that are common in the search literature must be updated. Conceptually, the selection, stopping, and purchase rules retain the property of maximum expected utility, but the rules anticipate more, and are more complicated. We summarize the revised decision rules below and provide more detailed equations in the next section. We provide derivations in Web Appendix D:

- Selection rule. If the customer is going to search (click), the customer will choose to click on the product with the highest reservation utility $z_{ij}$ derived from the system in Equations 3:

$$\begin{aligned} c_{ij} &= \sigma_{\psi_{ij}} \int_{\theta}^{\infty} T\left(\frac{R_i + \mu_{ij} + \xi_{ij} + \sigma_{\epsilon_{ij}} t}{\sigma_{\psi_{ij}}}\right) - T\left(\frac{R_i + \mu_{ij} + \xi_{ij} + \sigma_{\epsilon_{ij}} \theta}{\sigma_{\psi_{ij}}}\right) d\Phi_{\epsilon_{ij}}(t) \\ z_{ij} &= \sigma_{\psi_{ij}} T\left(\frac{R_i + \mu_{ij} + \xi_{ij} + \sigma_{\epsilon_{ij}} \theta}{\sigma_{\psi_{ij}}}\right) - R_i \end{aligned} \tag{3}$$

  That the second equation is a 1-to-1 mapping $\theta \to z_{ij}$, but to find $z_{ij}$, we must first solve the first implicit equation for $\theta$. Intuitively, before the click, customers do not know the value of $\epsilon_{ij}$, and thus their reservation utilities cannot depend on it. By computing the integral in Equation 3, customers estimate the expected difference between expected purchase utilities in Equation 2 with and without a new click.

- Stopping rule. The customer continues to browse until their maximal expected utility of clicked options (Equation 2) exceeds the maximal reservation utilities of unclicked options (Equation 3). This stopping rule is conceptually similar to the standard search framework, but the manner in which it is computed is changed.

- Purchase rule. When the stopping rule is reached, the customer purchases either the highest-expected-utility product from the set of all clicked products, or the outside option.
- Return rule. If the customer purchased a product (not the outside option), the customer keeps (does not return) the product if their utility for the chosen-and-inspected product is larger than the negative of return costs, $-\mathrm{R_i}$.

The return option changes the distribution of rewards and, hence, the reservation utilities for the customer. These changes have the potential to change the order in which the customer clicks on products and the potential to change the stopping rule and the return rule. In other words, when returns are allowed, the customer may search and purchase differently.

## Estimation of Model Parameters

We now address model estimation and demonstrate with synthetic data that unknown parameters can be recovered successfully from the highly nonlinear model.

***Parameters of the model***

Table 1 summarizes the functional form assumptions that are implicit in the search/purchase/return model of the customer journey. Table 1 also summarizes the parameters which need to be estimated.

Table 1: Overview of model parameters.

| | Distributional assumptions | Functional forms | Estimated parameters |
|---|---|---|---|
| Customer preference vector | $\boldsymbol{\beta}_i^u \sim \mathcal{N}(\boldsymbol{\beta}^u, \boldsymbol{\sigma}^\beta)$ | $\mu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_i^u$ | $\boldsymbol{\beta}^u, \boldsymbol{\sigma}^\beta$ |
| Pre-click preference shock | $\xi_{ij} \sim \mathcal{N}\left(0, \sigma_{\xi_{ij}}\right)$ | $\sigma_{\xi_{ij}} = 1$ | |
| Post-click preference shock | $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_{\epsilon_{ij}}\right)$ | $\sigma_{\epsilon_{ij}} = 1$ | |
| Post-purchase preference shock | $\psi_{ij} \sim \mathcal{N}\left(0, \sigma_{\psi_{ij}}\right)$ | $\log \sigma_{\psi_{ij}} = \mathbf{x}'_{ij}\boldsymbol{\beta}^\psi$ | $\boldsymbol{\beta}^\psi$ |
| Search costs | $c_{ij}$ | $\log c_{ij} = \mathbf{d}'_{ij}\boldsymbol{\beta}^c$ | $\boldsymbol{\beta}^c$ |
| Return costs | $R_{ij}$ | $R_{ij}$ | R |

Heterogeneity in the customer pre-click preference "shock" is essential for estimation, otherwise, there would be too little variation among the value of alternative clicks. Because the focus of the paper is on product returns, we set $\sigma_{\epsilon_{ij}} = \sigma_{\xi_{ij}} = 1$ for identification. We follow Ursu (2018) and assume search costs depend on the intercept and position of the product on the webpage.

The number of product features in fashion retail is huge. It is infeasible to parametrize the model using, without dimensionality reduction, all of the classical characteristics used in the search literature (category, price, rating). For example, in our data, there are 16 high-level product categories, which include broad categories like dresses and blouses. Even within dresses, there are products that have longer/shorter sleeves, different colors and patterns, etc. To make the model feasible while capturing the essential information about observable product attributes, we use seven-dimensional product embeddings as estimated with a deep-learning model to compress categorical information and product images into seven-dimensional vectors. The product embeddings capture the essence of the classical characteristics but in a lower-dimensional space. See Web Appendix B.

*Likelihood*

Let $V_i$ denote the number of products presented to the customer (for example, the number of products the customer sees on the main page of the website). From this set of products, the customer clicks on $C_i$ products according to the optimal search rules discussed in the previous section. Recall that the index j represents the order in which the customer clicks on the product (e.g., $j = 2$ denotes the second clicked product, and $j = C_i$ denotes the last clicked product). This notation implies that the customer did not click on products with $j > C_i$ enabling us to enumerate the order of non-clicked products randomly.

Consider the customer who was presented with $V_i$ products; clicked on $C_i$ products; purchased a product with index b and decided to return it. This sequence implies the following constraints where $\mathbb{I}[\text{constraint}]$ is the indicator function that takes on a value of 1 if the constraint is satisfied:

Click continuation. After clicking on option j, the customer continues clicking on other options if the value of exploring is more than the value of the best option in hand.

$$\forall j < C_i \quad \mathbb{I}\left[z_{ij+1} \geq \max_{s=j+2..V_i} z_{is}\right] \mathbb{I}\left[\max_{s=0..j} \omega_{is} < \max_{s=j+1..V_i} z_{is}\right] = 1 \tag{4}$$

Click stopping. The customer stops clicking when the maximal expected utility of clicked options is higher than the value of exploring the remaining options.

$$\mathbb{I}\left[\max_{s=0..C_i} \omega_{is} \geq \max_{s=C_i+1..V_i} z_{is}\right] = 1 \tag{5}$$

Purchase. Given the customer has clicked $C_i$ products and decides to stop clicking, the customer purchases a product if the expected utility of the purchased product is greater than the expected utility of all other clicked products including the outside option.

$$\mathbb{I}\left[\omega_{ib} \geq \max_{s=0..C_i} \omega_{is}\right] = 1 \tag{6}$$

Return. Given the customer bought the product b, the customer returns the product if the product utility is lower than the negative return cost $R_i$.

$$\mathbb{I}[\mu_{ib} + \xi_{ib} + \epsilon_{ib} + \psi_{ib} \leq -R_i] = 1 \tag{7}$$

Equations (4-7) define the set of constraints that must be satisfied to observe the given browsing session. Multiplication of the indicator functions for these conditions is the same as requiring all conditions to hold and produces a binary variable $W_i$ which takes 1 if and only if all constraints are satisfied. The case when the customer decides to keep the product or chooses the outside options closely follows the derivations in Equations (4-7). In Web Appendix F, we demonstrate that the set of Equations (4-7) could be rewritten in a more compact form in Equation (8):

$$\begin{gathered} W_i = \left[\prod_{j=1}^{C_i-1} \mathbb{I}\left[z_{ij} \geq z_{ij+1}\right]\right] \mathbb{I}\left[z_{iC_i} \geq \max_{s=C_i+1..V_i} z_{is}\right] \left[\prod_{j=0}^{C_i-1} \mathbb{I}\left[\omega_{ij} \leq \min\left\{z_{iC_i}, \omega_{ib}\right\}\right]\right] \\ \mathbb{I}\left[\omega_{iC_i} \leq \omega_{ib}\right] \mathbb{I}\left[\omega_{ib} \geq \max_{s=C_i+1..V_i} z_{is}\right] \mathbb{I}[\mu_{ib} + \xi_{ib} + \epsilon_{ib} + \psi_{ib} \leq -R_i] \end{gathered} \tag{8}$$

Because the researcher does not observe individual shocks $\xi_{ij}, \epsilon_{ij}, \psi_{ij}$ we obtain the probability of observing the given click sequence of customer i by integrating out these variables to determine the probability that all constraints are satisfied. This integration produces the log-likelihood function:

$$LL(\boldsymbol{\beta}) = \sum_{i}^{N} \log \iiint W(\boldsymbol{\xi}_i, \boldsymbol{\epsilon}_i, \boldsymbol{\psi}_i)\, dF(\boldsymbol{\xi}_i, \boldsymbol{\epsilon}_i, \boldsymbol{\psi}_i) \tag{9}$$

where $F(\boldsymbol{\xi}_i, \boldsymbol{\epsilon}_i, \boldsymbol{\psi}_i)$ represents the joint distribution of unobservable shocks.

***Estimation Procedures***

If computations were feasible, we could find the estimates of the parameters in Table 1 by maximizing the log-likelihood function in Equation 9. Unfortunately, direct maximization without either simplification or approximation is not feasible with today's computers. First, the reservation utilities $z_{ij}$ from Equation 3 cannot be computed directly because they are defined through implicit functions (we use the approximation described in Web Appendix E). Second, there is no known closed-form solution to the integral in Equation 9. Fortunately, we can draw on prior research on estimation of the search/purchase model which faced the same challenges. We considered the following methods.

*Accept-reject simulator* (Chen and Yao 2017). An accept-reject simulator replaces the true probability $P_i$ with a simulated probability $\widehat{P_i}$. In this approach, for given parameter estimates, we simulate B random draws of shocks from corresponding distributions and calculate the share of draws in which $W_i = 1$ (all constraints in Equation 8 are satisfied). The parameter vector corresponding to the largest share of draws is the maximum likelihood estimate. The challenge with this approach is that the nature of our browsing data makes $P_i$ close to zero and requires large values of B with a correspondingly substantial increase in computation time. Compounding the computational limit is the fact that this approach produces a non-smooth objective function that requires the use of non-gradient optimization methods (e.g., Nelder-Mead method). Such methods are substantially slower.

*Accept-reject simulator with smoothing* (Honka and Chintahunta 2017; Ursu 2018). Smoothing replaces the sharp constraints in the accept-reject simulation, such as $\mathbb{I}[a < b]$, with a continuous function of difference $b - a$. This approach punishes large violations of the constraints but allows small differences. While this approach is often feasible, it is not feasible for the search/purchase/return journey in our data. First, most of the constraints of the form $\mathbb{I}[a < b]$ have arguments a and b bounded from below. For example, $\omega_{ij}$ is bounded by $-R_i$ because $T(x) \to 0$ if $x \to +\infty$. When we attempt to impose these bounds, the difference $b - a$ does not translate well into a probability. Second, returns are represented by a single constraint, and we observe returns only for sessions that ended with a purchase. The "return constraint" constitutes a small proportion of all the constraints in the model. With smoothing, violation of the "return constraint" would have negligible impact on the final objective function effectively reducing the model to search/purchase rather than search/purchase/return.

*Partially closed-form integration.* The third approach recognizes that some, but not all variables, in the constraints can be integrated out with closed-form solutions. For example, only the return constraint contains the value of the shock $\psi_{ib}$. In Web Appendix G, we show that for the integral in Equation 9, all but one constraint can be integrated out. Only one constraint needs to be replaced with a smoothed version. Partially closed-form integration reduces the required number of draws B substantially and allows maximization with a gradient-based algorithm. In Web Appendix H, we demonstrate that this approach is more suitable than the other discussed above.

***Demonstration that Estimation is Feasible and Reproduces "Known" Parameters***

The search/purchase/return model in Equations 8 and 9 is complicated and nonlinear, and the estimation is not closed form. We would like to know the proposed estimation methods can recover "known" (and reasonable) parameters from synthetic data that are comparable to the data in our empirical setting. Real customers search, purchase, and return products. We would also like to examine if modeling the entire customer journey yields parameters closer to their true values than modeling just the search/purchase or the purchase/return sub-journeys (assuming of course that the model is specified correctly).

To examine parameter recovery and the implications of the full customer journey, we simulate 1,000 synthetic customers according to the search/purchase/returns customer-journey model. To create the synthetic data, we choose the parameters that represent the structure of the empirical data. Specifically, we assume that the retailer has an assortment of a total 100 different products split between two categories ($x_{ij} \in \{0,1\}$). Customers visiting the website see only a subsample of these products. The subsample consists of 48 randomly selected products where 80% of the listed products are from the second category ($x_{ij} = 1$).

We estimate three models:

1. The full search/purchase/returns customer journey.
2. The purchase/returns sub-journey in which the estimation does not use search data and the model does not include search parameters. Customers choose from the full set of products, not just those that they clicked on.
3. The search/purchase sub-journey in which the estimation does not use returns data and the model does not include returns parameters (e.g., $R \rightarrow \infty$).

We summarize the results of the estimation of all models in Table 2.

Table 2 demonstrates that the parameters of the full model can be recovered using the proposed estimation procedure. (The two alternative approaches discussed in the model estimation section performed substantially worse and required substantially more advanced computational resources). The full model recovers the true parameters with reasonable accuracy even with the relatively small size of the data sample.

Table 2 suggests that not modeling the full customer journey (when the full journey is the correct model) skews the parameters that are estimated and, of course, provides no data on the parameters that cannot be estimated. The detailed differences in parameter estimates among models and the direction of movement depend upon the parameters of the true model. However, the differences in Table 2 have face validity. The important insight is that not-modeling the full customer journey, leads to incorrect sub-journey parameter estimates.

Table 2: Estimation with Synthetic Data to Test if True Parameters can be Recovered.

| | | True value | Search/ Purchase/ Returns Model | Excluding Information on Search | Excluding Information on Returns |
|---|---|---|---|---|---|
| Value of outside option | $\beta_0^u$ | – 4.40 | – 4.38 | – 4.77 | – 3.68 |
| | | | (.10) | (5.04) | (.05) |
| Customer preference | $\beta_1^u$ | – .30 | – .31 | .08 | .39 |
| | | | (.16) | (1.12) | (.04) |
| Preference heterogeneity | $\sigma_1^u$ | .50 | .43 | .00 | .45 |
| | | | (.03) | (8.36) | (.03) |
| Search costs (intercept) | $\beta_0^c$ | – 7.00 | – 7.00 | | – 6.18 |
| | | | (.19) | | (.18) |
| Search costs (per click) | $\beta_1^c$ | .50 | .52 | | .51 |
| | | | (.03) | | (.03) |
| Post-purchase information | $\beta_0^\psi$ | 1.00 | 1.09 | 1.00 | |
| | | | (.28) | (.49) | |
| Return costs | $\log R$ | – 1.00 | – .93 | – .96 | |
| | | | (.24) | (1.37) | |

*Note*: *Bootstrapped SE in parenthesis (25 repetitions).*

Both alternative models (excluding information on search or on returns) lead to incorrect signs of the customer preference $\beta_1^u$. This implies that managers could potentially come to the wrong conclusion about which product customers prefer. The reasons for an incorrect sign of the parameters are different but both highlight the importance of modeling the complete journey.

The "no returns" model considers the popularity of the product but ignores the reason why the product is popular. Specifically, because $\beta_0^\psi > 0$, the expected purchase utility of this product is higher and thus customers purchase it more frequently. Hence, the estimated parameter value for $\beta_1^u$ is well above zero. The "no search" model assumes that the customer has enough energy to click on all products in the assortment while in reality browsing through a small subsample of products. Thus, this model would treat the high sales of products with $x_{ij} = 1$ as an indicator of "likable" product characteristics and thus assign a positive coefficient.

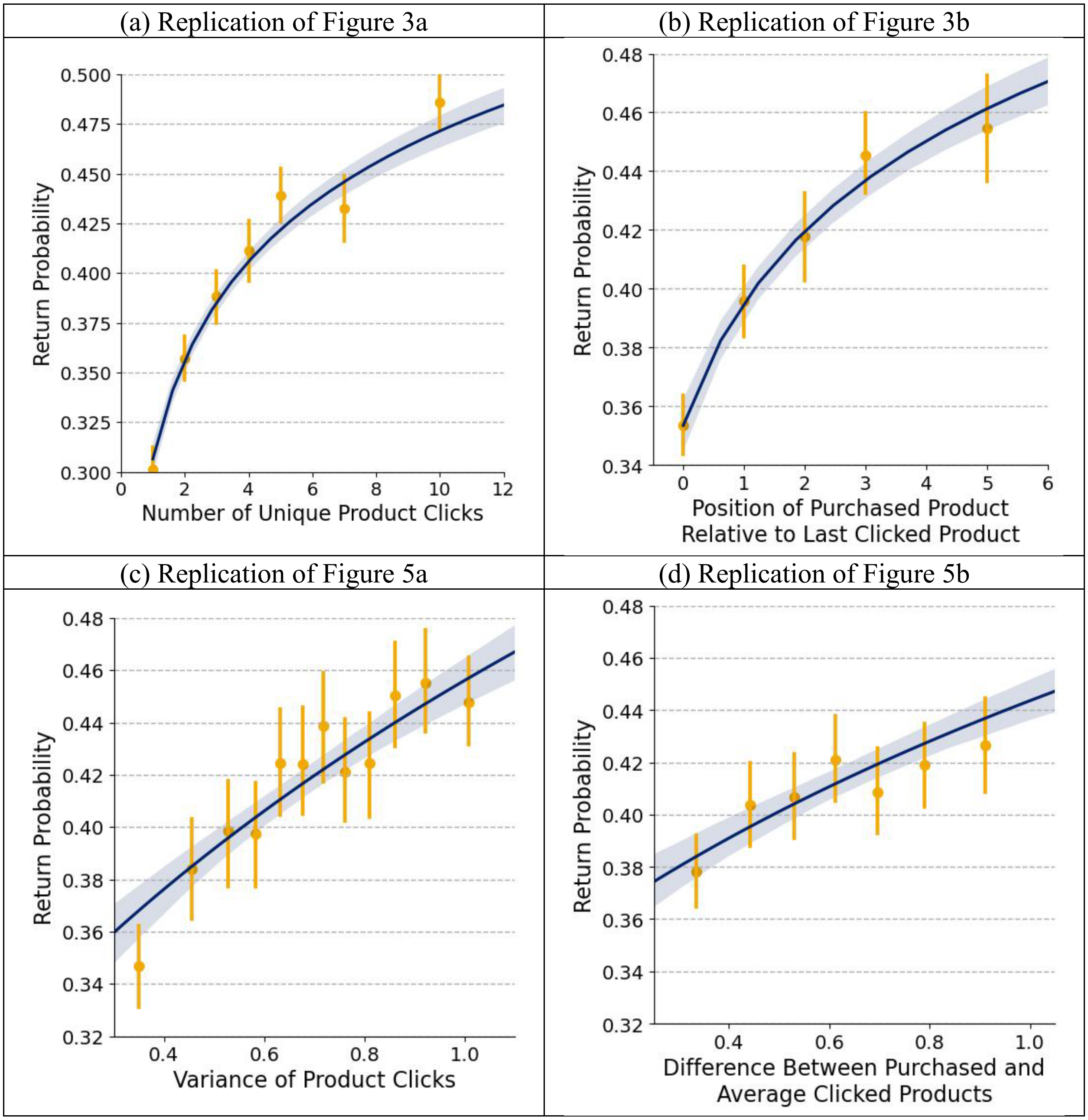
However, in reality, the high sales of the products with $x_{ij} = 1$ are due to their more frequent presence on the website (remember that 80% of listed products have $x_{ij} = 1$)

## Empirical Results for a Fashion Retailer

Having shown that we can recover known parameters with the method of partially-closed-form integration of the likelihood function, we estimate the proposed model of the search/purchase/returns customer journey. We provide details in Web Appendix H. Because the product attributes are summarized by deep-learning embeddings, the weights, and standard deviations are not interpretable — their primary purpose is to capture product information in the model so that we might focus on the customer journey. Website characteristics are interpretable: product position on the website affects the search costs and products located on the second row are less likely to be clicked by the customers.

We now use the empirically-estimated model of the customer journey to reexamine the model-free evidence discussed earlier in the paper. We then investigate how the estimated model could help the retailer to identify problematic products in the assortment and thereby improve the product display on the website. To replicate the model-free evidence, we use our estimated model to simulate the same customers as in the original data and plot the results in Figure 7. Overall, we observe that our model can replicate the qualitative implications of the figures with model-free evidence (Figures 3 & 5). Because model predictions are less susceptible to noise and generally serve as a regularized benchmark, we observe much smoother plots. In subsequent sections, we discuss in more detail each of the insights.

Figure 7: Replication of model-free evidence.




***Customers who make more clicks prior to purchase are more likely to return the product.***

Intuitively, because customers are searching optimally based on expected pre-click utility, the customers who click on many products are having difficulty in finding a product that justifies the risk of purchase and return. The purchase utility is more likely to be lower, perhaps just above the outside option. But returning the product after purchase may still be rational, thus a lower expected utility makes returns more likely. The length of the search session informs retailers how fast the customer found the product they like and longer sessions indicate that they were struggling with the decision.

More formally, consider two customers who face the same pre-click reservation utilities, $z_{ij}$'s, but whose realizations of the post-click utilities, $\omega_{ij}$, differ. Assume both customers have the same click sequences up to and including j. For the first customer $C_i = j$ and for the second customer $C_i = j + 1$. Equation 4 implies that the maximum of the post-click utilities is (weakly) less for the second customer than the first customer, because the bound in the second term is the maximum over fewer reservation utilities. The expected purchase utility $\omega_{ib}$ of the second customer would be lower and, all else equal, more likely to be below the negative of the return cost. Our model-free evidence illustrated in Figure 3a shows this and empirically-based customer-journey simulations suggest that customers with one additional click have 1.1 percentage point higher chance of returning the product.

***Customers who click on the last-clicked product are less likely to return it.***

From Figure 7b, it follows that products purchased as last click are less likely to be returned. In the customer-journey model, when a customer purchases the last clicked product, it implies that there is no value to additional search. The customer found a product that matches the customer's preferences (for example, a "dream t-shirt") sufficiently so that any expected improvement does not justify additional search. Intuitively, a last-click purchase is likely correlated with a better preference-match implying a lower return probability.

In the customer-journey model, b is an index of purchased products and $C_i$ is the index of the last product clicked. Consider two customers with two identical browsing sequences except that the first customer purchased product $b \neq C_i$ while the second one has $b = C_i$. In this case most of the constraints in Equation 8 would have a similar form. However, the constraint, $\omega_{ib} \leq z_{iC_i}$ applies only to the case when $b < C_i$. That is, the expected utility of a non-last-click purchase is bounded. On the other hand, the expected purchase utility for a last-click purchase is unbounded from above. Higher utility at purchase implies lower return probabilities (Equations 1 and 3), thus information about whether the last clicked product was purchased allows the firm to identify customers who found a well-matched product and are less likely to return the product. On average, customers who purchased the last-clicked product are 19% less likely to return products than customers who purchased a non-last-clicked product (among those who made at least two product clicks).

***Customers who apply search refining tools have a lower probability of return.***

Consider two customers searching for a dress. The first customer does not have well-defined preferences while the second customer strongly prefers a black dress of medium length and made from a natural fabric. The second customer applies a search filter to narrow the search while the first customer does not. Assume we observe that both customers buy the same dress. We know that the second customer's dress matches the first customer's preferences on at least three attributes. We do not know if the first customer found a dress to match the customer's preferences on these attributes, we only know that the first customer found a dress that represented tradeoffs on all of the available dress attributes.

More formally, by using search filters, the customer changes the distribution from which to sample the products (for example, browse only products made of natural fabrics). Typically, the application of search filters requires paying an additional search cost such as navigating through the menu, reading, and clicking. This implies that the customer faces a tradeoff: sample from a better distribution by paying additional search costs or sample from the default distribution for free. The second customer chose to pay the additional search costs, hence the second customer's pre-click distribution is a better match to that customer. Because the post-click distribution is correlated with the pre-click distribution, the expected value of the chosen product is likely higher for the second customer. As always, in the rational customer-journey model, higher expected utility implies lower return rates. To simulate the self-selection of filters, we simulate customers who have an oracle as to which filter to use. That is, customers have an option to use a filter on the first deep-learning embedding. Customers who choose to apply the filter would see products only with the first embedding greater than the $10^{th}$ percentile (10% of products are hidden). We set the filter application cost such that the share of customers choosing the filter is equal to 26% (matching the data). With this oracle-based simulation of the customer journey, customers who use a filter are 11% less likely to return a product than customers who do not use a filter. Although there could be other empirical explanations for customers' use of filters, our model suggests that the self-selection of customers is one of the possible mechanisms behind the filters and returns. Estimation of filter-applying costs goes beyond the scope of this paper, however, further research would extend the model to the pre-search stage and estimate the filter-applying costs directly. Figure 4 also shows that using filters is associated with a lower return probability.

***Customers who click on very different products are more likely to return the product.***

Intuitively, a customer who browses many products does not have a particular product in mind to purchase, therefore the customer chooses to browse the website hoping to find something interesting. Figure 7c provides evidence.

Mathematically, customers with a particular product in mind would have a high preference for some attributes of the product and lower preferences for other attributes (for example, black short-sleeved t-shirts). “Uncertain" customers have more uniform preferences over the attributes. The customer with focused preferences is more likely to click on products that look similar to their “black short-sleeved t-shirts,” while the customer with diffuse preferences would click on very different products. Intuitively, the customer search helps us to infer whether preferences are focused or diffused, which in turn reveals the likelihood of a high-match high-utility choice, which, as always, affects the probability of return.

**Empirically-based Policy Simulations: Changing Product Visibility on the Website**

As reviewed in the paper, there is a rich literature on returns policies such as whether to allow returns, whether to charge for returns, whether to allow free shipping, and whether to have a strict or lenient policy. The search/purchase/return customer journey model is complementary to these policy decisions. For example, the qualitative insights in the previous section are consistent with a strategy where a firm monitors a customer’s search and provides incentives to direct the search or reduce search costs to encourage the customer to search more products and achieve a great post-click utility. Our formal model is consistent with these strategies, but we would need to expand the model slightly to demonstrate causality.

However, our model allows the retailer to use search data to identify situations where the customer is sufficiently likely to return a product, that offering the product on the website is unprofitable. To illustrate how the search/purchase/return customer journey model might be used, we explore a scenario in which the retailer adds or removes products that are unprofitable from its website. This analysis extends the analyses reported by Dzyabura et al. (2023) who used product images to identify products whose return rates were so high that offering the products for sale was unprofitable. Dzyabura et al. (2023) project a net increase in profit of 8.3% for their data. Their results were based on the purchase/returns sub-journey.

Fashion retailers typically have a massive assortment of products (for example in our data we observe approximately 10,000 different products). Fully optimizing the assortment is a difficult combinatoric problem. There exist $2^{10,000} \simeq 10^{3,000}$ possible assortment combinations. Our goal in this paper is more modest. We seek to demonstrate that the empirically-based search/purchase/return customer journey model can identify products that are currently unprofitable and could be removed from the website. Following Dzyabura et al. (2023), this solution to the problem is feasible because there are relatively few interactions among products in the online retail environment. As long as the total percentage of products removed from the website is moderately small, removal should have little effect on the overall website traffic. Removal might even make the website more attractive because customers are more likely to find a match to their preferences.

We assume that the online retailer can update the website to hide products from view and do so at a low cost within established regulations. Furthermore, the retailer can test these website modifications with standard A/B tests. If purchase cycles were sufficiently long, the retailer might monitor search, purchases, and returns and remove products that are unprofitable, perhaps using A/B tests to confirm potential removal. However, in an industry where fashion cycles are short, the retailer might need to make decisions before sufficiently many products are returned. In our scenario, the retailer observes the customer journeys, estimates the search/purchase/returns model, and simulates various removal scenarios. Such an application is meant to be a proof-of-concept and assumes that fewer observations are needed to estimate the model than would be to obtain product-by-product return-rate estimates for all 10,000 products in the overall assortment.

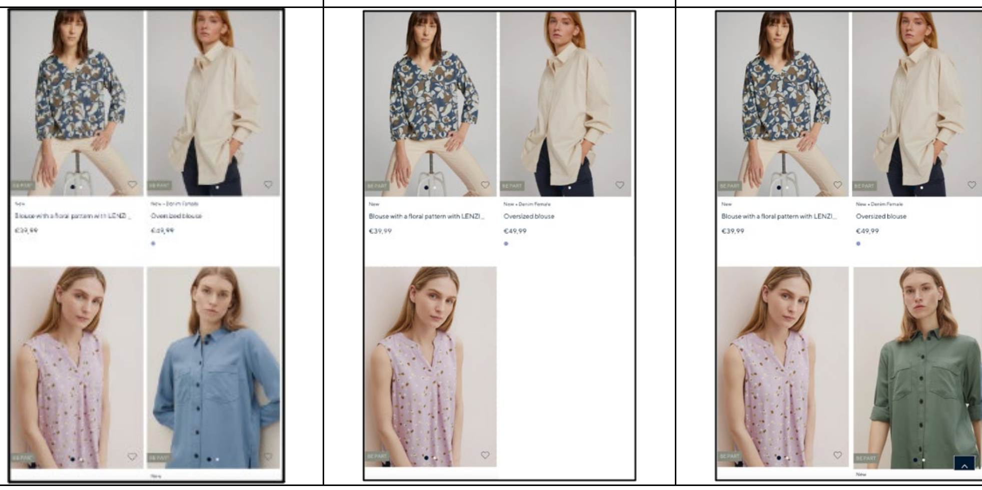
***Identifying Problematic Products in the Assortment***

We implement changes in the (simulated) assortment by removing products from the retailer's website. We assume customers behave according to the empirically-based customer-journey model. We randomly sample the unobserved (to the researcher) shocks from the corresponding distributions as the synthetic customers browse the website. We summarize our approach in six steps:

1. Simulate customers visiting the default version of the website (Figure 8a). Using the customer journey model, we compute baseline values for purchase and return rates.
2. Split all products into 256 equally-sized groups based on predicted return rate. Denote all products in group g as $S_g$.
3. Remove a product in set $S_g$ from the website to simulate hiding the product from view (Figure 8b). To avoid a number-of-products available confound, we replace the removed product with a random product from the same category as the removed one (Figure 8c).
4. Simulate customers' journeys on the updated website. Compute revised values for purchase and return rates.
5. Repeat Steps 2, 3, and 4 for each group g.
6. Compare baseline values (Step 1) with values when a product is removed (Step 4).

Figure 8: Changing the online assortment.

| (a) Baseline | (b) Product removed from baseline | (c) Random product substituted back into baseline |
|---|---|---|



In each simulation, the customer sees the same number of products on the website and each simulation removes a small portion of the assortment (approximately 0.4% given 256 groups of products). This ensures the change to the website is minor and is unlikely to cause disturbances in the customer journey. Our simulation strategy is supported by the literature where Boatright and Nunes (2001) found that reducing the assortment does not cause a reduction in the perceived variety (see also Broniarczyk et al. 1998).

In an environment where customers make rational decisions about choice, removing some products might increase overall sales from the website because removing those products makes it more likely that customers, by rational search, will find products that match their preferences better. Our results show that the suggested website changes reveal four segments of products, depending on how sales and returns changed after their removal. Table 3 summarizes the distribution of products over four segments:

- Segment 1 (12.9% of products in the assortment): Removing these products from the website leads to an overall higher purchase probability and lower return rate in the online shop. The retailer might consider removing these products.
- Segment 2 (37.2% of products in the assortment): Removing these products from the website leads to an overall lower purchase rate and higher return rate in the online shop. These are products that should not be removed.
- Segments 3 & 4 (49.9% of products in the assortment): Removing these products from the website leads to either a lower purchase rate but lower return rate or a higher purchase rate but lower return rate. Whether or not the retailer removes these products depends upon the retailer's profit margins and return costs.

Table 3: The relative size of product segments depending on their impact on purchase and return rates after removal from the online website.

| | **Lower Return Rate** | **Higher Return Rate** |
|---|---|---|
| **Higher Purchase Rate** | 12.9% | 26.9% |
| **Lower Purchase Rate** | 23.0% | 37.2% |

The number of products in each segment is not equal suggesting that the search/purchase/return model is picking up more than random noise. The search/purchase/return customer journey policy simulations estimate the total effect of removing a product, specifically, removing Segment-1 products makes it more likely for customers to find other products that match their preferences better. Intuitively, Segment-1 products likely have low sales but still take up space on the website (formally, Segment-1 products increase the search costs for other products). Moreover, in the event of a sale, sufficiently many customers dislike Segment-1 products at home and return them to the retailer for a full refund.

Table 3 highlights that removing at least some problematic products would have a positive impact on two retailer's key metrics: purchase and return rates. With more detailed data on margins and return costs, removing other products might be profitable. While the search/purchase/return customer journey model is promising, we recommend future research, and forward validation where some products are removed from an "A" website and not from a "B" website. Such A/B tests would test, and hopefully validate, the model as a practical tool to cull problematic products.

## Conclusion and Future Research

Online retailers, particularly fashion retailers, face high return rates and high return costs. Improving how a retailer manages product returns has a direct and considerable impact on the firm's bottom line. This paper explores the search/purchase/return customer journey to generate insights and suggests strategies by which a retailer can maximize profits. By modeling the full customer journey, we gain insights that enable the retailer to use search patterns to better predict returns. The search patterns are not causative, that is changing the patterns would not necessarily change return rates, but the search patterns do reveal when products are more likely to be returned.

We develop a rational model of the customer journey in the presence of a return option. Using data from a major European apparel online retailer, we provide a feasible means to estimate the parameters of the customer journey model. The empirically-based customer journey model provides explanations of observed model-free evidence on how search, purchase, and returns. More specifically, the model-free evidence and the empirically-based customer journey model consistently show that purchasing the last clicked product, browsing fewer products, using refinement tools, and browsing a more-focused variety of products are linked to a lower return probability. Empirically-based policy simulations suggest that some products can safely be hidden on the retailer's website to improve overall performance.

Future research. Data on the complete customer journey are rare. This paper illustrates what can be done with a formal model combined with such data. Hopefully, other researchers will obtain data on the search/purchase/returns customer journey in other product categories and explore implications further. Future research might model the retailers' decisions as endogenous and enable estimation in such data regimes. Our policy simulations are based on the estimated model. Future research might test the predictions.

## References


Abbey, James, Michael Ketzenberg, and Richard Metters (2018), "A More Profitable Approach to Product Returns," *MIT Sloan Management Review*, 60 (1), 1-6.

Anderson, Eric T., Karsten Hansen, and Duncan Simester (2009), "The Option Value of Returns: Theory and Empirical Evidence," *Marketing Science*, 28 (3), 405-423.

Bechwati, Nada N. and Wendy S. Siegal (2005), "The Impact of the Prechoice Process on Product Returns," *Journal of Marketing Research*, 42 (3), 358-367.

Bower, Amanda B. and James G. Maxham (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, Fernando, Monic Sun, and J. Miguel Villas-Boas (2012), "Optimal Search for Product Information," *Management Science*, 58 (11), 2037-2056.

Bronnenberg, Bart J., Jun B. Kim, and Carl F. Mela (2016), "Zooming In on Choice: How Do Consumers Search for Cameras Online?" *Marketing Science*, 35 (5), 693-712.

Che, Yeon-Koo (1996), "Customer Return Policies for Experience Goods," *The Journal of Industrial Economics*, 44 (1). 17-24.

Chen, Yuxin and Song Yao (2017), "Sequential Search with Refinement: Model and Application with Click-Stream Data," *Management Science*, 63 (12), 4345-4365.

Cui, Hailong, Sampath Rajagopalan, and Amy R. Ward (2020), "Predicting Product Return Volume Using Machine Learning Methods," *European Journal of Operational Research*, 281 (3), 612-627.

Davis, Scott, Michael Hagerty, and Eitan Gerstner (1998), "Return Policies and the Optimal Level of 'Hassle'", *Journal of Economics and Business*, 50 (5), 445-460.

Dukes, Anthony and Lin Liu (2016), "Online Shopping Intermediaries: The Strategic Design of Search Environments," *Management Science*, 62 (4), 1064-1077.

Dzyabura, Daria and John R. Hauser (2019), "Recommending Products When Consumers Learn Their Preference Weights," *Marketing Science*, 38 (3), 417-441.

Dzyabura, Daria, Siham El Kihal, John R. Hauser, and Marat Ibragimov (2023), "Leveraging the Power of Images in Managing Product Return Rates," *Marketing Science*, 42 (6), 1125-1142.

El Kihal, Siham and Edlira Shehu (2022), "It's Not Only What They Buy, It's Also What They Keep: Linking Marketing Instruments to Product Returns," *Journal of Retailing*, 98 (3), 558-571.

Hong, Yili (Kevin) and Paul A. Pavlou (2014), "Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents," *Information Systems Research*, 25 (2), 328-344.

Honka, Elisabeth and Pradeep K. Chintagunta (2013), "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *SSRN Electronic Journal*, Published online October 27, 2013, http://dx.doi.org/10.2139/ssrn.2345449.

Janakiraman, Narayan, Holly A. Syrdal, and Ryan Freling (2016), "The Effect of Return Policy Leniency on Consumer Purchase and Return Decisions: A Meta-Analytic Review," *Journal of Retailing*, 92 (2), 226-235.

Janssen, Maarten and Cole Williams (2023), "Consumer Search and Product Returns in E-Commerce," *American Economic Journal: Microeconomics (forthcoming).*

Ke, T. Tony, Zuo-Jun Max Shen, and J. Miguel Villas-Boas (2016), "Search for Information on Multiple Products," *Management Science*, 62 (12), 3576-3603.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010) “Online Demand Under Limited Consumer Search,” *Marketing Science*, 29 (6), 1001-1023.

Lee, Jinjoo, “Retailers’ Many Unhappy Returns,” *Wall Street Journal*, Published online January 5, 2022, https://www.wsj.com/articles/retailers-many-unhappy-returns-11641387605.

Moorthy, Sridhar and Kannan Srinivasan (1995), “Signaling Quality with a Money-Back Guarantee: The Role of Transaction Costs,” *Marketing Science*, 14 (4), 442-466.

Morozov, Ilya, Stephan Seiler, Xiaojing Dong, and Liwen Hou (2021), “Estimation of Preference Heterogeneity in Markets with Costly Search,” *Marketing Science*, 40 (5), 871-899.

Owen, David, “What Happens to All the Stuff We Return?” *The New Yorker*, Published August 14, 2023, https://www.newyorker.com/magazine/2023/08/21/the-hidden-cost-of-free-returns.

Padilla, Nicolas, Eva Ascarza, and Oded Netzer (2023), “The Customer Journey as a Source of Information,” *Social Science Research Network*, Published online November 21, 2023, https://doi.org/10.2139/ssrn.4612478.

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

Petersen, J. Andrew and V. Kumar (2010), “Can product returns make you money?” *MIT Sloan Management* Review, 51 (3), 95-99.

Petersen, J. Andrew and V. Kumar (2015), “Perceived Risk, Product Returns, and Optimal Resource Allocation: Evidence from a Field Experiment,” *Journal of Marketing Research*, 52 (2), 268-285.

Ren, Qitian and Kinshuk Jerath (2022), “Consumer Search and Product Returns,” *Working Paper.*

Sahoo, Nachiketa, Chrysanthos Dellarocas, and Shuba Srinivasan (2018), "The Impact of Online Product Reviews on Product Returns," *Information Systems Research*, 29 (3), 723-738.

Shehu, Edlira, Dominik Papies, and Soctt A. Neslin (2020), "Free Shipping Promotions and Product Returns," *Journal of Marketing Research*, 57 (4), 640-658.

Stigler, George J. (1961), "The Economics of Information," *Journal of Political Economy*, 69 (3), 213-225.

Ursu, Raluca M., Qingliang Wang, and Pradeep K. Chintagunta (2020), "Search Duration," *Marketing Science*, 39 (5), 849-871.

Ursu, Raluca M., Qianyun Zhang, and Elisabeth Honka (2022), "Search Gaps and Consumer Fatigue," *Marketing Science*, 42 (1), 110-136.

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

Weitzman, Martin L. (1979), "Optimal Search for the Best Alternative," *Econometrica*, 47 (3), 641-654.