

# Estimation of Preference Heterogeneity in Markets with Costly Search\*

**Xiaojing Dong**  
Santa Clara  
University

**Ilya Morozov**  
Stanford University

**Stephan Seiler**  
Stanford University

**Liwen Hou**  
Shanghai Jiao  
Tong University

This draft: March 27, 2018

We study the estimation of preference heterogeneity in markets where consumers engage in costly search to learn product characteristics. Costly search amplifies the way consumer preferences translate into purchase probabilities, generating a seemingly large degree of preference heterogeneity. We develop a search model that allows for flexible heterogeneity in preferences, and estimate its parameters using a unique panel dataset on the search and purchase behavior of consumers. The estimation results reveal that ignoring search costs leads to an overestimation of standard deviations of product intercepts by 30%. We show that this bias leads to incorrect inference about price elasticities and markups of sellers and has important consequences for optimal targeted marketing.

**Keywords:** Consumer Search, Preference Heterogeneity

\*We thank Bart Bronnenberg and Wes Hartmann for insightful comments. We also thank seminar participants at Frankfurt University, Santa Clara, Georgia Tech and Carnegie Mellon as well as conference participants at the 2018 Winter Marketing-Economics Summit. All errors are our own.

# 1 Introduction

One of the most well-known and widely studied stylized facts in quantitative marketing is strong persistence in the product choices of consumers (Rossi and Allenby, 1993; Rossi, McCulloch, and Allenby, 1996; Allenby and Rossi, 1999). Researchers typically interpret this persistence as evidence that consumers have heterogeneous preferences across products.<sup>1</sup> In this paper, we challenge the conventional wisdom that strong persistence in choices is driven by strong preference heterogeneity. We argue that in the presence of search frictions, even small differences in preferences across products can translate into highly persistent choices.

The contribution of this paper is twofold. First, we illustrate how preference heterogeneity and search frictions jointly determine persistence in consumer choices. In the presence of search costs, consumers first search their most preferred product and are less likely to evaluate other options. This search behavior concentrates purchase probabilities around each consumer’s preferred product. A researcher who ignores search frictions may incorrectly infer that consumers strongly prefer certain products, whereas in reality, slight preferences across products are amplified by the presence of search frictions. Because prior research finds that search costs are substantial in different product categories (De Los Santos, Hortacsu, and Wildenbeest, 2012; Seiler, 2013; Honka, 2014; Koulayev, 2014; Giuliatti, Waterson, and Wildenbeest, 2014), we conjecture this amplification effect is an important driver of choice persistence in many settings. As a result, preference heterogeneity in many markets may be less pronounced than previously thought.

Second, we show that accounting for the influence of search costs is quantitatively important. To illustrate this point, we estimate a model of search with flexible preference heterogeneity using search and purchase data from an online retailer. The estimation results reveal that ignoring search frictions leads to a 30% upward bias in the estimated standard deviations of product intercepts. Overestimating heterogeneity generates a seemingly large degree of product differentiation; as a result, we underestimate own-price elasticities and overestimate the markups of firms by 25%. Furthermore, when ignoring search, the scope for targeted marketing is overestimated because consumers appear more heterogeneous in their preference than they actually are. As a consequence, optimal personalized prices derived from a full-information model are more dispersed than prices based on the search model. This difference in prices affects the expected profit from targeting: personalized prices derived from the full-information model increase profits only by 1.1% relative to uniform pricing, whereas prices based on the search model increase profits by 13.2%. Hence, explicitly modeling consumer search is crucial for evaluating the degree of competition among sellers and for the implementation of optimal targeted marketing.

To identify preference heterogeneity separately from search costs, we collect rich data on consumer search and purchase behavior and develop a computationally efficient way to estimate our search model. Because our identification strategy requires panel data on the search and purchase decisions of consumers, we use a dataset in which some consumers search and purchase multiple

---

<sup>1</sup>Switching costs provide another possible explanation for choice persistence (Dubé, Hitsch, Rossi, and Vitorino, 2008; Dubé, Hitsch, and Rossi, 2009). We discuss the role of switching costs in more detail at the end of this section.

times in the same product category. The prior literature on the estimation of preference heterogeneity typically uses panel data on consumer purchases (Rossi and Allenby, 1993; Allenby and Rossi, 1999), which does not allow to explicitly study consumer search. At the same time, most of the search literature analyzes cross-sectional data on search and purchase behavior (Honka, 2014; Chen and Yao, 2016; Honka and Chintagunta, 2015), which does not include a panel dimension. To the best of our knowledge, ours is the first paper that combines these two approaches and uses panel data on search behavior to separately identify preference heterogeneity and search costs.

In terms of computational burden, the estimation of our search model is computationally intense because we have to integrate preference heterogeneity out of the likelihood function. To reduce this computational burden, we implement an importance-sampling estimator similar to the one discussed in Akerberg (2009). This approach removes the need to recompute individual likelihood contributions for each new guess of parameters, thus making estimation significantly faster.

This paper connects two strands of literature. Our model framework is similar to other structural models from the consumer search literature (Kim, Albuquerque, and Bronnenberg, 2010; De Los Santos, Hortacsu, and Wildenbeest, 2012; Honka, 2014; Honka and Chintagunta, 2015; Chen and Yao, 2016). However, these papers estimate search models using cross-sectional data and do not focus on estimating preference heterogeneity. On the other hand, there is a long history of research in marketing regarding estimation and identification of preference heterogeneity in markets where consumers are perfectly informed about product characteristics (Chintagunta, Jain, and Vilcassim, 1991; Rossi and Allenby, 1993; Rossi, McCulloch, and Allenby, 1996; Allenby and Rossi, 1999). We connect these two areas of research by re-examining the estimation of preference heterogeneity in markets with costly search.

Our paper is also related to the broader literature on persistence in consumer choices. Prior research suggests that choice persistence can be rationalized by the presence of preference heterogeneity (see papers cited above) or switching costs (Dubé, Hitsch, Rossi, and Vitorino, 2008; Dubé, Hitsch, and Rossi, 2009). We contribute to this literature by showing that search costs may also play a substantial role in generating choice persistence. It is important to note that the presence of search costs does not by itself generate persistence; instead, search costs amplify the way in which preference heterogeneity translates into persistent choices. To simplify exposition and estimation, in this paper we abstract away from switching costs.<sup>2</sup> Although this assumption may seem strong, prior research suggests that switching costs play a minor role in shaping consumer choices relative to heterogeneity in preferences (Dubé, Hitsch, and Rossi, 2010).

The rest of the paper proceeds as follows. In sections 2 and 3, we develop a search model and analyze the impact of search costs on consumers' purchase decisions and in turn on the estimates of preference heterogeneity. We then adapt the basic search model for a panel data setting and discuss identification in section 4, and describe our estimation strategy in section 5. Section 6 presents the

---

<sup>2</sup>The literature on the estimation of switching costs emphasizes the importance of flexible controls for preference heterogeneity, which are usually implemented in a Bayesian framework (Dubé, Hitsch, and Rossi, 2010). Because modeling search significantly increases the computational burden of estimation, we are not able to allow for the same degree of flexibility with regard to preference heterogeneity.

estimation results, whereas section 7 analyzes implications for optimal pricing. We offer concluding remarks in section 8.

## 2 General Model Framework

Suppose each consumer  $i$  conducts  $T_i$  search sessions, which we index by  $t = 1, \dots, T_i$ . In each session, she chooses exactly one product out of  $J$  available alternatives. No outside good exists, and the utility consumer  $i$  derives from product  $j$  in period  $t$  is the sum of a consumer-specific product intercept and a taste shock:<sup>3</sup>

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

We assume the consumer is imperfectly informed about product-specific utilities and must engage in costly search to resolve uncertainty. In particular, she knows product intercepts  $\xi_{ij}$  and the distribution  $F_\varepsilon(\varepsilon)$  from which the taste shocks are drawn, but must search to learn realizations of taste shocks  $\varepsilon_{ijt}$ . These taste shocks are assumed to be i.i.d. across consumers, search sessions, and products.

Following Weitzman (1979), we assume the consumer searches sequentially and incurs a cost  $c_i$  for each searched product. Upon finishing search, she chooses one product from the searched options. In this setting, the optimal search behavior can be characterized by a simple threshold rule derived in Weitzman's paper. To describe this rule, define reservation utility  $z_{ij}$  for consumer  $i$  and product  $j$  as

$$\int_{z_{ij}}^{\infty} (u_{ijt} - z_{ij}) dF(u_{ijt}) = c_i. \quad (2)$$

The reservation utility  $z_{ij}$  is the level of utility at which the consumer is indifferent between searching product  $j$  and receiving  $z_{ij}$  with certainty.<sup>4</sup> In the optimum, the consumer searches products in order of descending reservation utilities. At each step during this process, she continues searching as long as the maximum realized utility is lower than the reservation utility of the next product in the search sequence; otherwise, she stops searching and purchases the highest-utility product among the searched options.

## 3 Preference Heterogeneity and Search Costs

This section illustrates how preference heterogeneity and search costs jointly drive consumer choices. In addition, it explains how ignoring search costs may bias the estimates of preference heterogeneity.

---

<sup>3</sup>In our empirical model in section 4, we expand this framework by introducing a taste shock known prior to search, and add price to the utility function.

<sup>4</sup>The reservation utility is time invariant because it depends on the distribution but not the realization of the taste shock  $\varepsilon_{ijt}$ .

<b>Panel A: No Search Costs</b>									
Product	<u>True Intercepts</u>		<u>Purchase Probability</u>						
	A	B	A	B	A	B	A	B	
Consumer 1	0	-1	0.76	0.24					
Consumer 2	0	1	0.24	0.76					

<b>Panel B: Search Costs = 0.5</b>									
Product	<u>True Intercepts</u>		<u>Search Probability</u>		<u>Purchase Probability</u>		<u>Inferred Intercepts</u>		
	A	B	A	B	A	B	A	B	
Consumer 1	0	-1	1	0.12	0.91	0.09	0	-1.93	
Consumer 2	0	1	0.12	1	0.09	0.91	0	1.93	

<b>Panel C: Search Costs = 1</b>									
Product	<u>True Intercepts</u>		<u>Search Probability</u>		<u>Purchase Probability</u>		<u>Inferred Intercepts</u>		
	A	B	A	B	A	B	A	B	
Consumer 1	0	-1	1	0.03	0.97	0.03	0	-2.77	
Consumer 2	0	1	0.03	1	0.03	0.97	0	2.77	

Table 1: **The Impact of Search Costs on Purchase Shares: A Simple Example.**

We first describe a simple example with two consumers and two products to build intuition and then illustrate that the same results also apply to more general settings.

### 3.1 Simple Example

Consider a market with two consumers (1 and 2) choosing between two products (A and B). We normalize product intercepts for product A to zero (i.e.  $\xi_{1A} = \xi_{2A} = 0$ ) and assume consumer 1 prefers product A to product B ( $\xi_{1B} = -1$ ), whereas consumer 2 prefers product B to product A ( $\xi_{2B} = 1$ ). Taste shocks  $\varepsilon_{ijt}$  are assumed to be i.i.d. standard normal. Finally, we assume search costs are the same for both consumers and are equal to 0.5.

Our goal is to compare the purchase behavior of consumers under two scenarios: when they have perfect information about products, and when they engage in costly search. We start by considering the perfect-information scenario. In this scenario, the purchase probability of product A of consumer 1 equals

$$PurchProb_{1A}^{PerfectInformation} = Pr(0 + \varepsilon_{iAt} > -1 + \varepsilon_{iBt}) = 0.76, \quad (3)$$

which corresponds to a standard probit probability. We compute purchase probabilities for the other three consumer-product pairs in a similar way; these probabilities are reported in Panel A of Table 1. According to computed values, each consumer purchases the product with the higher intercept with probability 76% and buys the other product in the remaining 24% of cases.

We compare these purchase probabilities with the second scenario, in which consumers have imperfect information and engage in costly search. Unlike in the perfect-information scenario, now consumers actively decide whether to include different products in their consideration sets. For example, according to the optimal search rule, consumer 1 should first search product A and should only search product B if the realized taste shock of product A is below the reservation utility

of product B<sup>5</sup>; hence, the probability of searching product B equals

$$SearchProb_{1B} = Pr(0 + \varepsilon_{iAt} < z_{1B}) = 0.12,$$

where  $z_{1B}$  is the reservation utility from equation (2).<sup>6</sup> If the consumer searches product B, she then decides which product to buy based on the realized utilities of the two products. By contrast, if she does not search product B, she never learns the realization of the shock  $\varepsilon_{iBt}$  and purchases product A. Combining these two observations, we obtain the purchase probability for product A:

$$\begin{aligned} PurchProb_{1A}^{Search} = & Pr(0 + \varepsilon_{iAt} \geq z_{1B}) \\ & + Pr(0 + \varepsilon_{iAt} < z_{1B})Pr(0 + \varepsilon_{iAt} > -1 + \varepsilon_{iBt} \mid 0 + \varepsilon_{iAt} < z_{1B}) = 0.91. \end{aligned}$$

The first term relates to the case in which the consumer only searched product A, whereas the second term denotes the probability of purchasing product A after having searched both products. As before, we repeat this computation for all consumer-product pairs and report the results in Panel B of Table 1. In this scenario with search costs, consumers purchase the product with the higher intercepts with probability 91% and buy the product with the lower intercept only in 9% of cases.

Comparing purchase probabilities under two scenarios reveals that search costs tend to concentrate purchase probabilities around products with higher intercepts. This effect becomes even stronger with larger search costs, as is clear from Panel C of Table 1, where we report purchase probabilities for the case in which search costs are equal to 1 instead of 0.5. Intuitively, higher search costs make consumers less likely to search the products with lower intercepts, thus reducing the purchase probabilities for these products.

As a consequence of the shift in purchase probabilities when search is costly, consumers' choices start to look more extreme. To understand this point, consider how consumers 1 and 2 change their probability of buying product A as we increase search costs. In the perfect-information scenario, the purchase probabilities of both consumers equal 0.76 and 0.24, but these probabilities become 0.91 and 0.09 when search costs equal 0.5, and further increase to 0.97 and 0.03 when search costs increase to 1.0. Hence, higher search costs increase the dispersion of purchase probabilities of product A across consumers. Because the problem is symmetric, the same happens to purchase probabilities of product B.

This heterogeneity in purchase probabilities is driven by both preference heterogeneity and search costs. Ignoring search costs forces us to fully attribute heterogeneity in choices to preferences,

---

<sup>5</sup>Because match values  $\varepsilon_{ij}$  have identical distribution across products, the ranking of reservation utilities in this model coincides with the ranking of product intercepts  $\xi_{ij}$ .

<sup>6</sup>Reservation utilities do not have closed-form solutions for most distributions  $F_\varepsilon(\varepsilon)$ . Using a numerical approximation, we compute that  $z_{1B} = -1.19$ . Therefore, the consumer only searches product B if the realized utility from product A is below  $-1.19$ . Under the normality assumption, the probability of searching product B is then equal to  $\Phi(-1.19) = 0.12$ , where  $\Phi$  denotes the standard normal CDF.

thus generating an upward bias to the estimates of preference heterogeneity. To illustrate, we use purchase probabilities from Table 1 to infer product intercepts of the two consumers assuming perfect information. For example, when the purchase probability of consumer 1 for product B equals 0.91, we infer the intercept  $\hat{\xi}_{1B}$  by solving  $Pr(0 + \varepsilon_{iAt} > \hat{\xi}_{1B} + \varepsilon_{iBt}) = 0.91$ .<sup>7</sup> The intercept of consumer 2,  $\hat{\xi}_{2B}$ , can be computed in a similar fashion. According to the results, the inferred intercepts  $\hat{\xi}_{1B}$  and  $\hat{\xi}_{2B}$  are larger in absolute terms than the true product intercepts. For example, when search costs are 0.5, we infer the intercepts to be 1.93 and -1.93 as opposed to the true intercepts, 1 and -1. This bias is even more extreme when search costs equal 1, in which case the inferred intercepts are 2.77 and -2.77. Overall, these results suggest that ignoring search costs leads to an overestimation of preference heterogeneity.

### 3.2 General Case

The example in the previous section shows that ignoring search frictions may introduce bias in the estimates of preference heterogeneity. In this section, we show that the direction of this bias depends on the structure of consumers' preferences. When products are mainly horizontally differentiated (as they were in the example in the previous section) and consumers disagree on the pre-search rankings of products, the failure to account for search costs leads to overestimation of heterogeneity. By contrast, strong vertical differentiation may generate the opposite effect, leading to a downward bias in heterogeneity estimates. To illustrate, we develop a more general example with many consumers and products, and consider estimation of preference heterogeneity under different configurations of preferences.

We start by analyzing how search costs change consumers' choices in a general case with many products. The effect of search costs on choices can be best illustrated with two extreme cases: when search costs equal zero ( $c_i = 0$ ), and when they are very large ( $c_i \rightarrow \infty$ ). With zero search costs, purchase probabilities are simply given by the standard perfect-information formula similar to equation (3). By contrast, increasing search costs to infinity induces consumers to search only the product with the highest intercept. Searching other products in this case does not generate sufficient benefits to justify paying a large search cost. In the intermediary cases ( $c_i > 0$ ), increasing search costs increases the purchase share of the highest-intercept product and reduces the purchase share of the lowest-intercept product. The effect on purchase shares of all other products is not necessarily monotonic: the purchase probabilities of some products might first increase at low levels of search costs and then decrease as search costs become larger.

Given this behavior of purchase probabilities, whether higher search costs induce choices for a given product to be more heterogeneous across consumers is unclear. For instance, take the case of vertical differentiation where all consumers agree on their preferred product but also do not entirely dislike other products. In this case, all products gain positive market shares when search costs are small; however, for sufficiently large search costs, all consumers buy the same product despite having different preferences. If we ignored search frictions, we would conclude

---

<sup>7</sup>For this exercise, we treat the consumer-specific purchase probabilities as observable.

that consumer preferences are homogeneous despite the fact that true preferences exhibit some amount of heterogeneity. Thus, vertical product differentiation can lead to an underestimation of heterogeneity when we ignore search frictions.

Consider now the case of horizontal differentiation where not all consumers agree on the ordering of products, as in our earlier example. Increasing search costs concentrates purchase probabilities of each consumer around her most preferred product, similar to the vertical case. However, different consumers have different preferred products, so the presence of search costs makes choices of consumers more diverse. If we ignored search frictions, we would find substantial heterogeneity in preferences even when true heterogeneity is minimal. Therefore, ignoring search costs in the horizontal case can lead to an overestimation of preference heterogeneity.

To further explore the horizontal-differentiation case, consider the following simulation exercise for a market with three products.<sup>8</sup> We simulate a large number of consumers and draw a unique set of product intercepts for each of them. Specifically, for two products, we draw independent intercepts from the standard normal distribution, whereas the intercept of the third product is normalized to zero. Because the means of intercepts for all three products are the same, the differentiation of products is purely horizontal. For each consumer, we simulate search behavior for a large number of search sessions and compute consumer-specific purchase probabilities. In the left graph in Figure 1, we plot the distribution of resulting purchase probabilities across consumers for one of the products.<sup>9</sup> The solid line shows the distribution of purchase probabilities in the perfect-information case. Increasing search costs from 0 to 0.2 shifts the probability mass toward the extremes (dashed line). Specifically, the presence of search costs increases the mass of consumers with low and high purchase probabilities and decreases the mass of consumers with intermediate purchase probabilities. As we increase search costs, this pattern becomes more pronounced (dotted line). Further increasing search costs eventually generates extreme purchase behavior: consumers either buy the product with certainty (if it is their most preferred product) or never buy it. In this case, the purchase probability distribution consists of two mass points at zero and 1.

Next, we infer product intercepts assuming perfect information. The right-hand graph in Figure 1 shows the estimated density of inferred intercepts for different levels of search costs for one of the products. The solid line depicts the benchmark case with zero search costs; in this case, the distribution of inferred intercepts coincides with the true distribution. As search costs increase, the density of estimated intercepts shifts from the middle toward the extremes. Interestingly, introducing search costs has the strongest impact on the values close to the mean and affects inferred intercepts in such a way that the distribution becomes bi-modal. Thus, in this example of horizontal differentiation, ignoring search costs, leads to overestimation of preference heterogeneity.

In summary, when products are horizontally differentiated, incorrectly assuming perfect information generates an upward bias in heterogeneity estimates. When products are vertically differ-

---

<sup>8</sup>Simulations with more than three products generate similar qualitative results. To simplify exposition, in this section, we describe the simulation exercise with three products.

<sup>9</sup>Because we chose identical distributions of intercepts, purchase probabilities are distributed identically for all three products.

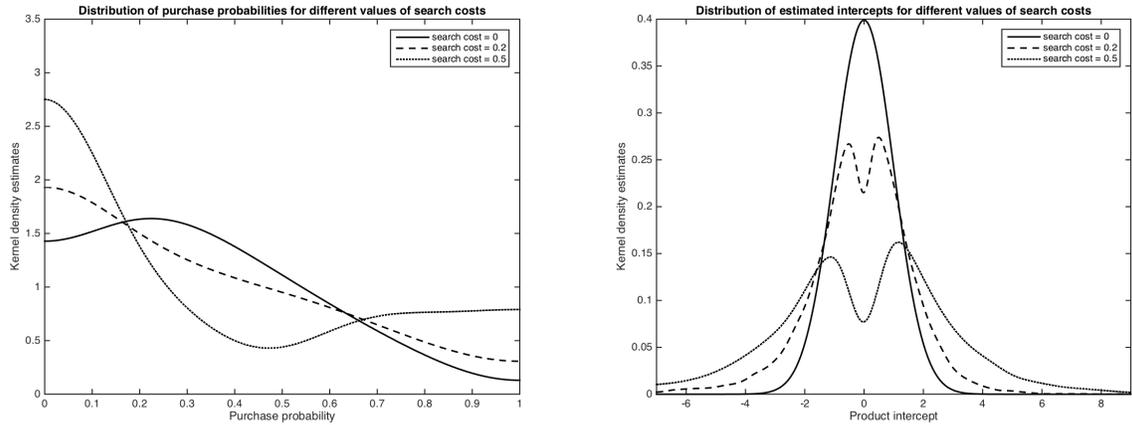


Figure 1: **Distribution of Purchase Probabilities and Estimated Product Intercepts for Different Values of Search Costs.**

entiated, however, ignoring search costs may lead to the opposite effect by generating a downward bias. In Appendix A, we provide more details for the case of vertically differentiated products.

We note the horizontal-differentiation case is more common than one might think. Consider, for example, an extended utility function that includes price. The preference ordering of products then depends on the utility consumers derive from each product net of the disutility from paying the price. In the common situation where higher quality products are sold at higher prices, the degree of horizontal relative to vertical differentiation net of price will be larger than one would infer from the distribution of product intercepts alone. Therefore, even markets with a clear utility ordering of products may exhibit predominantly horizontal differentiation once we account for price differences.

### 3.3 Other Sources of Choice Persistence

In the examples above, we assumed error draws  $\varepsilon_{ijt}$  are uncorrelated over time, that is, across search sessions. Although this assumption is common in many perfect-information demand models, it may feel less tenable here because the information consumers are looking for during search may be persistent over time. If so, then observing match values in the current period may also inform consumers about the distribution of future match values. A consumer who receives a high-match-value draw for a particular product will be more likely to search and purchase that product in the following sessions, whereas receiving a low-match-value draw will dissuade consumers from searching this product in the future. As a result, correlation in match values generates additional persistence: consumers repeatedly search and purchase the same set of products over time, relying on the information they collected in previous sessions.

In Appendix B, we show that allowing for correlated error terms in our model does not eliminate the amplification effect of search costs discussed above. Intuitively, in this extended model, persistence is driven both by time-invariant components (preferences and search costs) and by correlation

of error draws across search sessions. With sufficient data, we can use consumer-specific purchase shares to identify the time-invariant component, whereas any excess persistence in purchases between consecutive time periods should be attributed to correlated match values. We argued in the examples above that the time-invariant component should not be attributed entirely to preferences. The same logic still holds in the model with correlated errors: by fully attributing the time-invariant component to preferences, we will overestimate heterogeneity of consumers' preferences.

The same argument applies to other sources of state dependence including switching costs. With sufficient data, we can isolate the role of switching costs, a specific type of dependence between time periods, from the time-invariant component. The time-invariant component in turn should be explained by some combination of preference heterogeneity and search costs.

## 4 Empirical Model and Identification

To further illustrate that ignoring search costs leads to biased heterogeneity estimates, we now consider an empirical example. To this end, we first extend the sequential search model from section 2, making it suitable to estimation with panel data. The extended model includes an additional taste shock revealed prior to search and adds price to the utility function. We then derive restrictions that observed choices impose on model parameters and realizations of taste shocks and provide an informal discussion of identification.

### 4.1 Panel Data Framework

As before, suppose each consumer  $i$  conducts  $T_i$  search sessions, which we index by  $t = 1, \dots, T_i$ . In each session, she chooses exactly one product out of  $J$  available alternatives. The utility consumer  $i$  derives from purchasing a product  $j$  in search session  $t$  equals

$$u_{ijt} = \delta_{ijt} + \epsilon_{ijt} = (\xi_{ij} - \alpha_i p_{ijt} + \mu_{ijt}) + \epsilon_{ijt}. \quad (4)$$

In this expression,  $\xi_{ij}$  is a time-invariant intercept capturing consumer  $i$ 's preferences for product  $j$ ;  $p_{ijt}$  denotes the price of product  $j$  in session  $t$  of consumer  $i$ ;  $\alpha_i$  denotes the price coefficient; and  $\mu_{ijt}$  and  $\epsilon_{ijt}$  are idiosyncratic taste shocks, both distributed normally and iid with zero mean and variance  $\sigma_\epsilon^2$  and  $\sigma_\mu^2$ . Prior to searching, a consumer knows the realizations of  $\delta_{ijt}$  and the distribution of  $\epsilon_{ijt}$  but has to search in order to learn the realization of  $\epsilon_{ijt}$ .<sup>11</sup> As before, the consumer pays a fixed cost  $c_i$  every time she searches a new product. The model does not include an outside option.

The key difference relative to the framework in section 2 is the pre-search taste shock  $\mu_{ijt}$ . Absent this taste shock, an individual consumer in this model would always search products in the

<sup>10</sup>Consumers who purchase on the same day face identical prices. Price is indexed with an  $i$  subscript because  $t$  denotes a search session rather than calendar time.

<sup>11</sup>This assumption implies consumers observe prices before searching. The assumption is natural in our application, because users of the online store start the session by browsing a list of products that shows product names, photos, and prices. Hence, in our application, prices are known before the consumer starts the search process.

same order in all search sessions. Such a model would be rejected by the data, as we do observe individual consumers changing the order of search. The introduction of the shock  $\mu_{ijt}$  resolves this issue by adding a source of randomness in pre-search utilities. One can interpret the taste shock  $\mu_{ijt}$  as an unobserved information shock that changes the consumer’s propensity to search different products. Such a shock may include recommendations from friends or other product information acquired before the start of the search session. By contrast, the post-search taste shock  $\epsilon_{ijt}$  represents any information retrieved from the product page itself, such as information from customer reviews and detailed product descriptions.

To simplify estimation, we assume shocks  $\mu_{ijt}$  and  $\epsilon_{ijt}$  are independent across search sessions. This assumption effectively precludes consumers from learning over time and implies that new information is available in each search session. One can think of our model as capturing the behavior of frequent website users, who have already gone through a phase of initial exploration and learning during early interactions with the platform. Our model captures the “steady-state” behavior of these users after they have resolved the uncertainty about time-invariant characteristics of products. Hence, they only face uncertainty with regards to product characteristics that are changing over time, which is captured by the post-search shocks  $\epsilon_{ijt}$ . We also note that the key pattern we want to explore, namely, the amplification of preference heterogeneity due to search frictions, is unaffected by possible correlation in match values as we demonstrated in section 3.3. Finally, the assumption of uncorrelated errors is partly driven by data constraints: the relatively short panel dimension of our data precludes us from estimating the correlation in error terms.

## 4.2 Restrictions Imposed by Observed Choices

The search model above imposes a set of restrictions on the model’s parameters and utility shocks. These restrictions are generally associated with three different parts of the observed choices: order of search, decisions to continue or stop searching, and purchase decisions. Below, we derive restrictions of these three types. To simplify exposition, our derivations in this section suppress indices  $i$  and  $t$ .

### *Order of search*

In the optimal search rule, the consumer searches products in order of decreasing reservation utilities (see section 2). Because the shocks  $\epsilon_j$  are distributed identically across products, both reservation utilities  $z_j$  and pre-search utilities  $\delta_j$  are ordered in the same way. Therefore, the probability of observing a particular search order is identical to the probability that the ranking of pre-search utilities  $\delta_j$  is consistent with that order.

To formally state this result, we first introduce additional notation. Let  $\mathfrak{S} = \{1, \dots, J\}$  denote the set of available products, and let  $S \subseteq \mathfrak{S}$  be the set of searched products. Next, let  $M$  be the number of products in this set, so  $|S| = M$ ; and assume the function  $\pi : \{1, \dots, M\} \rightarrow \mathfrak{S}$  describes the observed order of search, so that  $\pi_1$  represents the product searched first,  $\pi_2$  is the product

searched second, and so on.<sup>12</sup>

When the consumer searches in order  $\pi$ , the ranking of pre-search utilities must be consistent with this order:

$$\delta_{\pi_1} \geq \delta_{\pi_2} \geq \dots \geq \delta_{\pi_M} \geq \delta_k \text{ for } \forall k \notin S. \quad (5)$$

Because we only observe search order for products that were actually searched, the data do not impose any restrictions on the ranking of utilities  $\delta_k$  for unsearched products. We only know that the pre-search utility of the product searched last,  $\delta_{\pi_M}$ , exceeds the pre-search utilities of all unsearched products. A more compact way to write inequalities in (5) is

$$\delta_{\pi_m} \geq \max_{l \in S_m} \delta_l \text{ for } m = 1, \dots, M,$$

where  $S_m = \mathfrak{S} \setminus \{\pi_k : k \leq m\}$  represents the set of products left unsearched after searching options  $\pi_1, \dots, \pi_m$ . These inequalities specify the restrictions imposed on parameters and realizations of taste shocks by the observed order of search  $\pi$ .

### *Continuation and stopping decisions*

The consumer continues searching if and only if the maximum realized utility among searched products is lower than the maximum reservation utility among unsearched options. If the consumer decides to search product  $\pi_m$ , it must be that the maximum realized utility among searched products  $\pi_1, \dots, \pi_{m-1}$  is lower than the reservation utility of product  $\pi_m$ , namely,  $z_{\pi_m}$ . This relationship must hold for all searched products except the first one, as our model does not include an outside option, and the consumer always searches at least one product:

$$\max\{u_{\pi_1}, \dots, u_{\pi_{m-1}}\} \leq z_{\pi_m} \text{ for } m = 2, \dots, M. \quad (6)$$

If the consumer decides to stop after searching product  $M$ , the maximum realized utility of searched products must exceed the reservation utilities of all unsearched products<sup>13</sup>:

$$\max\{u_{\pi_1}, \dots, u_{\pi_M}\} \geq \max_{k \notin S} z_k \quad (7)$$

The inequalities in (6) and (7) capture the restrictions the observed continuation and stopping decisions impose on the parameters and realizations of taste shocks.

### *Purchase decision*

Upon finishing the search, the consumer purchases the product with the highest realized utility among the searched options. Therefore, if the consumer searches products  $S$  and buys a certain

<sup>12</sup>Throughout the paper, we use  $\pi_k$  as a shorthand notation for  $\pi(k)$ .

<sup>13</sup>The consumer will also stop searching if she has exhausted all search opportunities. In this case, the stopping inequality is irrelevant.

product  $y \in S$ , then realized utilities must satisfy:

$$u_y \geq \max_{x \in S} u_x. \quad (8)$$

### 4.3 Identification

This section provides an informal discussion of identification and shows how one can use panel data to jointly identify preference parameters and search costs. We consider nonparametric identification of heterogeneity by assuming each consumer has a unique set of preference parameters and showing how panel data can identify these individual-level parameters. Our argument effectively assumes the data include an infinitely large number of search sessions per consumer, allowing us to observe the joint distribution of search and purchase decisions for each consumer.

In a panel dataset, we can clearly distinguish two types of moments in the data: those that identify search costs and those that identify preference parameters. In our model, the order of search depends only on preference parameters, suggesting we can identify preferences of a given consumer by analyzing her typical order of search. Once preferences are identified, we can use data on continuation and stopping decisions to recover search costs. Below, we discuss this identification argument in more detail. We assume for simplicity that price does not enter the utility function, and focus on the identification of product intercepts and search costs.

#### Preference Parameters

In our model, the consumer first searches the product with the highest pre-search utility. Hence, the probability that product  $k$  is searched first equals

$$Pr(\pi_1 = k) = Pr(\delta_k \geq \delta_j \forall j) = Pr(\xi_k + \mu_k \geq \xi_j + \mu_j \quad \forall j) \quad \text{for } \forall k. \quad (9)$$

Under the assumption of normally distributed taste shocks, these expressions are standard probit probabilities.<sup>14</sup> We can invert the system of equations in (9) to express the product intercepts  $\xi_k$  as a function of search probabilities  $Pr(\pi_1 = k)$ , as in Hotz and Miller (1993). Therefore, the knowledge of search probabilities  $Pr(\pi_1 = k)$  is sufficient to identify consumer-specific preference parameters.

The identities of other searched products provide additional information about preference parameters. Based on the inequalities in (5), the consumer chooses to search products in order  $\pi$  with probability

$$\begin{aligned} Pr(\pi) &= Pr(\delta_{\pi_1} \geq \delta_{\pi_2} \geq \dots \geq \delta_{\pi_M} \geq \delta_k \quad \text{for } \forall k \notin S) \\ &= Pr(\xi_{\pi_1} + \mu_{\pi_1} \geq \xi_{\pi_2} + \mu_{\pi_2} \geq \dots \geq \xi_{\pi_M} + \mu_{\pi_M} \geq \xi_k + \mu_k \quad \text{for } \forall k \notin S). \end{aligned} \quad (10)$$

---

<sup>14</sup>The data on first searches in our setting provide information similar to panel data on purchases in a perfect-information model.

This expression suggests that products that, on average, are searched early during the search process should have relatively high product intercepts. Similarly, the intercepts should be low for products that are searched at the end of the search session or not searched at all. Thus, using information about the complete order of search  $\pi$ , not just about products that are searched first, helps to identify preference parameters.

Importantly, the order of search is unaffected by search costs and only depends on the part of utility known prior to search. Hence, the distribution of search orders in (10) identifies preferences separately from search costs.

So far, we have ignored the price coefficient and discussed identification of product intercepts; however, identification of the price coefficient is straightforward. Namely, the extent to which products are searched earlier or later as a function of their price identifies price sensitivity  $\alpha$  as long as price variation is exogenous with respect to taste shocks. We maintain this exogeneity assumption in our empirical application. Because the model includes time-invariant product intercepts, identification of the price coefficient comes from price changes for the same product over time. Hence, our main identifying assumption is that the timing of price changes is uncorrelated with temporary preference shocks.

### Search Costs

After recovering preferences, we can use the average duration of search sessions to identify search costs. In particular, the expected number of searched products conditional on preferences  $\xi$ ,  $E(M|\xi)$ , decreases in search costs:

$$\frac{\partial E(M|\xi)}{\partial c} < 0.$$

To gain intuition, note that reservation utilities of products are decreasing in search costs. With higher reservation utilities, the consumer is less likely to continue searching in each step of the search process, so the expected search duration decreases. This monotonic relationship suggests that we can use the expected duration of the search session to identify search costs.

This identification argument relies on the knowledge of product intercepts  $\xi$ . If product intercepts are unknown, low duration of search can be rationalized both by high search costs and by large dispersion of product intercepts across products. Therefore, conditioning on  $\xi$  is crucial for identifying search costs from the expected length of search sessions.

### The Role of Choice Data

So far, we have shown that information on the order of search and on stopping decisions is sufficient to recover search costs, product intercepts, and the price coefficient. Hence, the choice data are not required for identification of model parameters. Below, we briefly outline an informal argument as to why consumers' choices provide little information, after conditioning on the observed search sequence. A formal treatment of this argument is beyond the scope of this paper.

Consider a market with four products – A, B, C, and D – and suppose a consumer searches in order A, B, and C but does not search D. We can analyze what happens to the purchase probabilities if search costs increase, holding the search sequence constant. Because search costs are now higher, the consumer only finds it optimal to search up to product C if the error realizations  $\varepsilon_A$  and  $\varepsilon_B$  are lower. At the same time, the consumer is more tempted to stop searching at C, so we infer that the error realization  $\varepsilon_C$  for product C is also lower. All in all, the error realizations and hence utility realizations of all three products in the search sequence are now lower, making it unclear how increasing search costs change purchase probabilities.

Similarly, consider increasing the intercept  $\xi_A$  of product A. On the one hand, the purchase probability of product A should increase, because this product has higher expected utility. On the other hand, the consumer finds it optimal to search up to product C only when the realization of  $\varepsilon_A$  is relatively low, implying that the purchase probability of product A should decrease. The net impact of increasing  $\xi_A$  on the conditional purchase probability A is therefore ambiguous and depends on the specific distributional assumptions of the model. In fact, constructing an example in which the purchase probability of product A is non-monotonic in the product intercept  $\xi_A$  is possible.

Hence, once we condition on the observed search sequence, no clear mapping exists from the purchase probabilities to preference parameters or search costs. This observation suggests that choice data provide little additional information to help identify parameters of the model.

## 5 Estimation

### 5.1 Parametrization

The identification argument in the previous section does not impose any restrictions on the distribution of preferences and search costs across consumers. However, in a typical application, the panel dimension of the data is unlikely to be long enough to nonparametrically identify individual preferences, and our application is not an exception. Therefore, we help identification by making several distributional assumptions.<sup>15</sup>

First, we assume that product intercepts  $\xi$  follow a multivariate normal distribution:

$$\xi \sim N(\bar{\xi}, \Sigma_\xi), \tag{11}$$

where  $\bar{\xi}$  is a vector of means and  $\Sigma_\xi$  is a diagonal  $J \times J$  matrix. Second, price sensitivity  $\alpha$  is assumed to be heterogeneous across consumers, capturing differences in the marginal utility of income. Specifically,  $\alpha$  follows a log-normal distribution to ensure the price coefficient is non-negative for all consumers:

---

<sup>15</sup>In Appendix D, we assess (parametric) identification in our particular setting. Specifically, we estimate our model based on a simulated dataset that mimics our actual data in terms of the number of consumers and number of search sessions per consumer. We also employ the same distributional assumptions regarding various model parameters that are outlined in this section.

$$\log \alpha \sim N(\bar{\alpha}, \sigma_\alpha^2), \quad (12)$$

where  $\bar{\alpha}$  and  $\sigma_\alpha^2$  denote the mean and the variance. Similarly, search costs  $c$  follow a log-normal distribution with mean  $\bar{c}$  and variance  $\sigma_c^2$ :

$$\log c \sim N(\bar{c}, \sigma_c^2). \quad (13)$$

We do not model observed heterogeneity for several reasons. In our empirical application, we do not have access to any demographic variables that might be used to model observed heterogeneity in preferences. This situation is relatively common in online markets where firms tend to have rich information on consumers' purchase histories but relatively little information on their demographic profiles. Furthermore, many studies in the literature on preference heterogeneity show that demographic variables explain only a small share of preference heterogeneity relative to the unobserved component (Rossi and Allenby (1993), Rossi, McCulloch, and Allenby (1996), Allenby and Rossi (1999)). Hence, modeling observed heterogeneity may not be crucial for explaining the observed behavior of consumers.

Finally, we fix the variances of taste shocks  $\mu$  and  $\epsilon$  by assuming  $\sigma_\mu^2 = 0.1$  and  $\sigma_\epsilon^2 = 1$ . Although the first normalization is necessary because it fixes the scale of utility, the second one is not required. We could in principle estimate  $\sigma_\epsilon^2$  from the data. However, increasing the variance of  $\epsilon$  leads to an increase in the benefits from search, which makes consumers search more. Reducing search costs  $c$  affects consumers' behavior in a similar way because it also leads to more search. Thus, the variance of the post-search shock can be identified separately from search costs only through functional form. To solve this issue, in our estimation, we fix  $\sigma_\epsilon^2$ . We also re-estimate the model based on different normalization of  $\sigma_\epsilon^2$  and find this approach has little impact on our key counterfactuals.<sup>16</sup>

## 5.2 Likelihood Function

To derive the likelihood function, we first summarize the restrictions that observed choices impose on model parameters and realizations of taste shocks. Let vector  $\theta = (\xi, \alpha, c)$  denote the consumer's type that describes her product intercepts, price sensitivity, and search costs. Additionally, let  $p$  denote the  $J \times 1$  vector of product prices. If a consumer with type  $\theta$  searches products  $S$  in order  $\pi$  and purchases product  $y \in S$ , the following inequalities must hold:

$$w_m^O(\theta, \mu, \epsilon, p) = \delta_{\pi_m} - \max_{l \in S_m} \delta_l \geq 0 \quad \text{for } m = 1, \dots, M \quad (14)$$

$$w_m^S(\theta, \mu, \epsilon, p) = z_m - \max\{u_{\pi_1}, \dots, u_{\pi_{m-1}}\} \geq 0 \quad \text{for } m = 2, \dots, M \quad (15)$$

---

<sup>16</sup>When changing the normalization of the variance by factor  $k$ , the estimated search cost changes (as expected) by roughly a factor  $1/k$ . Because search costs and the post-search-shock variance are not fully co-linear, other parameter estimates also change, but only very slightly.

$$w_{M+1}^S(\theta, \mu, \epsilon, p) = \max\{u_{\pi_1}, \dots, u_{\pi_M}\} - \max_{k \notin S} z_k \geq 0 \quad (16)$$

$$w^P(\theta, \mu, \epsilon, p) = u_y - \max_{x \in S} u_x \geq 0, \quad (17)$$

where (14) corresponds to the order of search, (15) and (16) describe inequalities for continuation and stopping decisions, and (17) contains purchase inequalities. To simplify notation, define function  $W(\theta, \mu, \epsilon, p)$  as

$$W(\theta, \mu, \epsilon, p) = \min\{w_1^O, \dots, w_M^O, w_2^S, \dots, w_{M+1}^S, w^P\}. \quad (18)$$

Note that  $W(\theta, \mu, \epsilon, p) \geq 0$  if and only if all the inequalities in (14)-(17) hold.

Adding subscripts  $i$  for consumers and  $t$  for search sessions, we now formulate the likelihood of the model. Suppose a researcher observes data on search and purchase decisions as well as prices of products at the time of each decision; we let  $D_{it}$  denote these data for consumer  $i$  and session  $t$ , and we let  $D$  denote the data for all consumers and sessions. Suppose the data  $D$  include  $N$  consumers, and each consumer  $i$  conducts  $T_i$  search sessions. The likelihood of observing search and purchase decisions of consumer  $i$  in period  $t$  given consumer's type  $\theta_i$  and prices  $p_{it}$  is

$$L_{it}(D_{it}|\theta_i, p_{it}) = Pr(W_{it}(\theta_i, \mu_{it}, \epsilon_{it}, p_{it}) \geq 0|\theta_i, p_{it}), \quad (19)$$

where the uncertainty on the right-hand side comes from taste shocks  $\mu_{it}$  and  $\epsilon_{it}$ , both of which are unknown to an econometrician. The vector of prices is indexed by  $i$  because consumers conduct their search sessions at different times.

Combining these individual likelihoods for different search sessions, we obtain the full likelihood of the data  $D$ :

$$L(D|\Omega, p) = \prod_{i=1}^N \int \left( \prod_{t=1}^{T_i} L_{it}(D_{it}|\theta_i, p_{it}) \right) \cdot dF(\theta_i|\Omega), \quad (20)$$

where  $\Omega = (\bar{\xi}, \bar{\alpha}, \bar{c}, \text{diag}(\Sigma_\xi), \sigma_\alpha, \sigma_c)$  denotes the vector of parameters describing the distribution of types  $\theta_i$ , and  $p$  summarizes price vectors different consumers faced in different search sessions. Because types  $\theta_i$  are unobserved, the likelihood function in (20) contains an expectation of consumer-specific likelihoods with respect to the joint distribution of  $\theta_i$ . This joint distribution,  $F(\theta_i|\Omega)$ , is fully defined by the distributional assumptions in (11)-(13).

### 5.3 Importance Sampling

A straightforward approach to estimation would be to maximize the simulated version of the likelihood (20) with respect to parameters  $\Omega$ . In our application, this approach is impractical because we need to recompute the simulated likelihood for each new guess of parameters; this repeated computation tends to be computationally burdensome because we have to recompute reservation

utilities and simulate consumers' decisions for each draw of types and taste shocks. To solve this issue, we estimate the model using the importance-sampling method proposed in Akerberg (2009). First, we rewrite the likelihood function, multiplying and dividing the consumer and period-specific likelihoods by some function  $g(\theta_i)$ :

$$L(D|\Omega, p) = \prod_{i=1}^N \int \left( \prod_{t=1}^{T_i} L_{it}(D_{it}|\theta_i, p_{it}) \cdot \frac{f(\theta_i|\Omega)}{g(\theta_i)} \right) g(\theta_i) d\theta_i$$

Importantly, the function  $g(\theta_i)$ , the so-called proposal density, does not depend on the values of unknown parameters  $\Omega$ . This observation suggests that instead of drawing consumers' types from  $f(\theta_i|\Omega)$ , we can take  $N_M$  draws of types from the proposal density  $g(\theta_i)$  and approximate the likelihood function with a simulated counterpart:

$$\tilde{L}(D|\Omega, p) = \prod_{i=1}^N \frac{1}{N_M} \sum_{m=1}^{N_M} \left( \prod_{t=1}^{T_i} \tilde{L}_{it}(D_{it}|\theta_i^m, p_{it}) \cdot \frac{f(\theta_i^m|\Omega)}{g(\theta_i^m)} \right), \quad (21)$$

where  $\theta_i^m$  denotes the  $m$ -th draw of types for consumer  $i$ , and  $\tilde{L}_{it}(D_{it}|\theta_i^m, p_{it})$  is the simulated consumer and session-specific contribution to the likelihood:

$$\tilde{L}_{it}(D_{it}|\theta_i^m, p_{it}) = \frac{1}{N_S} \sum_{s=1}^{N_S} 1 \{W_{it}(\theta_i^m, \mu_{it}^s, \epsilon_{it}^s, p_{it}) \geq 0\}. \quad (22)$$

In this expression,  $\mu_{it}^s$  and  $\epsilon_{it}^s$  are draws of taste shocks. We estimate parameters  $\Omega$  by maximizing the simulated expression  $\tilde{L}(D|\Omega, p)$  in (21). Appendix C describes details of the estimation procedure.

The importance-sampling approach has several important advantages over a simple frequency estimator. First, importance sampling significantly reduces the computational burden of estimation. The only part of the likelihood in (21) that depends on parameters  $\Omega$  is the weights  $f(\theta_i^m|\Omega)/g(\theta_i^m)$ , so recomputing likelihood contributions  $\tilde{L}_{it}(D_{it}|\theta_i^m, p_{it})$  for each new guess of parameters  $\Omega$  is not necessary. Instead, we can precompute these likelihood contributions for draws of types  $\theta_i^m$  and taste shocks  $\mu_{it}$  and  $\epsilon_{it}$ , and use these precomputed values during maximization. Precomputing likelihood contributions significantly reduces the computational burden and allows us to use a relatively large number of draws per consumer. In our application, we set  $N_S = N_M = 100$ , which results in at least 10,000 simulation draws per consumer.

The second advantage of the importance sampling method is that the resulting objective function is smooth in parameters because the weights  $f(\theta_i^m|\Omega)/g(\theta_i^m)$  in (21) are continuous and differentiable in  $\Omega$ . This smoothness allows us to use derivative-based optimization methods to estimate the model. Note that a frequency estimator would generate a discontinuous function, making gradient-based methods impractical.<sup>17</sup>

<sup>17</sup>An alternative approach to smoothing the objective function is to use a kernel-smoothed frequency estimator as in Honka (2014) and Honka and Chintagunta (2015); however, in our application, we found such an approach to work significantly slower than importance sampling as it requires recomputing likelihoods for each new guess of

One practical consideration is the choice of proposal density  $g(\theta_i)$ . Ideally, we would choose the proposal density that coincides with the true density of types  $\theta_i$ ,  $f(\theta_i|\Omega_0)$ . Because we do not know the true values of parameters  $\Omega_0$ , in practice, we choose  $g(\theta_i) = f(\theta_i|\tilde{\Omega}_0)$ , where  $\tilde{\Omega}_0$  are initial values of parameters for our estimation. To obtain reasonable initial values, we choose  $\tilde{\Omega}_0$  that roughly matches four types of moments to their data counterparts: purchase probabilities  $Pr(y_{it} = k)$ , probabilities of first searches  $Pr(\pi_{1,it} = k)$ , persistence in purchases  $Pr(y_{it} = k|y_{i,t-1} = k)$ , and expected duration of search sessions  $E(M_{it})$ .

## 6 Data and Results

### 6.1 Data and Descriptive Statistics

We use data from the Chinese online store of a large international chain of cosmetic stores. The layout of the store’s website is similar to other online retailers. Figure 2 shows the typical layout of an online store that looks similar to the one we analyze. Consumers can enter search terms or use category tags or filters to narrow down the set of relevant products. Consumers are then presented with a list of products that contains basic information including product names, prices, photos, and short descriptions. They can then click on each product to visit the product page, which contains a more detailed product description as well as customer reviews.

The sample covers the full year of 2014 and includes detailed information on browsing and purchase activity of all website users during this year. We observe the date and time of each page visit, types of visited pages (main page, product page, shopping cart, etc), and identities of products described on each visited page. In addition, we have information on purchases made through the website. Unique consumer identifiers allow us to follow consumers over time and match their browsing activity to related purchases.<sup>18</sup>

In our analysis, we focus on the category of moisturizers. Moisturizers are cosmetic products designed to prevent and treat dry skin, protect sensitive skin, and improve skin tone and texture. The category is suitable for our analysis because consumers purchase moisturizers on a regular basis, so we can observe some of them making several purchases in this category. Observing repeated search sessions and purchases is crucial for our estimation strategy because the panel dimension of the data helps identify the time-invariant component of preferences. Another advantage of choosing moisturizers is a relatively small number of products in this category. To simplify estimation, we focus on the 10 most popular moisturizers, which account for more than 60% of category sales.

We consider a consumer to be searching a given product if she visits the corresponding product page. A search session is defined retrospectively from the purchase in which it ended. Specifically, the search session includes all searches that occurred within one week before a given purchase.<sup>19</sup>

---

parameters.

<sup>18</sup>Our data provider is tracking the browsing and purchase activity of consumers, using browser cookies. Each time a consumer logs in, the website ties browsing history from a given cookie to the account of this consumer. In our data, we observe anonymized identifiers of consumers that correspond to unique accounts.

<sup>19</sup>Over 80% of all searches occur within one week before some purchase. When two purchases of the same consumer

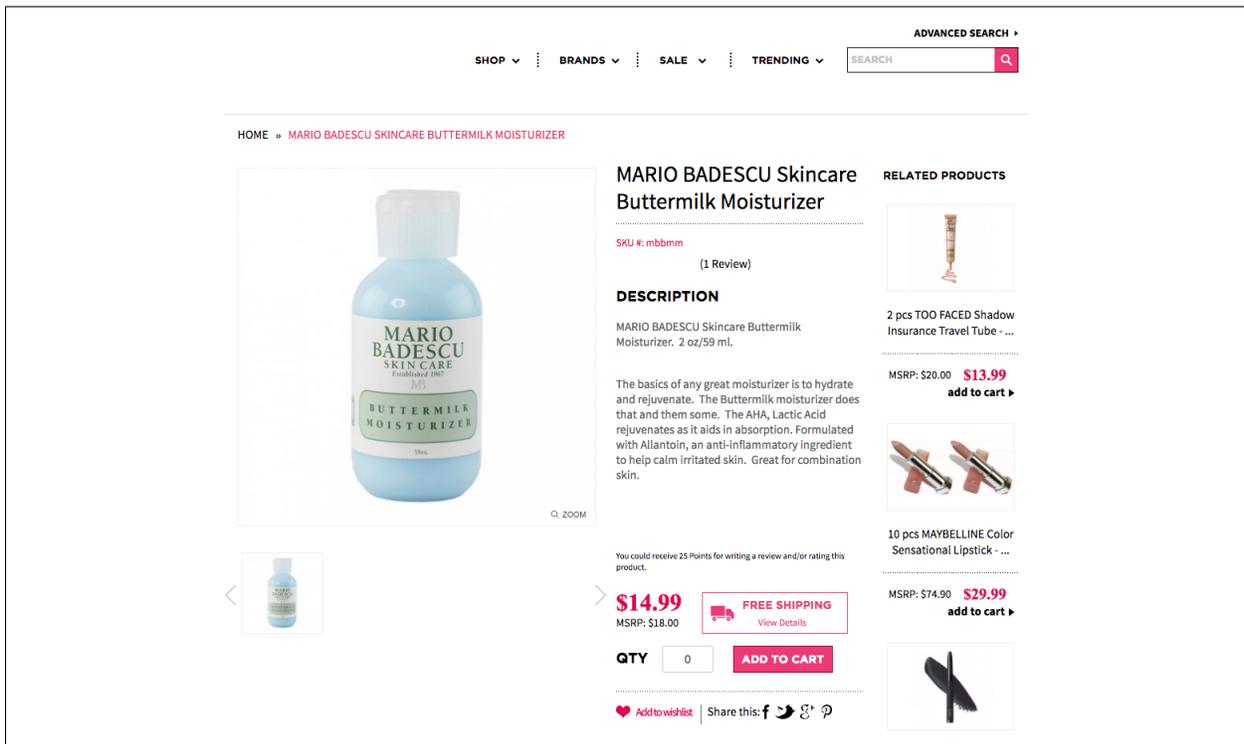
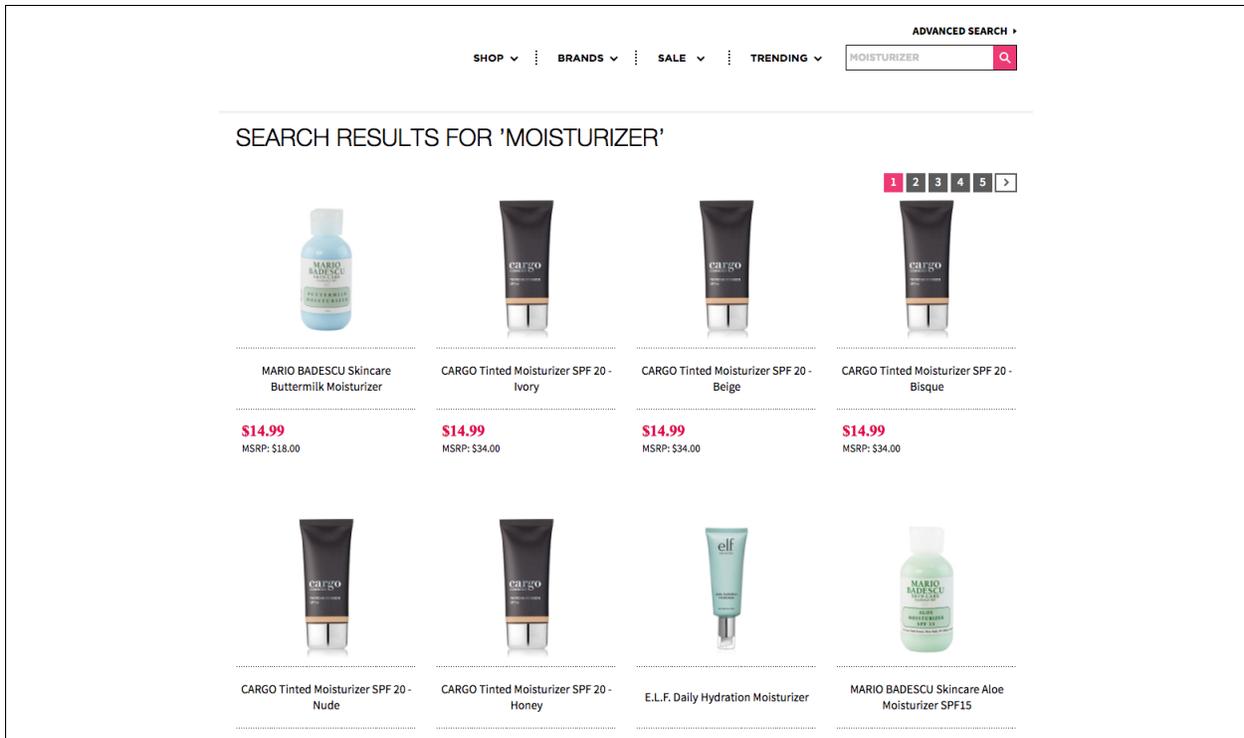


Figure 2: **Example of Webpage Layout.** The two figures show the layout of a page with search results (top figure) and a product page (bottom figure) from an online store similar to the one in our sample.

are less than seven days apart, we attribute only searches that occurred between two purchases to the later purchase.

---

**Panel A: Search Behavior**

Average Number of Searches per Session		1.37
Number of Searches	1	72.82
(Percentage)	2	19.98
	3	5.44
	$\geq 4$	1.77

---

**Panel B: Persistence**

Purchase in $t =$ Purchase in $t-1$	0.67
First search in $t =$ First search in $t-1$	0.62

---

**Panel C: Product-Level**

Product	Brand	Avg. Price (Yuan)	Purch. Share	Search Share	Repeat Purch. Prob.	Repeat Search Prob.	Repeat Search Prob. ( $>1$ Prod. Searched)
1	Caudalie	249.9	0.220	0.271	0.73	0.73	0.55
2	Clinique	444.5	0.213	0.275	0.56	0.63	0.52
3	Laneige	215.0	0.093	0.108	0.75	0.80	0.64
4	Clinique	340.0	0.090	0.183	0.57	0.57	0.45
5	SK-II	650.0	0.079	0.106	0.58	0.63	0.38
6	Caudalie	284.0	0.072	0.106	0.73	0.63	0.52
7	Loccitane	260.0	0.070	0.083	0.68	0.69	0.27
8	Caudalie	226.5	0.059	0.089	0.79	0.67	0.36
9	For Beloved One	560.0	0.056	0.066	0.52	0.42	0.24
10	Clinique	340.0	0.048	0.080	0.70	0.47	0.33

---

Table 2: **Descriptive Statistics.**

Because our empirical model does not include an outside option, we do not consider searches that did not end in a purchase. The final sample includes 4,010 search sessions conducted by 3,577 consumers. Of all consumers, 358 (10%) made at least two purchases, and some consumers bought as many as eight moisturizers during the sample period.

Table 2 provides descriptive statistics on search behavior and presents some evidence of preference heterogeneity. Turning to search behavior (see Panel A of Table 2), we note that consumers search little, and the average search session contains only 1.37 searched products. Out of all search sessions, 72.8% finish after one search, 20% finish after two searches, and the remaining 7.2% end after three or more searches. This observation suggests search costs are high or benefits from search are limited (i.e., the variance of post-search shocks  $\epsilon_{ijt}$  is small). Furthermore, the number of searches per session varies significantly across consumers, indicating heterogeneity in search costs and search benefits.

Next, we investigate persistence in search and purchase behavior across different search sessions of the same consumer (Panel B of Table 2). Consumers tend to make persistent decisions: 67% of search sessions end in the same purchase as the previous search session. Moreover, search behavior exhibits a similar persistence, as 62% of the first searches are products that have also been searched first during the previous session. The persistence in both purchase and search behavior indicates substantial preference heterogeneity across consumers.

Finally, Panel C of Table 2 presents descriptive statistics at the product level. Prices vary substantially across products and range from 215 to 650 yuan (US\$34.4 and US\$104). Market shares also differ significantly, indicating some degree of vertical differentiation. We also report measures of repeat purchase and search behavior at the product level in the last three columns. Interestingly, even products that garner only a small market share tend to be purchased by consumers who have already purchased the same product in the past. For instance, the repeat purchase probability for the most popular product (22% market share) and the least popular product (4.8% market share) are almost identical: 0.73 versus 0.70.

## 6.2 Estimation Results

Table 3 presents estimation results from the structural model of search. The estimated parameters include means and standard deviations of product intercepts  $\xi$  for nine products (the intercept for the first product is normalized to zero) as well as parameters that determine the distribution of price coefficients  $\alpha$  and search costs  $c$ . To ease interpretation, we report the mean and standard deviation of the price coefficient and search costs rather than the location and variance parameter of the corresponding log-normal distribution. Note we cannot easily monetize the search-cost estimate, because search costs are identified only in relation to the fixed magnitude of the normalized variance  $\sigma_c^2$  (see the discussion in section 5.1). We also present estimation results for a demand model with perfect information in the last three columns of Table 3. The perfect-information model does not include search costs but is otherwise identical to the search model. To estimate this model, we use only purchase data and ignore data on search.

Our discussion in section 3 suggests that ignoring search costs may lead to overestimation of preference heterogeneity. To establish whether this bias arises in our setting, we first need to make the estimated parameters from both models comparable. Due to different normalizations, the natural way to compare estimates is to monetize parameter estimates.<sup>20</sup> We provide the monetized values of product intercepts (net of price) and standard deviations for the two models in columns (3) and (6) of Table 3.<sup>21</sup> According to the results, the variances of product intercepts are almost uniformly overestimated when we ignore search costs: on average, standard deviations are

<sup>20</sup>Because the price coefficient is heterogeneous as well, we take draws from the distribution of estimated parameters and compute monetized preference parameters for each set of draws. We report the average value of monetized preference parameters across simulation draws.

<sup>21</sup>As noted above, we do not monetize the search-cost parameter due to the normalization of the post-search shock variance.

	<i>Search Model</i>			<i>Model w/o Search</i>		
	Coefficients	Standard Errors	Monetized Parameters (Units: Yuan)	Coefficients	Standard Errors	Monetized Parameters (Units: Yuan)
$\bar{\xi}_2$	-0.270	0.040	-49.2	-1.078	0.078	-89.6
$\bar{\xi}_3$	-3.281	0.188	-598.9	-10.190	0.358	-849.4
$\bar{\xi}_4$	-0.686	0.043	-125.0	-1.682	0.064	-140.1
$\bar{\xi}_5$	-0.898	0.055	-163.9	-2.456	0.098	-204.7
$\bar{\xi}_6$	-0.849	0.048	-155.1	-3.276	0.129	-272.7
$\bar{\xi}_7$	-2.386	0.131	-435.4	-6.194	0.220	-516.7
$\bar{\xi}_8$	-3.321	0.138	-605.9	-7.208	0.249	-601.0
$\bar{\xi}_9$	-1.307	0.052	-238.3	-3.358	0.097	-280.0
$\bar{\xi}_{10}$	-1.684	0.078	-307.5	-4.667	0.168	-388.6
$\sigma_{\xi_2}$	0.800	0.051	146.9	2.309	0.099	197.6
$\sigma_{\xi_3}$	2.963	0.142	546.7	9.019	0.284	789.0
$\sigma_{\xi_4}$	0.841	0.040	154.9	1.604	0.059	139.8
$\sigma_{\xi_5}$	0.904	0.041	166.6	2.386	0.076	208.1
$\sigma_{\xi_6}$	0.807	0.028	148.6	3.088	0.090	269.3
$\sigma_{\xi_7}$	1.931	0.088	356.8	5.086	0.149	446.9
$\sigma_{\xi_8}$	2.383	0.096	441.1	5.243	0.149	465.0
$\sigma_{\xi_9}$	1.062	0.045	196.2	2.811	0.067	246.8
$\sigma_{\xi_{10}}$	1.204	0.072	222.8	3.434	0.109	303.9
$Mean(\alpha)$	0.0055	0.0005	-	0.0125	0.0005	-
$StdDev(\alpha)$	0.0005	0.0002	-	0.0027	0.0003	-
$Mean(c)$	0.411	0.067	-	-	-	-
$StdDev(c)$	0.042	0.015	-	-	-	-

Table 3: **Estimation Results.** The unit of observation is a search session. The sample contains 3,577 consumers and 4,010 search sessions. Monetized intercepts are reported net of price, i.e.,  $\bar{\xi}_j^{Monet} = (\bar{\xi}_j/\alpha) - p_j$ .

overestimated by 29.8%.<sup>22</sup>

### Model fit

We examine model fit by comparing several key predictions with their empirical counterparts. For this purpose, we split the sample of 3,577 consumers into a training and a holdout sample. The training sample includes 1,800 randomly selected consumers, whereas the remaining consumers constitute the holdout sample. We find that parameter estimates for the training sample are similar to those based on the full sample. Table 4 reports the simulated predictions and their empirical counterparts for the hold-out sample.

<sup>22</sup>Out of nine variance terms, only the variance for product 4 is slightly larger in the search model.

I. Purchase probabilities			II. Search probabilities (based on first searches)		
Product	Data	Simulated	Product	Data	Simulated
1	0.213	0.226	1	0.206	0.214
2	0.214	0.207	2	0.211	0.208
3	0.094	0.091	3	0.087	0.097
4	0.100	0.111	4	0.090	0.101
5	0.077	0.073	5	0.092	0.084
6	0.075	0.068	6	0.076	0.058
7	0.068	0.069	7	0.072	0.073
8	0.059	0.054	8	0.069	0.060
9	0.058	0.053	9	0.053	0.052
10	0.042	0.047	10	0.043	0.055

III. Persistence Prob.			VI. Number of searches	
	Data	Simulated	Average	Simulated
Purchase	0.643	0.553	1.448	1.414
First Searches	0.692	0.785		

Table 4: **Out-of-sample Fit of the Search Model.** We estimate the model on a training sample, which includes 1,800 randomly selected consumers. The holdout sample consists of the remaining 1,777 consumers.

We first evaluate the fit in terms of market shares for both first searches and purchase behavior in panels I and II and find that the model predictions fit the holdout data reasonably well. Similarly, the average number of searches predicted by the model is similar to the actual search-spell length in the holdout sample. The model does slightly worse in terms of fitting persistence in searches and purchases. Persistence in purchases is underestimated, whereas persistence in search is overestimated. We note the model contains no parameter that specifically caters to matching persistence, such as a switching cost. Hence, the degree of persistence predicted by the model is entirely driven by preference heterogeneity and search costs.

## 7 Optimal Pricing

### 7.1 Price Elasticities and Markups

The bias in heterogeneity estimates may affect our inference about price elasticities and market power of firms. Because we overestimate the degree to which preferences are different across consumers, we mistakenly conclude that most consumers have strong preferences for certain products and therefore respond little to price changes.<sup>23</sup> As a result, we underestimate demand elasticities and arrive at the wrong conclusion that competition among firms is mild.

This point is illustrated in Table 5, where we report price elasticities from the search and perfect-

<sup>23</sup>Several authors have emphasized that higher dispersion of consumer preferences often leads to less elastic demand (Anderson and Renault, 1999; Zhou, 2017).

Product	Own-price Elasticities	
	Search	No Search
1	-2.317	-1.935
2	-2.933	-2.504
3	-0.675	-0.506
4	-2.732	-3.001
5	-4.761	-4.175
6	-2.621	-1.723
7	-1.292	-1.112
8	-0.985	-0.997
9	-4.125	-3.517
10	-2.640	-2.081

Table 5: **Estimated Own-price Elasticities.**

information models. Consistent with the logic above, the perfect-information model underestimates own-price elasticities of all 10 products. This bias in elasticities, in turn, affects our inference about markups of firms: under the assumption of Bertrand-Nash pricing, our estimation results from the perfect-information model imply an average markup of 226.6 yuan, whereas the search model predicts an average markup of only 181.8 yuan. Therefore, by ignoring search, we overestimate markups of firms approximately by 25%. Overall, our results suggest that, to correctly understand the competitive environment, one needs to be aware of search frictions and the way these frictions affect pricing strategies. If a researcher incorrectly infers from high choice persistence that products are highly differentiated, he will tend to underestimate how strongly firms compete with each other.

From a managerial perspective, these results also imply that firms should set lower prices under the search model than under the perfect-information model.

## 7.2 Personalized Prices

Because targeted marketing generates higher profits when consumers have substantially different preferences, the optimal way to target consumers critically depends on the estimates of preference heterogeneity. Overestimation of preference heterogeneity when ignoring search therefore leads to suboptimal targeting. To illustrate this point, we consider the example of personalized pricing in which the firm charges different prices to different consumers based on their histories of searches and purchases.

We first develop a general framework to analyze personalized pricing under different models of consumer behavior and for different consumer histories. To this end, we start by deriving the profit function of the firm for a given consumer  $i$  (the expressions below omit index  $i$  to avoid clutter). Suppose the firm observes a history of decisions taken by a given consumer. The history consists of all products searched in previous sessions, the order in which these products were searched, and purchase decisions. Let  $H = \{y_t, \pi_t\}_{t=1}^T$  denote this history, where  $T$  is the number of search sessions the consumer conducted in the past; and let  $p^H = \{p_{1t}, \dots, p_{Jt}\}_{t=1}^T$  denote prices the

consumer faced in these previous search sessions. The firm infers consumer type  $\theta$  from history  $H$  and  $p^H$ . Assuming the prior distribution of the firm coincides with the true distribution of types  $f(\theta|\Omega_0)$ , the posterior distribution of types  $\theta$  given history  $H$  follows Bayes' rule:

$$f(\theta|H, p^H, \Omega_0) = \frac{L(H|p^H, \theta)f(\theta|\Omega_0)}{\int L(H|p^H, \theta)dF(\theta|\Omega_0)}, \quad (23)$$

where  $L(H|p^H, \theta)$  is the likelihood of observing history  $H$  given historical prices  $p^H$  and type  $\theta$ , and  $\Omega_0$  is a vector of the true values of structural parameters. Given this posterior distribution of types,  $f(\theta|H, p^H, \Omega_0)$ , the firm sets the price  $p_j$  that maximizes its expected profits:

$$\Pi(p_j, H) = \sum_{k \in \mathcal{F}} \int (p_k - mc_k) Pr(y = k|\theta) f(\theta|H, p^H, \Omega) g(\Omega) d\theta d\Omega. \quad (24)$$

Here,  $p_k$  is the price of product  $k$ ,  $mc_k$  denotes marginal costs,  $\mathcal{F}$  describes the set of all products sold by the firm, and  $Pr(y = k|\theta)$  is the purchase probability conditional on the consumer's type. The firm faces two sources of uncertainty: about type  $\theta$ , which is captured by the posterior density  $f(\theta|H, p^H, \Omega)$ ; and about preference parameters  $\Omega$ , which is reflected in the density  $g(\Omega)$ . We approximate posterior beliefs about type  $\theta$  by taking draws of types from the distribution  $f(\theta|H, p^H, \Omega)$ , and  $g(\Omega)$  is approximated using the asymptotic distribution of the maximum likelihood estimates  $\hat{\Omega}$  (see Appendix E for details). In addition, we infer marginal costs  $mc_j$  by assuming the current prices arise from a Nash Equilibrium. When inferring marginal costs, we assume firms set their prices based on the perfect-information demand model.

To evaluate the profits from different pricing regimes, we use the expected profit function based on the search model and histories of search and purchase behavior. By construction, this profit function is maximized when the seller sets personalized prices based on the search model and full histories of searches and purchases. Below, we derive the optimal personalized prices  $p_{ij}$  for each product separately, assuming the firm sets prices for product  $j$  while holding fixed the (uniform) prices of competing products and prices of other products sold by the same firm. This simplifying assumption allows us to abstract away from potential equilibrium effects and focus on assessing the scope for targeting.

As a benchmark, we first compute personalized prices based on the perfect-information model. Specifically, we use estimates from the perfect-information model and infer consumer types  $\theta$  only from purchase histories  $H = \{y_t\}_{t=1}^T$ . Figure 3 shows the resulting distribution of personalized prices for one product in our sample.<sup>24</sup> The distribution is bimodal with one mode located to the left, and the other to the right of the current uniform price of 650 yuan. Intuitively, consumers roughly fall into two groups. One group consists of consumers who have bought the product in the past; we infer that these consumers have large intercepts and hence large willingness to pay for the product, which suggests setting a high price. The other group consists of consumers who have

<sup>24</sup>The graph refers to product 5, which has an average price of  $\bar{p}_5 = 649.9$  during our sample period, and for which the inferred marginal cost equals  $\hat{mc}_5 = 493.5$ .

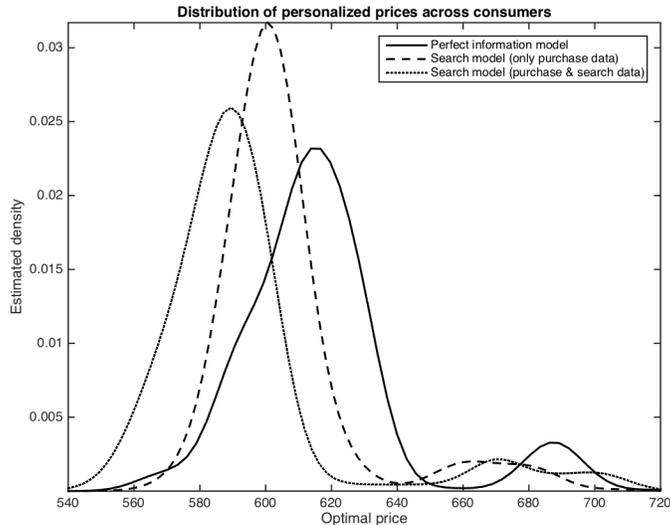


Figure 3: **Distribution of Personalized Prices.** The graph plots the kernel densities of personalized prices under three scenarios: (a) perfect information model, (b) search model in which types are inferred based only on purchase data, and (c) search model in which types are inferred from both search and purchase data.

never bought the product. Hence, these consumers most likely have a low willingness to pay and should be charged a low price. To quantify the losses from using this incorrect model, we insert perfect-information personalized prices for each product into the profit function from the search model and report the resulting change in profits in Table 6. The results show that targeting based on the incorrect model increases expected profits on average by 1.1% relative to uniform pricing.

We next analyze the personalized prices derived from a model of consumer search. To isolate the effect of switching to the right model of consumer behavior, we only allow the firm to set prices based on purchase histories as in the previous exercise. The results in Figure 3 suggest the distribution of personalized prices has a similar shape as the one generated by the perfect-information model but is shifted to the left. Consistent with the discussion in the previous section, this shift occurs because the search model predicts lower price elasticities, thus suggesting lower optimal prices than those obtained from the perfect-information model. Furthermore, these prices have a lower variance than those generated by the full-information model. This difference arises because the perfect-information model overestimates preference heterogeneity, thus overestimating the scope for targeting. Overall, according to the second column in Table 6, setting personalized prices based on the search model and purchase histories on average increases profits by 8.4% relative to uniform pricing.

Finally, we derive optimal personalized prices based on the search model, but now the firm is allowed to set prices based on both purchase and search histories. Adding search data to the history of consumer decisions leads to more precise inference about consumers' types  $\theta_i$ , thus generating higher profits from personalized pricing. Specifically, our results suggest that, compared to uniform

Product	Perfect Information Model	Search Model (Purchase Data Only)	Search Model (Purchase and Search Data)
1	0.019	0.026	0.067
2	0.029	0.049	0.085
3	0.041	0.230	0.312
4	0.002	0.008	0.023
5	0.016	0.120	0.168
6	0.008	0.014	0.025
7	0.009	0.172	0.264
8	-0.072	0.069	0.136
9	0.004	0.015	0.027
10	0.054	0.143	0.209
Average	0.011	0.084	0.132

Table 6: **Expected Gains from Personalized Pricing under Different Models and Consumer Histories.** The gains from charging personalized prices are computed relative to the expected profits from optimal uniform prices. The table reports gains for three scenarios: (a) pricing based on the perfect-information model, (b) pricing based on the search model with type inference based only on purchase data, and (c) pricing based on the search model with type inference based on both search and purchase data.

pricing, the expected profits of the average firm increase by 13.2%.

In summary, in our application, firms gain little by charging personalized prices based on the perfect-information model. The average firm gains only 1.1%, and for the firm producing product 8, this incorrect targeting actually reduces profits by 7.2%. Switching to the right model and using search data significantly increases profits of all firms. Compared to uniform pricing, the average seller earns 8.4% higher profits when she estimates the search model but uses only purchase data, and 13.2% higher profits when she uses the search model and complete data on search and purchase decisions. Furthermore, as expected, the gains from personalized pricing correlate with the estimated standard deviations of intercepts in column 3 of Table 3. Products with the largest heterogeneity in the intercepts, such as products 3 and 7, gain as much as 26%-31% of profits from offering personalized prices, whereas firms selling products with little heterogeneity, such as products 4 and 6, do not have much scope for targeting and gain only 2.3%-2.5% from personalized pricing.

## 8 Conclusion

This paper studies the estimation of preference heterogeneity in a setting where consumers are imperfectly informed and have to engage in costly search. We show that in the common setting where products are horizontally differentiated, ignoring search costs leads to an overestimation of preference heterogeneity. In our empirical exercise, we find that by ignoring search costs, we overestimate standard deviations of product intercepts by 30%. This bias has important consequences

for optimal price setting. When ignoring search frictions, own-price elasticities are underestimated and markups are overestimated. Hence, when not taking search into account, we underestimate the intensity of competition between firms. In addition, personalized prices generated based on the perfect-information model have larger variance than prices based on the model of costly search. As a result, personalized prices computed under the incorrect assumption of perfect information increase expected profits by only 1.1%. Personalized prices from the search model, by contrast, increase expected profits by 13.2%. Thus, our results suggest that ignoring search costs leads to sub-optimal targeting strategies.

## References

- ACKERBERG, D. A. (2009): “A new use of importance sampling to reduce computational burden in simulation estimation,” *QME*, 7(4), 343–376.
- ALLENBY, G. M., AND P. E. ROSSI (1999): “Marketing Models of Consumer Heterogeneity,” *Journal of Econometrics*, 89(1), 57–78.
- ANDERSON, S. P., AND R. RENAULT (1999): “Pricing, product diversity, and search costs: A Bertrand-Chamberlin-Diamond model,” *The RAND Journal of Economics*, pp. 719–735.
- CHEN, Y., AND S. YAO (2016): “Sequential Search with Refinement: Model and Application with Click-stream Data,” *Management Science*, p. forthcoming.
- CHINTAGUNTA, P. K., D. C. JAIN, AND N. J. VILCASSIM (1991): “Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data,” *Journal of Marketing Research*, 28(4), 417–428.
- DE LOS SANTOS, B. I., A. HORTACSU, AND M. WILDENBEEST (2012): “Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior,” *American Economic Review*, 102(6), 2955–2980.
- DUBÉ, J.-P., G. J. HITSCH, AND P. E. ROSSI (2009): “Do Switching Costs Make Markets Less Competitive?,” *Journal of Marketing Research*, 46(4), 435–445.
- (2010): “State dependence and alternative explanations for consumer inertia,” *The RAND Journal of Economics*, 41(3), 417–445.
- DUBÉ, J.-P., G. J. HITSCH, P. E. ROSSI, AND M. A. VITORINO (2008): “Category Pricing with State-Dependent Utility,” *Marketing Science*, 27(3), 417–429.
- ELBERG, A., P. M. GARDETE, R. MACERA, AND C. NOTON (2017): “Dynamic Effects of Price Promotions: Field Evidence, Consumer Search, and Supply-Side Implications,” .
- GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN (2004): “Bayesian data analysis. Texts in statistical science series,” .
- GIULIETTI, M., M. WATERSON, AND M. WILDENBEEST (2014): “Estimation of search frictions in the British electricity market,” *The Journal of Industrial Economics*, 62(4), 555–590.
- HONKA, E. (2014): “Quantifying search and switching costs in the US auto insurance industry,” *The RAND Journal of Economics*, 45(4), 847–884.
- HONKA, E., AND P. K. CHINTAGUNTA (2015): “Simultaneous or sequential? search strategies in the us auto insurance industry,” *Marketing Science*, 36(1), 21–42.
- KIM, J. B., P. ALBUQUERQUE, AND B. J. BRONNENBERG (2010): “Online Demand under Limited Consumer Search,” *Marketing Science*, 29(6), 1001–1023.
- KOULAYEV, S. (2014): “Search for differentiated products: identification and estimation,” *The RAND Journal of Economics*, 45(3), 553–575.
- ROSSI, P. E., AND G. M. ALLENBY (1993): “A bayesian approach to estimating household parameters,” *Journal of Marketing Research*, 30(2), 171–182.

- ROSSI, P. E., R. E. MCCULLOCH, AND G. M. ALLENBY (1996): “The value of purchase history data in target marketing,” *Marketing Science*, 15(4), 321–340.
- SEILER, S. (2013): “The impact of search costs on consumer behavior: A dynamic approach,” *Quantitative Marketing and Economics*, 11(2), 155–203.
- WEITZMAN, M. L. (1979): “Optimal Search for the Best Alternative,” *Econometrica*, 47(3), 641–654.
- ZHOU, J. (2017): “Competitive bundling,” *Econometrica*, 85(1), 145–172.

## A Preference Heterogeneity and Search Costs

We have noted in section 3.2 that the way search costs affect purchase probabilities depends on whether products are horizontally or vertically differentiated, and we discussed the case of horizontal differentiation in detail. In this appendix, we provide some additional technical details on the simulation exercise and present simulation results for the case of vertical differentiation.

### A.1 Simulation Procedure

In both cases (horizontal and vertical differentiation), we assume there are three products and we simulate search and purchase behavior for a set of 100,000 consumers, whose intercepts are drawn from the distributions specified below / in section 3.2. We simulate the search and purchase behavior of each consumer for 10,000 search sessions and for a fixed value of search costs and compute purchase probabilities, as well as inferred product intercepts (under the assumption of perfect information) that rationalize purchase probabilities of each consumer.

In section 3.2, we report two graphs for the horizontal-differentiation case: the distribution of purchase probabilities and the distribution of inferred intercepts. Both graphs show kernel density estimates that we compute using a gaussian kernel and the Silverman’s rule-of-thumb bandwidth parameter. Additionally, we adjust the estimated distribution of purchase probabilities to account for the fact that probabilities are bounded between zero and one. To this end, we reflect the estimated density around two points (zero and one), compute kernel density estimates based on the augmented data, and then truncate the estimates between zero and one.

### A.2 Vertical Differentiation

To illustrate the impact of search costs in the case of vertical differentiation, we normalize the intercept of product 1 to zero, and draw intercepts for products 2 and 3 from independent normal distributions with means 1 and 2, both with a standard deviation of 0.2. Due to these distributional assumptions, most consumers agree product 3 is the best product and product 2 is second-best. As before, we compute the distributions of purchase probabilities and inferred product intercepts using simulation. Due to the differences in mean preferences across products, the distributions now differ across products and we report them for both product 2 and product 3.

The top graphs in Figure F2 show the distribution of purchase probabilities. Under the assumption of full information, modal purchase probabilities are about 0.65 for product 3 and 0.27 for product 2. An increase in search costs moves the purchase probabilities of different consumers in the same direction. As search costs increase, consumers shift their purchases from the less preferred products 1 and 2 toward the most preferred product 3. Therefore, the probability mass in the two graphs shifts in opposite directions. The figure for product 1 is left out, but looks similar to the one for product 2.

The bottom graphs in Figure F2 display the distribution of inferred intercepts for product 2 and product 3. As expected, the increase of search costs shifts the inferred distribution of product

3 intercepts to the right, as this product is purchased more frequently. The impact on the inferred distribution of intercepts for product 2 is less clear, because the search-cost increase makes product 3 purchases more likely relative to product 2. However, at the same time, product 2 becomes more likely to be purchased relative to product 1. In our specific example, the inferred distribution of product 2 intercepts shifts to the right.

In summary, ignoring search costs in the vertical-differentiation case results in biased estimates of preference parameters. The inferred distributions of intercepts are located to the right of the true distributions and have larger variances. We would therefore overestimate both means and variances of intercepts for products 2 and 3 when not taking search behavior into account. In general, the direction of the bias is less clear in the vertical case, and in the most extreme case of high search costs and purely vertical differentiation, we would find purchases concentrated around one product and would conclude that utility of all consumers is very high for that specific product.

## B Simulation Exercise with Correlated Errors

Consider an example with two products  $j = A, B$  and a continuum of consumers whose preferences are defined by (1) as before, except now we assume match values  $\varepsilon_{ijt}$  of individual products follow an AR(1) process. Specifically,  $\varepsilon_{ijt} = \rho\varepsilon_{ij,t-1} + \mu_{ijt}$ , where  $\rho \in [0, 1)$  is a correlation parameter and  $\mu_{ijt}$  are iid standard normal utility shocks. When  $\rho = 0$ , match values are uncorrelated over time, in which case, the model is equivalent to the example in section 3.2. By contrast, when  $\rho > 0$ , consumers can use match values observed in the previous sessions to predict the current values of  $\varepsilon_{ijt}$ .

To show that omitting search leads to biased estimates of heterogeneity, we generate consumers' decisions using this extended model and estimate individual-specific product intercepts from the generated data. In doing so, we assume consumers know the process followed by  $\varepsilon_{ijt}$  and update their beliefs about match values accordingly. To simplify computation, we assume that consumers make myopic decisions, not taking into account that their current search decisions will change their information set in the future. Additionally, we assume product intercepts for product A are normalized to zero for all consumers ( $\xi_{iA} = 0$ ), whereas intercepts for product B follow a standard normal distribution. Using these assumptions, we simulate behavior of  $S = 1,000$  consumers in  $T = 500$  search sessions under different values of  $\rho$ , assuming substantial search costs ( $c = 0.5$ ). We then infer product intercepts  $\xi_{iB}$  of different consumers by matching simulated purchase shares to those predicted by the perfect-information model with correlated match values.<sup>25</sup>

As we illustrate in Figure F1, ignoring search leads to overestimation of heterogeneity even when  $\rho$  is relatively large (e.g., when  $\rho = 0.9$  as in the bottom panel). The bias is more pronounced under low values of  $\rho$ , but somewhat decreases as match values  $\varepsilon_{ijt}$  become more correlated. However, even under strong correlation, the estimates of heterogeneity are substantially biased. Therefore,

---

<sup>25</sup>In our estimation, we assume the researcher knows the value of  $\rho$ . This assumption allows us to ensure any bias in estimated product intercepts arises because we assumed an incorrect model of behavior, and not because we have to estimate an additional parameter  $\rho$  from the data.

this extended example illustrates that introducing consumer learning does not affect our main qualitative conclusions.

## C Estimation Details

In section 5, we outlined the main features of the importance-sampling estimator; this section provides additional details. With regards to simulation draws, we set  $N_S = N_M = 100$ . That is, for each consumer, we take 100 draws from the proposal distribution of types  $g(\theta_i)$ . For each of these draws, we take 100 draws from the distributions of utility shocks  $\varepsilon_{it}$  and  $\mu_{it}$  for each consumer-session pair.

To compute individual likelihoods in (22), we need to first calculate reservation utilities  $z_{ijt}$  for each set of random draws  $\theta_i$  and  $\mu_{it}$ . Calculating reservation utilities is computationally burdensome, so in practice, we use approximations. Recall that  $z_{ijt}$  is defined from the indifference condition (which is equivalent to equation (2)):

$$z_{ijt} = -c_i + P(u_{ijt} \geq z_{ijt})E(u_{ijt}|u_{ijt} \geq z_{ijt}) + P(u_{ijt} < z_{ijt})z_{ijt}.$$

Under our assumption that  $\varepsilon_{ijt}$  follows a normal distribution with standard deviation  $\sigma_\varepsilon$ , this condition can be written in a more convenient form:

$$z_{ijt} = (z_{ijt} - \delta_{ijt})\Phi\left(\frac{z_{ijt} - \delta_{ijt}}{\sigma_\varepsilon}\right) + \sigma_\varepsilon \cdot \phi\left(\frac{z_{ijt} - \delta_{ijt}}{\sigma_\varepsilon}\right) + \delta_{ijt} - c_i,$$

where  $\Phi(\cdot)$  and  $\phi(\cdot)$  represent the Gaussian cdf and pdf. The right-hand side of this expression is a contraction mapping with respect to  $z_{ijt}$ , so we can compute reservation utilities through a fixed-point iteration procedure (see Elberg, Gardete, Macera, and Noton, 2017). To ease computation, we pre-compute a lookup table that stores  $z_{ijt}$  for a large number of combinations of  $\delta_{ijt}$  and  $c_i$ . We compute relevant reservation utilities from this lookup table using linear interpolation. This method is computationally inexpensive, because we only need to compute the lookup table once before the estimation. The grid for  $\delta_{ijt}$  covers values from -20 to 20 with a step-size of 0.01, and the grid for  $c_i$  is from 0 to 5 with a step-size of 0.001. Hence, our lookup table is a  $2500 \times 1000$  matrix of pre-computed reservation utilities.

## D Simulation Exercise

Our identification arguments were phrased in terms of individual-level parameters that can be identified from a long time-series of data for each consumer. However, our actual data contain only a limited panel dimension, and hence we constrain individual parameters to be drawn from specific distributions, which we define at the beginning of section 5.1. To test whether we are able to recover the structural parameters of such a model with a dataset of the size (in terms of number of consumers and search sessions per consumer) of our actual data, we implement a simulation

exercise.

We simulate the behavior of 3,500 consumers, of which 3,000 conduct only one search session and the remaining 500 conduct two search sessions. Choosing this sample size and panel structure allows us to mimic the actual dataset we use for estimation and helps us understand whether the panel dimension available in our data suffices to recover preference heterogeneity. We consider a model with  $J = 10$  products and fix standard deviations of random utility shocks at  $\sigma_\varepsilon = \sigma_\mu = 1$ . Column 1 of Table F1 reports the values of structural parameters  $\Omega$  used to generate the dataset. The prices of products are drawn from the standard normal distribution, independent across consumers, search sessions, and products.

Table F1 presents the results of this simulation exercise. Overall, coefficients are precisely estimated; standard errors are small, and the majority of estimates lie within two standard errors from the truth. We conclude that we are able to successfully recover structural parameters of the search model in the simulated sample.

## E Computation of Personalized Prices

### E.1 Computing Expected Profits

We compute expected profits as

$$\Pi(p_j, H) = \sum_{k \in \mathcal{F}} \int (p_k - mc_k) Pr(y = k | \theta) f(\theta | H, p^H, \Omega) g(\Omega) d\theta d\Omega. \quad (25)$$

This profit function captures uncertainty regarding both type  $\theta$  and structural parameters  $\Omega$ . To compute personalized prices, we first need to calculate the integrals in (25). These integrals do not have closed-form solutions and have to be approximated using simulation. We take bootstrap estimates  $\hat{\Omega}^b$  and treat them as draws from the distribution  $g(\Omega)$ .<sup>26</sup> For each of these draws  $\hat{\Omega}^b$ , we use the Metropolis-Hastings algorithm to take  $S$  draws of types  $\theta_s^b$  from the posterior distribution  $f(\theta | H, p^H, \hat{\Omega}^b)$ . The details of our sampling algorithm are described in Appendix E.2 below. Next, we use the resulting draws of types and compute the probability that the consumer buys product  $j$  for each of these draws. The expected profits in (25) can then be approximated with the following expression:

$$\Pi(p_j, H) \approx \frac{1}{BS} \sum_{k \in \mathcal{F}} \sum_{b=1}^B \sum_{s=1}^S (p_k - mc_k) Pr(y = k | \theta_s^b) \quad (26)$$

Note that  $B$  is the number of bootstrap estimates, and  $S$  is the number of types we draw for each of these bootstrap estimates.

---

<sup>26</sup>Bootstrap estimates were used earlier in order to compute standard errors. We re-use them here.

## E.2 Taking Draws $\theta_s^b$

To compute the expected profits in (26), we take draws  $\theta_s^b$  from the distribution  $f(\theta|H, p^H, \hat{\Omega}_b)$  using the standard Metropolis-Hastings algorithm. The convergence of the Markov chain is diagnosed using two different methods. First, we initiate several chains from different starting points and visually examine whether they converge to the same stationary distribution. Second, we use a Gelman-Rubin convergence statistic for the parallel chains and terminate the burnout period only when this statistic starts taking values sufficiently close to one.<sup>27</sup> Based on these two tests, we set the burnout period to be  $S_{BURN} = 1000$  draws. To diminish the effect of the starting distribution, we discard the first  $S_{BURN}$  draws and take the next  $S_{MAIN} = 10,000$  draws to construct the main sample. Because draws in this sample exhibit strong autocorrelation (the correlation between neighboring draws is around 0.9-0.95), we take only each 100th draw, thus thinning the chain and reducing autocorrelation to 0.2-0.3. The final sample consists of 100 draws of types  $\theta_s^b$  for each  $\hat{\Omega}^b$ .

---

<sup>27</sup>The Gelman-Rubin statistic reflects the ratio of between-chain variance to within-chain variance. Once all initiated chains have converged, the two variances should become similar and the Gelman-Rubin statistic should start taking values close to one (Gelman, Carlin, Stern, and Rubin, 2004).

## F Additional Tables and Figures

	Truth	Estimates	Standard Errors
$\bar{\xi}_2$	-0.200	-0.218	0.006
$\bar{\xi}_3$	-0.400	-0.385	0.007
$\bar{\xi}_4$	-0.600	-0.605	0.007
$\bar{\xi}_5$	-0.800	-0.799	0.007
$\bar{\xi}_6$	-1.000	-0.998	0.009
$\bar{\xi}_7$	-1.200	-1.208	0.008
$\bar{\xi}_8$	-1.400	-1.392	0.009
$\bar{\xi}_9$	-1.600	-1.596	0.007
$\bar{\xi}_{10}$	-1.800	-1.792	0.009
$\sigma_{\xi_2}$	0.200	0.189	0.006
$\sigma_{\xi_3}$	0.200	0.193	0.005
$\sigma_{\xi_4}$	0.200	0.190	0.004
$\sigma_{\xi_5}$	0.200	0.195	0.005
$\sigma_{\xi_6}$	0.200	0.189	0.005
$\sigma_{\xi_7}$	0.200	0.195	0.004
$\sigma_{\xi_8}$	0.200	0.199	0.006
$\sigma_{\xi_9}$	0.200	0.201	0.006
$\sigma_{\xi_{10}}$	0.200	0.193	0.006
$Mean(\alpha)$	0.056	0.060	0.001
$StdDev(\alpha)$	0.030	0.032	0.001
$Mean(c)$	0.519	0.522	0.006
$StdDev(c)$	0.159	0.158	0.004

Table F1: **Estimates from Simulation Exercise.** The table reports estimates obtained from the simulated sample (column 2) together with the true values of parameters used to generate the sample (column 1). The simulated sample contains 3,500 consumers, of which 3,000 conduct only one search session, and the remaining 500 conduct two search sessions. We set  $N_S = N_M = 100$ .

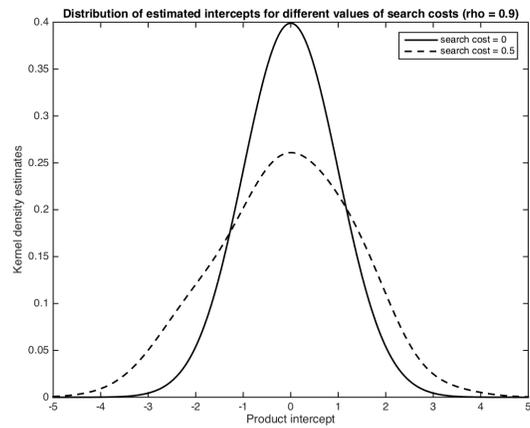
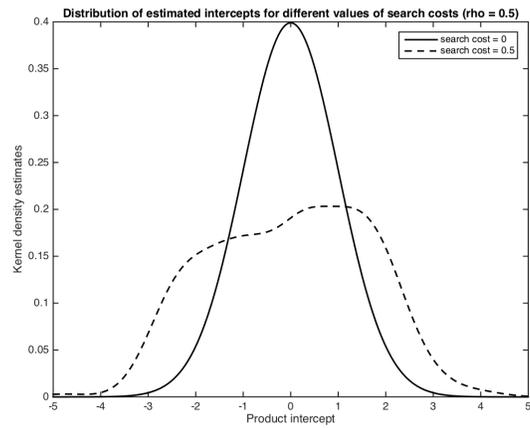
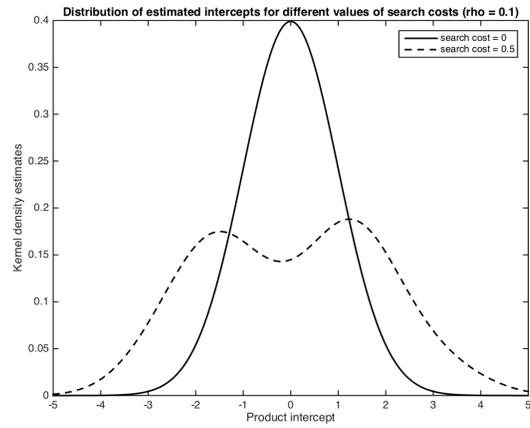


Figure F1: Distribution of Estimated Product Intercepts for Different Values of the Correlation Parameter  $\rho$ .

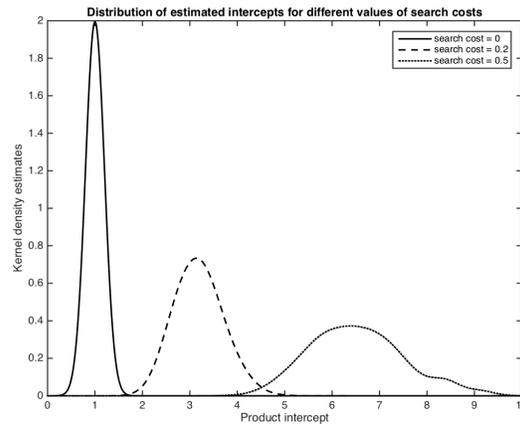
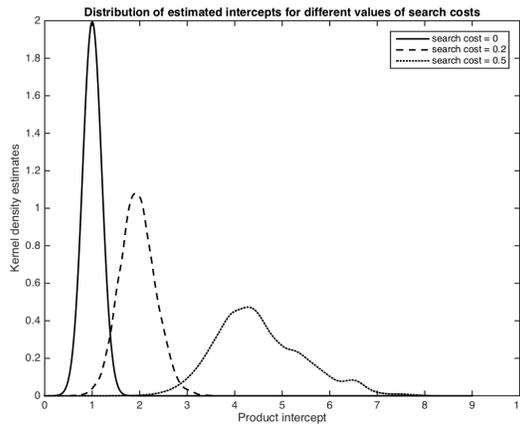
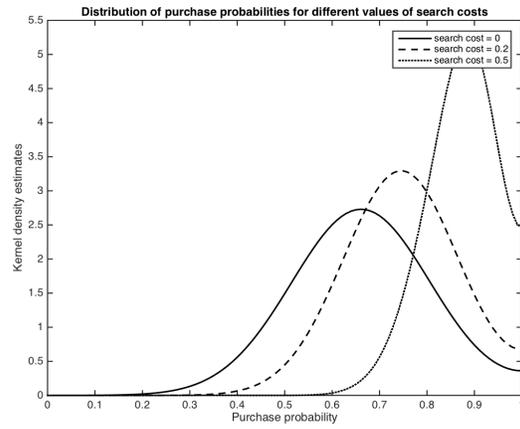
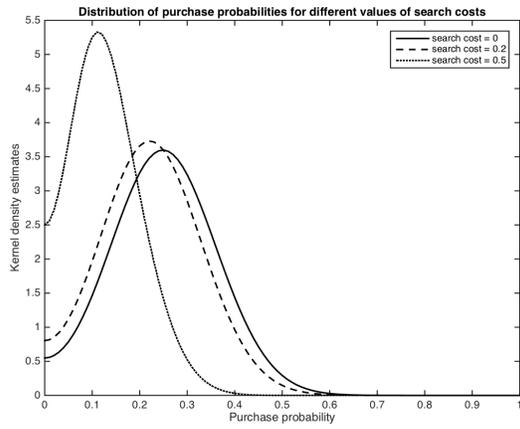


Figure F2: **Vertical Differentiation Case.** The two top graphs display kernel density estimates for the distribution of purchase probabilities under different values of search costs for product 2 (left) and 3 (right). The two bottom graphs show the corresponding kernel density estimates for the distribution of inferred product intercepts.