Marketing Science

Estimation of Preference Heterogeneity in Markets with Costly Search

Ilya Morozov , Stephan Seiler , Xiaojing Dong , Liwen Hou

http://pubsonline.informs.org/journal/mksc

Estimation of Preference Heterogeneity in Markets with Costly Search

Ilya Morozov,[a] Stephan Seiler,[b,c] Xiaojing Dong,[d] Liwen Hou[e]

[a] Northwestern University, Evanston, Illinois 60208; [b] Imperial College London, London SW7 2BU, United Kingdom; [c] Centre for Economic Policy Research, London EC1V 0DX, United Kingdom; [d] Santa Clara University, Santa Clara, California 95053; [e] Shanghai Jiao Tong University, 200240 Shanghai, China
Contact: ilya.morozov@kellogg.northwestern.edu, https://orcid.org/0000-0003-1185-4215 (IM); s.seiler@imperial.ac.uk, https://orcid.org/0000-0001-8636-1840 (SS); xdong1@scu.edu, https://orcid.org/0000-0002-6866-9443 (XD); lwhou@sjtu.edu.cn (LH)

Abstract: We study the estimation of preference heterogeneity in markets in which 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 preference heterogeneity and estimate its parameters using a unique panel data set on consumers' search and purchase behavior. The results reveal that when search costs are ignored, the model overestimates standard deviations of product intercepts by 53%. We show that the bias in heterogeneity estimates leads to incorrect inference about price elasticities and seller markups and has important consequences for personalized pricing.

1. Introduction

One of the most well-known and widely studied stylized facts in the marketing literature is that consumers exhibit strong persistence in product choices (Rossi and Allenby 1993, Rossi et al. 1996, Allenby and Rossi 1999). Researchers often interpret this persistence as evidence that consumers have heterogeneous preferences over existing products. In this paper, we challenge the conventional wisdom that strong persistence in choices is necessarily driven by strong preference heterogeneity. We argue that, in the presence of search frictions, even mild preferences over products can translate into highly persistent choices.[1]

This paper's contribution is twofold. First, we demonstrate how preference heterogeneity and search frictions jointly determine persistence in consumer choices. When search is costly, 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. In reality, however, slight differences in preferences for products are amplified by the presence of search frictions. This amplification is especially important in light of recent research that finds substantial search costs in a broad range of product categories (De Los Santos et al. 2012, Seiler 2013,

Giulietti et al. 2014, Honka 2014, Koulayev 2014). As a result, preference heterogeneity in many markets may be less pronounced than previously thought.[2]

Second, we show that accounting for search frictions is quantitatively important. To illustrate this point, we estimate a search model with flexible preference heterogeneity using search and purchase data from a large online retailer. The results reveal that ignoring search frictions leads to a 53% upward bias in the estimated preference heterogeneity, measured by the standard deviations of product intercepts. Because this bias implies a seemingly large degree of product differentiation, we underestimate absolute values of own-price elasticities by 40% and overestimate the markups of firms by 19%. Furthermore, because consumers appear more heterogeneous in our estimates than they actually are, we also overestimate the scope for personalized pricing. Optimal personalized prices are more dispersed in a perfect information model than in a model that explicitly accounts for search frictions. This difference in personalized prices affects the expected profit from personalization: although pricing based on the search model increases profits by 9.1%, pricing based on the perfect information model actually reduces profits by 0.6%. Hence, accounting for search frictions is crucial for evaluating the degree of competition among sellers and for implementing optimal personalized pricing.

We estimate our model using a unique panel data set on consumer search and purchase behavior. We observe each consumer visiting the retailer's website multiple times, and our data set describes search and purchase behavior during each of these visits. This panel dimension is particularly useful in our setting for several reasons. Because preference heterogeneity leads to persistence in choices over time, observing consumers repeatedly helps us estimate preference heterogeneity. We explicitly build our identification argument on this panel structure and show that, with a sufficiently long panel, we can nonparametrically identify how preferences and search costs are distributed across consumers. Moreover, the panel dimension of the data allows us to analyze choice persistence and study how this persistence is jointly driven by preferences and search costs.

Estimating our model is computationally burdensome because we have to integrate consumer heterogeneity out of the likelihood function, which is, by itself, difficult to compute. To reduce the computational burden, we implement an importance sampling procedure similar to that proposed by Ackerberg (2009). This approach removes the need to recompute individual likelihood contributions for each guess of parameters, thus substantially reducing computational costs. We show that this approach makes estimation significantly faster than alternative methods and can, therefore, prove useful to other researchers who want to estimate flexible models of consumer search.

This paper connects two strands of literature. On the one hand, our model of sequential search is similar to other structural models from the consumer search literature (Kim et al. 2010, De Los Santos et al. 2012, Honka 2014, Chen and Yao 2017, Honka and Chintagunta 2017). These papers, however, estimate search models using cross-sectional data and do not focus on estimating preference heterogeneity. On the other hand, there is a rich marketing literature that focuses on the estimation and identification of preference heterogeneity but assumes that consumers are perfectly informed about existing products and their characteristics (Chintagunta et al. 1991, Rossi and Allenby 1993, Rossi et al. 1996, Allenby and Rossi 1999). We connect these two areas of research by reexamining the estimation of preference heterogeneity in markets with costly search.

Our paper is also related to the broader literature on explaining persistence in consumer choices. Prior research suggests that choice persistence can be rationalized by the presence of preference heterogeneity (see previously cited papers) or switching costs (Dubé et al. 2008, 2009). We contribute to this literature by showing that search costs may also generate choice persistence. We show that, although the presence of search costs does not generate persistence by itself, search costs amplify the way preference heterogeneity translates

into persistent choices. To simplify exposition and estimation, we abstract away from switching costs. Prior research suggests that switching costs play a modest role in shaping consumer choices relative to search costs (Honka 2014) or heterogeneity in preferences (Dubé et al. 2010).[3]

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, choice persistence, and estimated preference heterogeneity. We then adapt this search model to accommodate panel data and discuss the identification of the model's parameters in Section 4. We describe our estimation strategy in Section 5. Section 6 presents the estimation results, and Section 7 analyzes implications for personalized pricing. We offer concluding remarks in Section 8.

2. General Model Framework

Suppose a consumer i conducts T_i search sessions, which we index by $t = 1, \ldots, T_i$. In each session, the consumer chooses exactly one product out of J available alternatives. The utility consumer i derives from product j in session t is the sum of a consumer-specific product intercept and a taste shock:

$$u_{ijt} = \xi_{ij} + \varepsilon_{ijt}. \tag{1}$$

We assume the consumer is imperfectly informed about product-specific utilities and must engage in costly search to resolve uncertainty. In particular, the consumer knows product intercepts ξ_{ij} and the distribution $F_\varepsilon(\varepsilon)$ from which the independent and identically distributed (i.i.d.) taste shocks ε_{ijt} are drawn, but the consumer must search to learn the realizations of these taste shocks. The taste shocks are i.i.d. across consumers, products, and search sessions.[4] There is no outside option, so the consumer always chooses one of the available options.[5]

Following Weitzman (1979), we assume the consumer searches sequentially and incurs a cost c_{ij} for each searched product. Upon finishing search, the consumer 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 (1979). Define reservation utility of consumer i for product j as the unique value z_{ij} that solves the equation[6]

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

That is, 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.[7] In the optimum, the consumer searches products in order of descending reservation utilities. At each step during this

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

	Panel A: No search costs					
	True intercepts		*Purchase probability*			*Persistence*
Product	A	B	A	B		
Consumer 1	0	−1	0.76	0.24		0.64
Consumer 2	0	1	0.24	0.76		0.64

	Panel B: Search costs = 0.5								
	True intercepts		*Search probability*		*Purchase probability*		*Persistence*	*Inferred intercepts*	
Product	A	B	A	B	A	B		A	B
Consumer 1	0	−1	1	0.12	0.91	0.09	0.84	0	−1.93
Consumer 2	0	1	0.12	1	0.09	0.91	0.84	0	1.93

	Panel C: Search costs = 1								
	True intercepts		*Search probability*		*Purchase probability*		*Persistence*	*Inferred intercepts*	
Product	A	B	A	B	A	B		A	B
Consumer 1	0	−1	1	0.03	0.97	0.03	0.94	0	−2.77
Consumer 2	0	1	0.03	1	0.03	0.97	0.94	0	2.77

Note. Persistence is defined as the probability that a consumer purchases the same product in consecutive time periods.

process, the consumer continues searching as long as the maximum realized utility is lower than the reservation utility of the next product in the search sequence; otherwise, the consumer stops searching and purchases the highest utility product among the searched options.

3. Preference Heterogeneity and Search Costs

In this section, we describe how preference heterogeneity and search costs jointly drive consumer choices. Additionally, we show that search costs increase choice persistence and demonstrate how ignoring search costs may bias the estimates of preference heterogeneity. We first build intuition using a simple example with two consumers and two products and then show that our conclusions also apply to a more general setting.

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 ($\xi_{2B} = 1$). The taste shocks ε_{ijt} are i.i.d. standard normal.

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, consumer 1 purchases product A with probability

$$PurchProb_{1A}(c = 0) = Pr(0 + \varepsilon_{iAt} \geq -1 + \varepsilon_{iBt}) = 0.76, \tag{3}$$

which corresponds to a standard probit purchase probability. We compute purchase probabilities for the other three consumer–product pairs in a similar way and report the results in Panel A of Table 1. According to the computed values, each consumer purchases the product with the higher intercept with probability 76%. The probability of purchasing the same product in consecutive time periods is then equal to 64%.

Next, we consider a second scenario in which consumers engage in costly search and face a search cost of $c = 0.5$. In contrast to their behavior in the perfect information scenario, consumers now actively decide which products they want to search and in which order they want to search them. For example, according to the optimal search rule, consumer 1 first searches product A and only searches product B if the realized taste shock of product A is below product B's reservation utility.[8] Hence, this consumer only searches product B with probability

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

where z_{1B} is the reservation utility from Equation (2).[9] If the consumer searches product B, the consumer then decides which product to buy based on the realized utilities of the two products. By contrast, if the consumer does not search product B, the consumer never learns the realization of the shock ε_{iBt} and

purchases product A. Combining these two observations, we obtain that the consumer purchases product A with probability

$$PurchProb_{1A}(c = 0.5) = Pr(0 + \varepsilon_{iAt} \geq z_{1B})$$
$$+ Pr(0 + \varepsilon_{iAt} < z_{1B}) \cdot Pr(0 + \varepsilon_{iAt} \geq -1$$
$$+ \varepsilon_{iBt}|0 + \varepsilon_{iAt} < z_{1B})$$
$$= 0.91.$$

The first term captures the case in which the consumer only searches product A, whereas the second term captures 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 costly search, each consumer purchases the highest intercept product with probability 91% and purchases the other product only with probability 9%.

Comparing purchase probabilities under the 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, which we illustrate in Panel C of Table 1 in which we increase search costs to $c = 1.0$. Intuitively, higher search costs make consumers less likely to search the products with lower intercepts, thus reducing the purchase probabilities for these products. This shift in purchase probabilities toward high-intercept products also increases choice persistence, that is, the probability that the consumer chooses the same product in consecutive time periods. Although this probability equals 0.64 in the perfect information scenario, it increases to 0.84 when the search cost equals $c = 0.5$ and further increases to 0.94 when the search cost equals $c = 1.0$.

With costly search, consumers make more extreme choices. To see this 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. However, these probabilities become 0.91 and 0.09 when search costs equal 0.5 and shift to 0.97 and 0.03 when search costs increase to 1.0. Hence, costly search increases the dispersion of purchase probabilities for product A across consumers. By symmetry, we see the same increase in the dispersion of purchase probabilities for 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, thus generating an upward bias in 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 the equation $Pr(0 + \varepsilon_{iAt} > \hat{\xi}_{1B} + \varepsilon_{iBt}) = 0.91.$[10] We compute the intercept of consumer 2, $\hat{\xi}_{2B}$, 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, which are more dispersed than the true values −1 and 1. Raising search costs to one exacerbates this bias, generating the inferred intercepts −2.77 and 2.77. As a whole, these results show that ignoring search costs leads to an overestimation of preference heterogeneity and that the magnitude of the bias is increasing in consumers' search costs.

3.2. Product-Specific Search Costs

To this point, we have assumed identical search costs across products. One might think, however, that search costs may vary across products. For example, an online retailer may make some products more salient than others (e.g., via advertising or product rankings), thus lowering these products' search costs. To illustrate the implications of product-specific search costs, we repeat the exercise from the previous section but assume that search costs are asymmetric across products. In particular, we assume a search cost of 0.5 for product A and 2.0 for product B. Given those search cost values, both consumers find it attractive to search product A first. Therefore, the search order of consumer 2 no longer aligns with consumer 2's preferences and is driven by the difference in product-specific search costs. As illustrated in Table 2, both consumers start by searching product A and are more likely to buy it than product B. As a result, when costly search is ignored, we mistakenly conclude that these consumers' preferences are similar, which leads to an underestimation of preference heterogeneity.

As this specific example shows, introducing asymmetric search costs may reverse the conclusions from Section 3.1 and may lead to an underestimation (rather than overestimation) of preference heterogeneity. We, therefore, explicitly model product-specific search costs in our application (see Section 4). Even when allowing for possible asymmetries in search costs, however, we find that preference heterogeneity is significantly overestimated when ignoring search frictions.

3.3. Many Products and Continuous Distribution of Heterogeneity

The examples we have discussed thus far are somewhat stylized in that they feature only two consumers and two products. We now present a more general example with a structure that is more closely related to the empirical model we use for estimation. This example mimics the preference structure of random-coefficient demand models that have been used extensively in prior literature to estimate preference heterogeneity. We use the same utility specification as before (see

Table 2. Purchase Behavior When Search Costs Are Asymmetric Across Products

	Panel A: No search costs					
	True intercepts			Purchase probability		
Product	A	B		A	B	
Consumer 1	0	−1		0.76	0.24	
Consumer 2	0	1		0.24	0.76	

|	Panel B: Search cost A = 0.5/search cost B = 2							
	True intercepts		Search probability		Purchase probability		Inferred intercepts	
Product	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	1	0.16	0.84	0.16	0	−1.41

Equation (1)) but consider a market with three products and assume that the consumer-specific intercepts for each of these products follow a standard normal distribution. We simulate search behavior for a large number of consumers and for a large number of search sessions for each consumer.[11]

In the left-hand graph in Figure 1, we plot the distribution of resulting purchase probabilities across consumers for one of the products.[12] 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 by increasing the mass of consumers with low and high purchase probabilities and decreasing the mass of consumers with intermediate purchase probabilities (dashed line). As we further increase search costs, this pattern becomes more pronounced (dotted line). As in our previous example, the increase in search costs also increases choice persistence. The average choice persistence increases from 0.44 in the perfect information case to 0.65 in the case in which search costs are 0.5.

Next, we infer product intercepts assuming perfect information. The right-hand graph in Figure 1 shows the density of inferred intercepts for one of the products under different search costs. 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. Thus, as in our previous example with two consumers and two products, ignoring search costs leads to an overestimation of preference heterogeneity.

We also consider a similar example in the case of a market that is characterized by strong vertical differentiation of products. To introduce vertical differentiation, we assume that the intercepts follow a normal distribution with an identical standard deviation of 0.2 and with means zero, one, and two. That is, most consumers now agree regarding which products look ex ante more attractive, but they differ slightly in the exact values of their product-specific intercepts. We

Figure 1. Distribution of Purchase Probabilities and Estimated Product Intercepts for Different Values of Search Costs
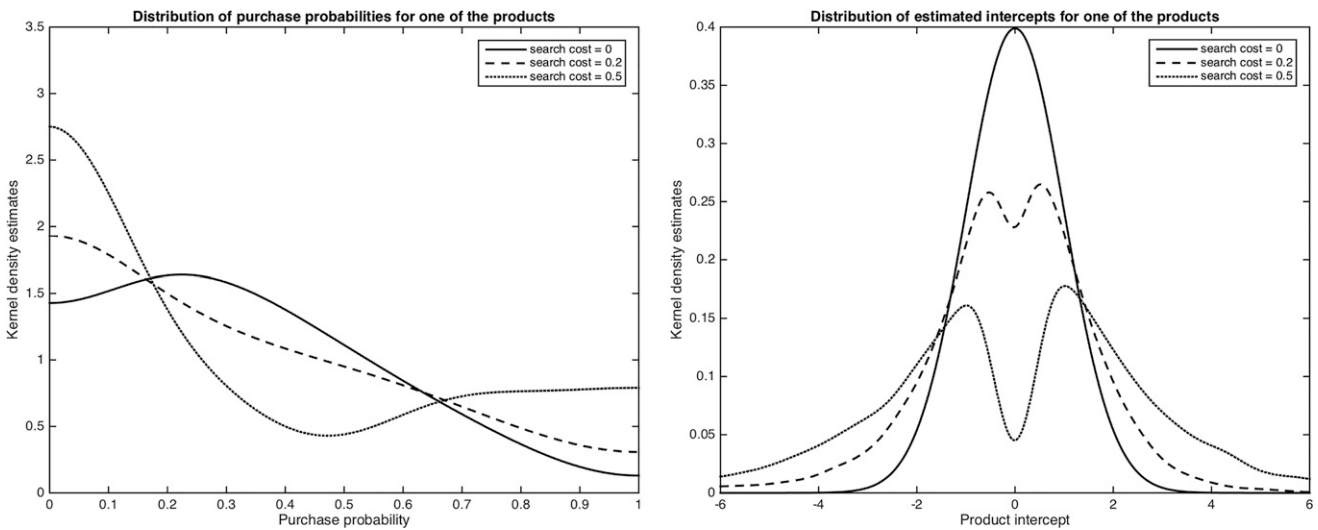
Note. The two graphs display kernel density estimates for the distribution of purchase probabilities and inferred product intercepts.

Figure 2. Vertical Differentiation Case
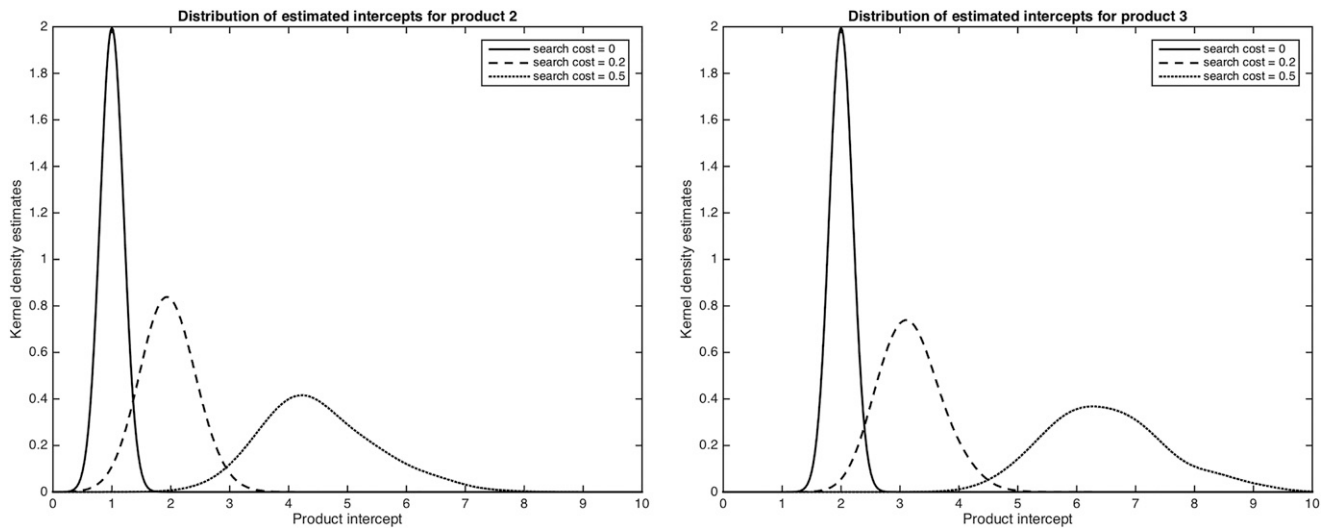
Note. The two graphs display kernel density estimates for the distribution of inferred product intercepts for products 2 (left) and 3 (right).

report the distributions of inferred intercepts for products 2 and 3 in Figure 2. For both products, the variance of inferred intercepts increases when search costs are higher. Moreover, the distributions are shifted to the right under costly search relative to the perfect information case. Hence, in the vertical case, ignoring search leads to a bias in both the variance of inferred intercepts and mean intercepts. Although our focus is on the estimates of preference heterogeneity, that is, the variance of intercepts, we later also document a bias in inferred mean intercepts in our empirical application. Average persistence in choices increases from 0.51 in the perfect information case to 0.78 in the case in which search costs are equal to 0.5.

3.4. General Case: Search Costs Increase Persistence and Inferred Preference Heterogeneity

The preceding examples illustrate that, when search costs increase, consumer choices tend to become more persistent. As a result, consumers appear to have strong preferences for certain items, which may lead to an overestimation of preference heterogeneity when the researcher ignores costly search. In this section and in Appendices A and B, we show that higher search costs lead to higher choice persistence and inferred preference heterogeneity under a wide range of assumptions about model primitives. We emphasize that this relationship between search costs, choice persistence, and inferred preference heterogeneity only applies to a setting in which search costs are common across products. As the example outlined in Section 3.2 shows, our findings do not apply to product-specific changes in search costs; therefore, it is an empirical question whether search cost asymmetries are large

enough such that preference heterogeneity is underestimated. In our application, we find that, even when allowing for product-specific search costs, preference heterogeneity is significantly overestimated.

Suppose a consumer chooses from J products indexed by j and that the indirect utility is $u_j = \xi_j + \varepsilon_j$, where ξ_j is product j's intercept observed prior to search and ε_j is the taste shock revealed during search. Assume ε_j is independently and identically distributed across products and sessions according to some continuous distribution $F(\varepsilon)$. In this setting, we establish two results.

1. In the sequential search model with uniform search costs across products, increasing search costs by $\Delta > 0$ leads to an increase in choice persistence for any value of Δ.

2. In the sequential search model with uniform search costs across products, increasing search costs by $\Delta > 0$ leads to an increase in the inferred heterogeneity of product intercepts for any value of Δ.

Intuitively, because search is directed, the consumer always starts by searching products with high intercepts. As search costs increase, the consumer increasingly focuses on searching a small set of products with the largest intercepts, gradually converging to the state in which the consumer only searches and purchases the product with the largest intercept ξ_j. As a result, the consumer's choices become increasingly concentrated around a few products, which increases the probability of purchasing the same product in consecutive search sessions. In Appendix A, we provide a formal proof of this result, which holds for all continuous distributions of taste shocks $F(\varepsilon)$. The result also applies to markets in which consumers search over prices, in which case we can interpret taste shocks ε_j

as a consumer's preference over prices and $F(\varepsilon)$ as a price distribution consistent with the consumer's beliefs. We also show in the same appendix that this result holds for the simultaneous search model of Chade and Smith (2006).

Furthermore, when search costs are positive, the bias in the inferred intercept of a given product, $(\hat{\xi}_j - \xi_j)$, increases in the true intercept's value ξ_j. It, therefore, follows that, for any pair of consumers (denoted as consumers 1 and 2) with $\xi_{1j} > \xi_{2j}$, inferred intercepts are further apart from each other than true intercepts so that $\hat{\xi}_{1j} - \hat{\xi}_{2j} > \xi_{1j} - \xi_{2j}$. Therefore, search costs increase the dispersion in inferred intercepts across consumers. Furthermore, the bias in inferred intercepts is negative (positive) if the true intercept value is negative (positive). Therefore, mean intercepts are biased away from zero. Finally, the absolute value of the bias increases in search costs, thus leading to a stronger increase in dispersion if search costs are higher. In Appendix B, we use an extensive set of simulations to show that this result holds for the utility specification outlined and under various assumptions regarding the distribution of taste shocks ε_j.[13]

4. Empirical Model and Identification

To further illustrate that ignoring search costs leads to biased heterogeneity estimates, we consider an empirical application. Our general strategy in the empirical application is to use a data set most suitable for recovering a flexible distribution of preference heterogeneity. To this end, we use a unique panel data set in which each consumer visits the retailer's website multiple times and in which we observe search and purchase behavior during each consumer visit.

To estimate the model's parameters from such data, we first extend the sequential search model from Section 2, making it compatible with panel data. The extended model adds price to the utility function and includes an additional taste shock revealed prior to search. We then derive restrictions that observed choices impose on model parameters and realizations of taste shocks and finish by discussing the identification of model parameters.

4.1. Panel Data Framework

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

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

In this expression, ξ_{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;[14]

α_i denotes the price coefficient; and μ_{ijt} and ε_{ijt} are idiosyncratic taste shocks, both distributed normally and i.i.d. with zero mean and variances σ_ε^2 and σ_μ^2. Prior to searching, a consumer knows the realizations of δ_{ijt} and the distribution of ε_{ijt} but has to search in order to learn the realization of ε_{ijt}.[15] The consumer pays a fixed cost c_{ij} when the consumer searches product j. We assume that search costs may differ across products, reflecting that some products might be ranked higher on the web page and, hence, are more salient to consumers (Ursu 2018). As before, the model does not include an outside option.

The key difference relative to the framework in Section 2 is the presearch taste shock μ_{ijt}. Absent this taste shock, an individual consumer who encounters the same product prices in each search session would always search products in the same order. Introducing the shock μ_{ijt} resolves this issue by adding a source of randomness in presearch utilities. One can interpret the taste shock μ_{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 postsearch taste shock ε_{ijt} represents any information retrieved from the product page itself, such as information from customer reviews and detailed product descriptions.

In our main specification, we assume shocks μ_{ijt} and ε_{ijt} are independent across search sessions. This assumption implies that new information is available in each search session. We later relax this assumption in Section 6.4 in which we reestimate the model, allowing for information to be correlated across search sessions. We find that the estimates of preference heterogeneity change little compared with those from our main specification.

We normalize the variances of taste shocks at $\sigma_\mu^2 = 1$ and $\sigma_\varepsilon^2 = 1$. The first normalization is necessary to fix the scale of utility. We could, in principle, relax the second normalization and estimate σ_ε^2 from the data, but it is generally hard to estimate this variance separately from search costs. Increasing the variance σ_ε^2 leads to an increase in the benefits from search, which makes consumers search more, thus producing an effect similar to that of reducing search costs. We address this issue by normalizing the variance σ_ε^2. Experimenting with different values of σ_ε^2 reveals that this second normalization has little impact on our qualitative results. In particular, when we change the normalized standard deviation σ_ε by a factor of k, estimated search costs change roughly by a factor $1/k$. Because search costs and the variance of postsearch shocks are not fully collinear, other parameter estimates also change, albeit only marginally. Most importantly, in our application, we find that the monetized values of preference

heterogeneity estimates, the primary object of interest in this paper, remain almost unchanged when altering the value of σ_ε^2.[16] In Appendix C, we discuss in more detail why these two variance normalizations are necessary for the estimation.

4.2. Restrictions Imposed by Observed Choices
The search model imposes a set of restrictions on the model's parameters and utility shocks. In this section, we derive three sets of restrictions associated with different decisions: in which order to search products, when to stop searching, and what to purchase.

4.2.1. Search Order.
According to the optimal search rule, the consumer searches products in order of decreasing reservation utilities (see Section 2). Let $\mathfrak{I} = \{1,\dots, J\}$ denote the set of available products, and let $S_{it} \subseteq \mathfrak{I}$ be the set of products that consumer i searched during search session t. Also let M_{it} be the number of products in this set so that $|S_{it}| = M_{it}$. Without loss of generality, throughout this section, we index products in the order in which they were searched by the consumer in a given search session, that is, product $j = 1$ is searched first, $j = 2$ is searched second, and so on.

When the consumer searches products S_{it}, the ranking of reservation utilities z_{ijt} must be consistent with the observed search order:

$$z_{1it} \geq z_{2it} \geq \dots \geq z_{M_{it}, it} \geq z_{kit} \text{ for } \forall\, k \notin S_{it}. \quad (5)$$

Because we only observe search order for products that were actually searched, the data does not impose any restrictions on the ranking of reservation utilities for unsearched products. We only know that the reservation utility of the product searched last, $z_{M_{it}, it}$, exceeds the reservation utilities of all unsearched products. A more compact way to write the inequalities in (5) is

$$z_{mit} \geq \max_{l \in S_{mit}} z_{lit} \text{ for } m = 1,\dots, M_{it}, \quad (6)$$

where $S_{mit} = \mathfrak{I} \setminus \{j : j \leq m\}$ represents the set of products left unsearched after searching options $1,\dots, m$. These inequalities specify the restrictions that the observed order of search imposes on parameters and realizations of taste shocks.

4.2.2. 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 m, it must be that the maximum realized utility among searched products $1,\dots, (m-1)$ is lower than the reservation utility of product m, namely z_{mit}. This relationship must hold for all searched products

except the first one because, in our model, the consumer always searches at least one product:

$$\max\{u_{1it},\dots, u_{(m-1)it}\} \leq z_{mit} \text{ for } m = 2,\dots, M_{it}. \quad (7)$$

If the consumer decides to stop after searching product M_{it}, the maximum realized utility of searched products must exceed the reservation utilities of all unsearched products:[17]

$$\max\{u_{1it},\dots, u_{M_{it}, it}\} \geq \max_{k \notin S_{it}} z_{kit}. \quad (8)$$

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

4.2.3. 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_{it} and buys a product $y_{it} \in S_{it}$, this product's realized utility should be the largest among all searched products:

$$u_{y_{it}, it} \geq \max_{j \in S_{it}} u_{jit}. \quad (9)$$

4.3. Identification
We build our identification argument around the panel dimension of the data set. In particular, we discuss how individual-level preference parameters $(\xi_{i1},\dots, \xi_{iJ}, \alpha_i)$ and search costs $(c_{i1},\dots, c_{iJ})$ can be identified for one individual consumer when the researcher observes the probability distribution of search and purchase decisions for that consumer. Given that these individual-level parameters are identified, it then follows that the distribution of these parameters across individuals is nonparametrically identified. Because our argument requires us to observe the probability distribution of searches and purchases *for each consumer*, the argument effectively assumes that the data include an infinitely large number of search sessions per consumer. Note, however, that the data in our empirical application only contain a moderate panel dimension. Therefore, instead of estimating individual-level parameters for each consumer, we estimate the parameters of the distribution from which preferences and search costs are drawn.

To the best of our knowledge, when only cross-sectional data are available, the prior literature provides no formal identification argument for the nonparametric identification of consumer heterogeneity. Several papers argue that, when only cross-sectional data are available, the researcher needs to rely on functional form assumptions to identify unobserved preference heterogeneity. For example, Honka (2014) argues that preference heterogeneity can be identified through

deviations of demand patterns from the independence of irrelevant alternatives property, which follows from the assumption that error terms follow the extreme value distribution in her model. Honka and Chintagunta (2017) and Yavorsky et al. (2021) only allow for observed heterogeneity and argue that unobserved heterogeneity in preference is generally not identified with cross-sectional data.[18] Unlike these papers that do not focus explicitly on estimating unobserved heterogeneity, the primary goal of our paper is to study the effect of search frictions on estimates of preference heterogeneity. Therefore, it is particularly important in our setting to achieve nonparametric identification through the use of panel data.

4.3.1. Inferring Reservation Utilities.

To simplify exposition, we first assume that price does not enter the utility function and focus on the identification of product intercepts and search costs. During search session t, consumer i searches products in order of decreasing reservation utilities z_{ijt} as defined in Equation (2). Following Kim et al. (2010), we decompose the reservation utility as $z_{ijt} = \bar{z}_{ij} + \mu_{ijt} = \xi_{ij} + g(c_{ij}) + \mu_{ijt}$, where $g(c_{ij})$ is a known function that monotonically decreases in the search cost and depends only on the search cost c_{ij} and the distribution of preference shocks ε_{ijt}.[19] Going forward, we refer to the deterministic part of the reservation utility $\bar{z}_{ij}$ as *mean reservation utility*. The probability that consumer i first searches product k can be expressed as

$$Pr(i \text{ searches product } k \text{ first})$$
$$= Pr(z_{kit} \geq z_{jit} \quad \forall j) = Pr(\bar{z}_{ik} + \mu_{ikt} \geq \bar{z}_{ij} + \mu_{ijt} \quad \forall j). \quad (10)$$

Because we assume preference shocks μ_{ijt} follow the standard normal distribution, the expression in (9) corresponds to the standard probit probability. However, in contrast to standard discrete choice models in which probit expressions describe purchase probabilities, the expressions here capture the probabilities that different products are searched first within a search session.[20] In a setting with panel data, we can observe first-search probabilities for each consumer i and each product j. We emphasize that, without panel data, we would only observe one search session per consumer; in that case, we would not be able to compute these consumer-specific probabilities. Following Hotz and Miller (1993), we invert the observed first-search probabilities $Pr(i \text{ searches product } 1 \text{ first}), \ldots, Pr(i \text{ searches product } J \text{ first})$ to infer the values of $\bar{z}_{ij}$. Hence, for each consumer i, we infer the $\bar{z}_{ij}$ term for all products.

The identities of other searched products provide additional information about model parameters. Based on the inequalities in (5), consumer i chooses to search products in a specific order with probability

$$Pr(i \text{ searches in order } 1, \ldots, M_{it})$$
$$= Pr(z_{mit} \geq \max_{l \in S_{mit}} z_{lit} \text{ for } m = 1, \ldots, M_{it}). \quad (11)$$

This expression suggests that products that are often searched early during the search process should have higher mean reservation utilities $\bar{z}_{ij}$. In addition, the mean reservation utility should be low for products that are searched at the end of the search session or not searched at all. This additional information about the complete search order, not just about first searches, helps us estimate preference parameters.

4.3.2. Separating Preferences from Search Costs.

Next, we decompose the mean reservation utilities into preferences and search costs. The probability that consumer i in session t stops after searching M_{it} products is given by

$$Pr(i \text{ stops after searching } M_{it} \text{ products})$$
$$= Pr(\max\{u_{1it}, \ldots, u_{M_{it}, it}\} \geq \max_{k \notin S_{it}} z_{kit}). \quad (12)$$

Note that the stopping decision depends on the realized utilities of products that have been searched and reservation utilities of products that have not yet been searched. Intuitively, search costs for products that have already been searched are not relevant with regard to the stopping decision, and only realized utilities of searched products matter. The irrelevance of search costs of already searched products allows us to separately identify preferences and search costs. In particular, suppose we increase search costs for one specific product c_{ij} but, at the same time, increase this product's intercept ξ_{ij} such that $\bar{z}_{ij}$ remains unchanged.[21] Because $\bar{z}_{ij}$ is unchanged, the expected search order remains unchanged as well. However, the realized utility of product j is now higher because of its higher intercept; therefore, according to Equation (11), the consumer is more likely to stop searching after having searched this specific product. Hence, conditional on $\bar{z}_{ij}$, stopping decisions of consumers conditional on searching a specific product allow us to identify the product's intercept. Finally, having identified $\bar{z}_{ij}$ and ξ_{ij}, we can find the value of c_{ij} such that the relationship $\bar{z}_{ij} = \xi_{ij} + g(c_{ij})$ holds. This logic can be applied repeatedly for different i and j, implying that we can identify product intercepts and product-specific search costs for all consumers and products.

We have so far ignored the price coefficient and discussed identification of product intercepts; however, identification of the price coefficient is straightforward. The extent to which a consumer searches cheaper products earlier in a search session identifies the consumer's price coefficient α_i 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.

In summary, our identification argument establishes that, with panel data, we can identify the product intercepts ξ_{ij}, price coefficient α_i, and search costs c_{ij} for each individual consumer. Because this logic can be applied to each consumer, we are able to nonparametrically identify the distribution of preferences and search costs across consumers.

4.3.3. The Role of Purchase Data. The arguments presented thus far suggest that search order and stopping decisions are sufficient to identify preferences and search costs for a given consumer. We note, however, that consumers' purchases conditional on search decisions provide additional variation to estimate preferences separately from search costs. The corresponding argument is somewhat involved, so we relegate a more detailed discussion to Appendix D. The key pattern we outline there is that increasing a product's intercept makes it more likely that this product is purchased, conditional on being searched. This observation holds as long as the product in question was not searched last within the search session. In other words, increasing the intercept increases the product's recall probability, defined as the probability of purchasing a product searched earlier in the search sequence. At the same time, changing a product's search cost does not change the product's recall probability. Therefore, product-specific recall rates provide additional information to recover preferences separately from search costs.

5. Estimation
5.1. Parameterization
Our identification argument assumes that the data include an infinite number of search sessions per consumer. In our actual data set, however, the panel dimension is much longer for some consumers than others as the number of sessions per consumer ranges from 2 to 35. To handle data for consumers with a limited panel dimension, we impose additional assumptions on the distribution of consumer heterogeneity that allow us to pool information across consumers. Therefore, unlike the earlier nonparametric argument in which panel data identify consumer-specific search costs and preference parameters, panel data in our empirical application help us estimate the hyperparameters governing the distribution of consumer heterogeneity.

First, we assume that product intercepts ξ_{ij} for each product j follow a normal distribution:

$$\xi_{ij} \sim N(\overline{\xi}_j, \sigma^2_{\xi_j}), \qquad (13)$$

where $\overline{\xi}_j$ is the mean intercept for product j and $\sigma^2_{\xi_j}$ denotes the variance of the product intercept that captures unobserved preference heterogeneity. Second, we assume that price coefficients α_i also vary across consumers, capturing the differences in the marginal utility of income. To ensure price coefficients are positive for all consumers, we assume that α_i follows a log-normal distribution such that

$$\log \alpha_i \sim N(\bar{\alpha}, \sigma^2_{\alpha}), \qquad (14)$$

where $\bar{\alpha}$ and σ^2_{α} denote the mean and the variance. Similarly, we assume that the search cost c_{ij} for product j follows a log-normal distribution with a product-specific mean $\overline{c}_j$ and common variance σ^2_c:

$$\log c_{ij} \sim N(\overline{c}_j, \sigma^2_c). \qquad (15)$$

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 as predictors for consumers' preferences and search costs. This situation is relatively common in online markets in which firms tend to have rich information on consumers' purchase histories but relatively little information on their demographics. 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 et al. 1996, Allenby and Rossi 1999). Hence, modeling observed heterogeneity may not be crucial for explaining the observed behavior of consumers.

We also note that, in several papers, search costs have been parameterized as a function of the time at which the search occurs (Koulayev 2014, Chen and Yao 2017) or as a function of the product's ranking on the website (Ursu 2018). Unfortunately, we do not have access to any characteristics of this kind, and hence, we do not model search costs as a function of observable consumer characteristics.

5.2. Likelihood Function
To derive the likelihood function, we first summarize the restrictions that observed choices impose on parameters and taste shocks. Let vector $\boldsymbol{\theta_i} = (\xi_{i1}, \dots, \xi_{iJ}, \alpha_i, c_{i1}, \dots, c_{iJ})$ denote the consumer's type that describes the consumer's product intercepts, price sensitivity, and product-specific search costs. Additionally, let $\boldsymbol{p}_{it}$ denote a $J \times 1$ price vector and let $\boldsymbol{\mu}_{it}$ and $\boldsymbol{\varepsilon}_{it}$ denote $J \times 1$ vectors of taste shocks. If a consumer with type $\boldsymbol{\theta}_i$ searches products S_{it} in order $\{1, \dots, M_{it}\}$ and purchases product $y_{it} \in S_{it}$, the following inequalities must hold:

$$w^O_{mit}(\boldsymbol{\theta_i}, \boldsymbol{\mu_{it}}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it}) = z_{mit} - \max_{l \in S_{mit}} z_{lit} \geq 0$$

$$\text{for } m = 1, \dots, M_{it}, \qquad (16)$$

$$w_{mit}^S(\boldsymbol{\theta}_i, \boldsymbol{\mu}_{it}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it}) = z_{mit} - \max\{u_{1it}, \ldots, u_{(m-1)it}\} \geq 0$$
$$\text{for } m = 2, \ldots, M_{it}, \quad (17)$$

$$w_{(M_{it}+1)it}^S(\boldsymbol{\theta}_i, \boldsymbol{\mu}_{it}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it})$$
$$= \max\{u_{1it}, \ldots, u_{M_{it},it}\} - \max_{k \notin S_{it}} z_{kit} \geq 0, \quad (18)$$

$$w_{it}^P(\boldsymbol{\theta}_i, \boldsymbol{\mu}_{it}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it}) = u_{y_{it}, it} - \max_{j \in S_{it}} u_{jit} \geq 0, \quad (19)$$

where (15) corresponds to the order of search, (16) and (17) describe inequalities for continuation and stopping decisions, and (18) captures purchase inequalities. To simplify notation, we define function $W_{it}(\boldsymbol{\theta}_i, \boldsymbol{\mu}_{it}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it})$ as

$$W_{it}(\boldsymbol{\theta}_i, \boldsymbol{\mu}_{it}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it}) =$$
$$\min\{w_{1it}^O, \ldots, w_{M_{it},it}^O, w_{2it}^S, \ldots, w_{(M_{it}+1)it}^S, w_{it}^P\}. \quad (20)$$

so that $W_{it}(\boldsymbol{\theta}_i, \boldsymbol{\mu}_{it}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it}) \geq 0$ if and only if all the inequalities in (15)–(18) hold.

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. Let D_{it} denote these data for consumer i and session t, and let D denote the data for all consumers and sessions. Suppose the data D includes N consumers, and each consumer i conducts T_i search sessions. The likelihood of observing search and purchase decisions of consumer i in session t given consumer's type $\boldsymbol{\theta}_i$ and prices $\boldsymbol{p}_{it}$ is

$$L_{it}(D_{it}|\boldsymbol{\theta}_i, \boldsymbol{p}_{it}) = Pr(W_{it}(\boldsymbol{\theta}_i, \boldsymbol{\mu}_{it}, \boldsymbol{\varepsilon}_{it}, \boldsymbol{p}_{it}) \geq 0|\boldsymbol{\theta}_i, \boldsymbol{p}_{it}), \quad (21)$$

where the uncertainty on the right-hand side comes from taste shocks $\boldsymbol{\mu}_{it}$ and $\boldsymbol{\varepsilon}_{it}$, both of which are unobserved to a researcher.

Combining individual likelihoods, we can write the full likelihood of the data D as

$$L(D|\boldsymbol{\Omega}, \boldsymbol{p}) = \prod_{i=1}^N \int \left(\prod_{t=1}^{T_i} L_{it}(D_{it}|\boldsymbol{\theta}_i, \boldsymbol{p}_{it}) \right) \cdot dF(\boldsymbol{\theta}_i|\boldsymbol{\Omega}), \quad (22)$$

where $\boldsymbol{\Omega} = (\bar{\alpha}, \sigma_\alpha, \bar{c}, \sigma_c, \bar{\xi}_1, \ldots, \bar{\xi}_j, \sigma_{\xi_1}, \ldots, \sigma_{\xi_j})$ denotes the vector of hyperparameters that determine how types $\boldsymbol{\theta}_i$ are distributed across consumers. Because types $\boldsymbol{\theta}_i$ are unobserved to the researcher, the likelihood function in (21) contains an expectation of consumer-specific likelihoods with respect to the joint distribution of $\boldsymbol{\theta}_i$. This joint distribution, $F(\boldsymbol{\theta}_i|\boldsymbol{\Omega})$, is fully defined by the distributional assumptions in (12)–(14).

5.3. Importance Sampling

A straightforward approach to estimation would be to maximize the simulated version of the likelihood in (21) with respect to parameters $\boldsymbol{\Omega}$. In our application, this approach is impractical because we need to recompute the simulated likelihood for each new guess of

parameters. This repeated computation is computationally burdensome because we need 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 Ackerberg (2009). We rewrite the likelihood function by multiplying and dividing individual likelihoods by some density function $g(\boldsymbol{\theta}_i)$, which we term proposal density:

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

Importantly, we choose the function $g(\boldsymbol{\theta}_i)$ such that it does not depend on hyperparameters $\boldsymbol{\Omega}$. This observation suggests that, instead of drawing consumers' types from the original density $f(\boldsymbol{\theta}_i|\boldsymbol{\Omega})$, we can take N_M draws of types from the proposal density $g(\boldsymbol{\theta}_i)$ and approximate the likelihood function with a simulated counterpart:

$$\tilde{L}(D|\boldsymbol{\Omega}, \boldsymbol{p}) = \prod_{i=1}^N \frac{1}{N_M} \sum_{m=1}^{N_M} \left(\prod_{t=1}^{T_i} \tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \boldsymbol{p}_{it}) \cdot \frac{f(\boldsymbol{\theta}_i^m|\boldsymbol{\Omega})}{g(\boldsymbol{\theta}_i^m)} \right). \quad (23)$$

Here, $\boldsymbol{\theta}_i^m$ denotes the mth draw of types for consumer i, and $\tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \boldsymbol{p}_{it})$ is the simulated consumer and session-specific likelihood contribution:

$$\tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \boldsymbol{p}_{it}) = \frac{1}{N_S} \sum_{s=1}^{N_S} 1\{W_{it}(\boldsymbol{\theta}_i^m, \boldsymbol{\mu}_{it}^s, \boldsymbol{\varepsilon}_{it}^s, \boldsymbol{p}_{it}) \geq 0\}. \quad (24)$$

In this expression, $\boldsymbol{\mu}_{it}^s$ and $\boldsymbol{\varepsilon}_{it}^s$ are draws of taste shocks. We can then estimate parameters $\boldsymbol{\Omega}$ by maximizing the simulated expression $\tilde{L}(D|\boldsymbol{\Omega}, \boldsymbol{p})$ in (22). Appendix E describes the details of this estimation procedure.

The importance-sampling approach has several key advantages over a frequency estimator. First, this approach significantly reduces the computational burden of estimation. The only part of the likelihood in (22) that depends on parameters $\boldsymbol{\Omega}$ is the weights $f(\boldsymbol{\theta}_i^m|\boldsymbol{\Omega})/g(\boldsymbol{\theta}_i^m)$, which implies that we do not need to recompute the likelihood contributions $\tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \boldsymbol{p}_{it})$ for each new guess of parameters $\boldsymbol{\Omega}$. Instead, we can precompute these likelihood contributions for the generated draws of types $\boldsymbol{\theta}_i^m$ and taste shocks $\boldsymbol{\mu}_{it}$ and $\boldsymbol{\varepsilon}_{it}$ and use these 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 $N_S = 1{,}000$ and $N_M = 1{,}000$, which results in at least 1,000,000 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 $f(\boldsymbol{\theta}_i^m|\boldsymbol{\Omega})/g(\boldsymbol{\theta}_i^m)$ in (22) are continuous and differentiable in $\boldsymbol{\Omega}$. This

smoothness allows us to use gradient-based optimization methods instead of slower derivative-free alternatives.[22]

One practical consideration is the choice of the proposal density $g(\boldsymbol{\theta}_i)$. Ideally, we would choose the proposal density that coincides with the true density of types $\boldsymbol{\theta}_i$, $f(\boldsymbol{\theta}_i|\boldsymbol{\Omega}_0)$. Since we do not know $\boldsymbol{\Omega}_0$, in practice, we choose $g(\boldsymbol{\theta}_i) = f(\boldsymbol{\theta}_i| \tilde{\boldsymbol{\Omega}}_0)$, where $\tilde{\boldsymbol{\Omega}}_0$ are the initial values of parameters used as a starting point for our estimation. To obtain reasonable initial values, we choose $\tilde{\boldsymbol{\Omega}}_0$ that roughly matches four types of moments to their data counterparts: purchase probabilities, probabilities of first searches, persistence in purchases, and the expected number of searched products per session.

6. Data and Results
6.1. Data and Descriptive Statistics
We use data from a large international chain of cosmetics stores. The data contain search and purchase activity of consumers in the chain's Chinese online store. The layout of the store's website is similar to that of other online retailers. Figure 3 shows a typical layout of an online store that looks similar to that of the store we analyze. Consumers can enter search terms or use category tags or filters to narrow down the set of relevant products. They are then presented with a product list displaying basic information, such as product names, prices, photos, and short descriptions. Consumers can click on any product to visit the product page, which contains a more detailed product description as well as customer reviews.

Our data set includes detailed information on browsing and purchase activity of all website users in 2014. We observe the time stamp of each page visit, types of visited pages (e.g., main page, product page, shopping cart), and identities of products described on each visited page. We also observe purchases made through the website. In addition, the data include unique consumer identifiers, which we use to track consumers over time and match their browsing activity to related purchases.[23]

We focus on the category of cosmetic face masks. This category is suitable for our analysis for several reasons. Because consumers purchase face masks regularly, we can observe many of them making repeat purchases. Observing repeated purchases and corresponding search sessions is crucial for our empirical strategy because the panel dimension of the data helps us identify preference heterogeneity (see our argument in Section 4.3). Another advantage of choosing face masks is that most of this category's search and purchase activity is concentrated around a small number of products. To simplify estimation, we focus on the 10 most popular moisturizers, which account for more than 50% of category sales.

We consider a consumer to be searching a given product if the consumer visits the corresponding product page.[24] We define each search session retrospectively from the purchase in which it ended. Specifically, the search session includes all searches that occurred within one week before a given purchase.[25] Because it is difficult to define search sessions for consumers without purchases, we exclude these consumers from the data. For computational reasons, we only include consumers who made at least two purchases. The final estimation sample includes 11,540 search sessions conducted by 2,303 consumers. The average consumer in our data conducted five search sessions (i.e., made five purchases). Some consumers made as many as 35 purchases during the observation period.

Table 3 describes the main patterns of consumer behavior and presents some evidence of preference heterogeneity. Turning to search behavior (see Panel A of Table 3), we note that consumer search is limited. Consumers search, on average, only 2.42 products in a search session. Out of all search sessions, 17.9% finish after one search, 41.2% end after two searches, and only the remaining 40.9% end after three or more searches.

Next, we investigate persistence in searches and purchases across different search sessions of the same consumer (Panel B of Table 3). Consumers tend to make persistent decisions: 73% of search sessions end in the same purchase as the previous search session. Moreover, search behavior exhibits a similar persistence as 88% of the first searches are products that were also searched first during the previous session. The persistence in both purchases and searches suggests that consumers have substantially different preferences over the available products.

Finally, Panel C of Table 3 presents descriptive statistics at the product level. Prices vary substantially across products and range from 60.2 to 488.1 yuan (US$8.6 and US$69.8). Market shares are also significantly different, indicating some degree of vertical differentiation. We also report the frequency of consumers' repeat purchases conditional on purchasing specific products in the past. 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 probabilities for the most popular product (30% market share) and the least popular product (2.2% market share) are remarkably similar: 0.76 versus 0.70.

6.2. Estimation Results
Table 4 presents estimation results from our search model. The estimated parameters include means and standard deviations of product intercepts ξ_{ij} for nine products (the intercept for the first product is normalized to zero) as well as parameters that determine the

Figure 3. (Color online) Examples of a Category Page and a Product Page
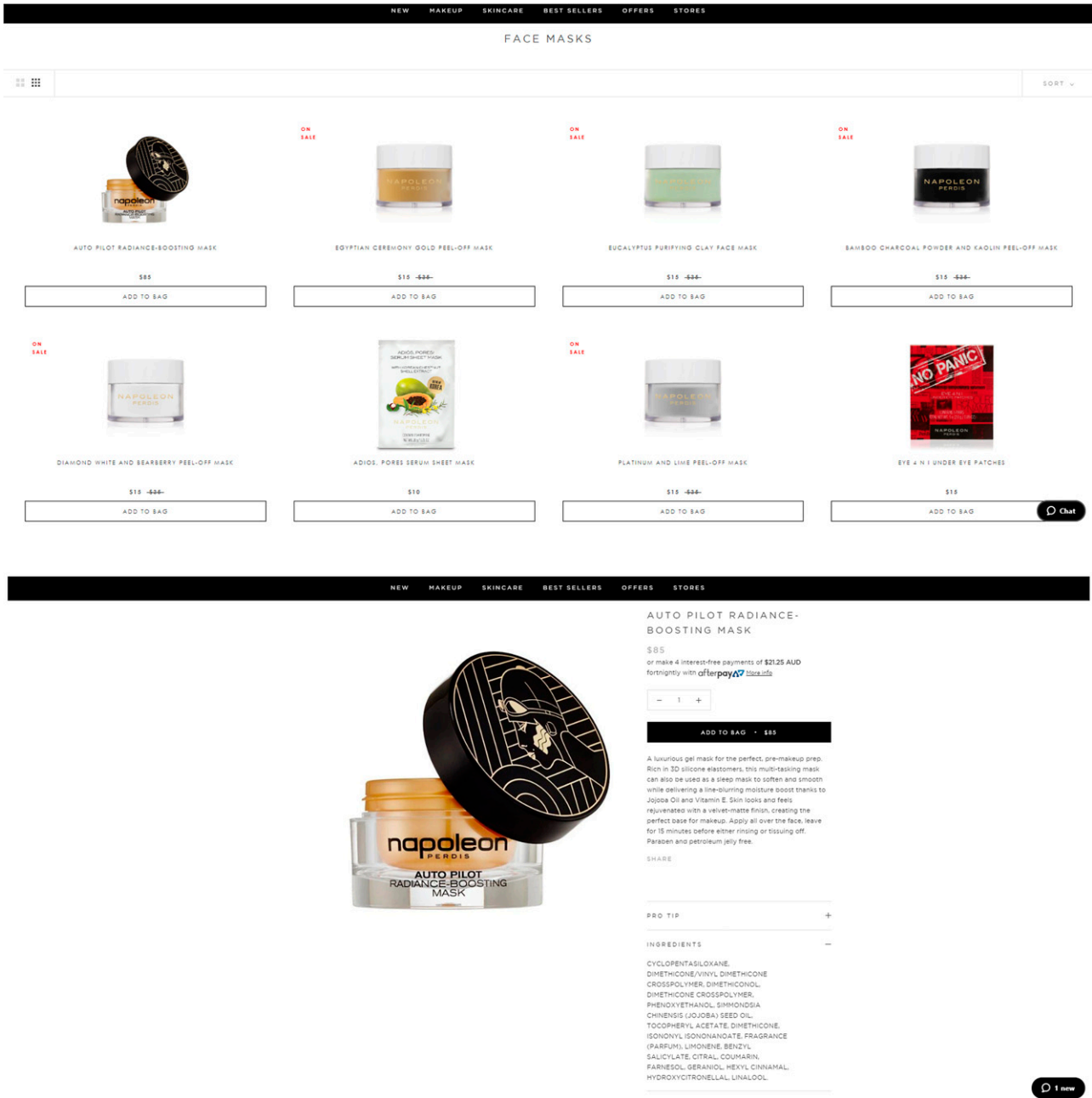
Notes. The two figures show examples of a category page (top figure) and a product page (bottom figure). To preserve the anonymity of our data provider, we show examples of webpages from another store that looks similar.

distribution of the price coefficients α_i and search costs c_{ij}. To ease interpretation, we report the implied means and standard deviations of price coefficients and search costs rather than the location and variance parameters of the corresponding log-normal distributions. Note that we cannot monetize estimated search costs because they are identified only in relation to the normalized variance σ_ξ^2 (see the discussion in Section 5.1 and in Appendix C).[26] We also present estimates from the perfect information model in the last three columns of Table 4. This model

assumes zero search costs for all consumers and products and is estimated using only purchase data, but it is otherwise identical to our search model.

Our analysis 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. A natural way to compare estimates is to monetize them in both models.[27] We report the monetized values of product intercepts and

Table 3. Descriptive Statistics for the Category of Cosmetic Face Masks

Panel A: Search behavior		
Average number of searches per session		2.42
Number of searches per session	1	17.93%
	2	41.18%
	3	27.32%
	≥ 4	13.57%

Panel B: Persistence	
Purchase in t = Purchase in $t-1$	0.73
First search in t = First search in $t-1$	0.88

		Panel C: Product level		
Product	Brand	Average price (yuan)	Purchase share	Search share	Repeat purchase probability
1	Sephora	63.0	0.297	0.237	0.762
2	Dr. Jart	400.0	0.103	0.100	0.718
3	For Beloved One	76.9	0.092	0.104	0.743
4	Benefit	373.2	0.091	0.101	0.715
5	Caudalie	420.7	0.040	0.037	0.696
6	Bliss	488.1	0.101	0.110	0.724
7	Boscia	60.2	0.085	0.104	0.721
8	Bliss	158.0	0.048	0.058	0.712
9	Sephora	230.0	0.022	0.024	0.702
10	Biotherm	60.2	0.116	0.121	0.711

standard deviations in columns (3) and (6) of Table 4. According to the results, the perfect information model uniformly overestimates the variances of product intercepts. For the average product, the standard deviation is overestimated by 53%. Furthermore, we find that estimated mean intercepts are almost uniformly biased away from zero.

With regard to search cost estimates, we find that estimated search costs differ somewhat across products. As discussed in Section 3.2, strong asymmetries in search costs across products could potentially lead to an underestimation of preference heterogeneity, whereas relatively similar search costs typically imply that heterogeneity is overestimated when we ignore search. Our empirical application falls into the latter category: we find that omitting costly search from the model leads to an overestimation of preference heterogeneity for all products. To further explore the role of asymmetric search costs, we also estimate a restricted model with common search costs across products. We report the results from this specification in Table 5 next to the results from our baseline model with product-specific search costs. We find that forcing search costs to be identical across products still leads to an overestimation of preference heterogeneity relative to the perfect information model.

6.3. Out-of-Sample Fit
We examine model fit by comparing several key predictions to their empirical counterparts. To this end, we split the sample of 2,303 consumers into an 80% training sample of 1,842 consumers and a 20% holdout

sample of 461 consumers. Table 6 reports the simulated predictions and their empirical counterparts for the holdout sample. We first evaluate fit in terms of purchase and first search probabilities in Panels I and II and find that the model predicts the hold-out data reasonably well. Similarly, it predicts the average number of searches reasonably well. The model does slightly worse in terms of fitting persistence in searches and purchases: both measures of persistence are underestimated. We note that the model contains no parameters that specifically cater to matching persistence, for example, switching costs. Hence, the predicted degree of persistence is entirely driven by preference heterogeneity and search costs.

We also report measures of in-sample statistical fit at the bottom of Table 4. However, these measures for the models with and without search are not directly comparable because we are not merely fitting different models to the same data. We are also using more data and a different likelihood function in the model with costly search. In particular, we are maximizing the joint-likelihood of search and purchase decisions in the search model and the likelihood of purchase decisions in the model without search.

6.4. Extension: Persistent Information Across Search Sessions
In our baseline model, we assume that taste shocks ε_{ijt} are independent across search sessions. Therefore, any information a consumer obtains about a product in a given search session does not predict utility from the

Table 4. Estimation Results from the Search Model and the Perfect Information Model

	Search model			Perfect information model		
	Coefficients	Standard errors	Monetized parameters	Coefficients	Standard errors	Monetized parameters
$\bar{\xi}_2$	−4.38	1.52	−154.0	−4.76	0.17	−199.6
$\bar{\xi}_3$	−5.42	1.08	−190.6	−5.45	0.10	−228.8
$\bar{\xi}_4$	−9.54	1.89	−335.6	−6.47	0.15	−271.4
$\bar{\xi}_5$	−8.12	2.25	−285.6	−8.57	0.27	−359.9
$\bar{\xi}_6$	−7.01	0.75	−246.7	−5.87	0.17	−246.8
$\bar{\xi}_7$	−6.25	1.58	−219.8	−6.81	0.10	−286.1
$\bar{\xi}_8$	−7.84	0.95	−275.8	−8.65	0.21	−363.2
$\bar{\xi}_9$	−7.24	0.84	−254.3	−10.43	0.20	−437.6
$\bar{\xi}_{10}$	−4.74	0.67	−166.9	−4.58	0.07	−192.6
σ_{ξ_2}	4.56	0.99	162.5	5.07	0.09	224.2
σ_{ξ_3}	4.68	1.44	166.8	5.07	0.11	215.2
σ_{ξ_4}	3.20	1.49	119.1	5.67	0.09	241.1
σ_{ξ_5}	4.08	1.04	147.5	5.39	0.10	230.2
σ_{ξ_6}	5.15	1.96	184.0	5.47	0.08	232.1
σ_{ξ_7}	4.34	0.88	155.6	5.55	0.07	235.8
σ_{ξ_8}	4.57	1.55	164.3	6.02	0.13	256.5
σ_{ξ_9}	4.72	1.31	169.4	6.87	0.10	293.4
$\sigma_{\xi_{10}}$	3.67	0.75	131.0	4.34	0.07	184.1
$Mean(\alpha)$	0.0288	0.0087		0.0240	0.0014	
$StdDev(\alpha)$	0.0031	0.0015		0.0024	0.0009	
$Mean(c_1)$	0.0064	0.0044				
$Mean(c_2)$	0.0021	0.0015				
$Mean(c_3)$	0.0010	0.0008				
$Mean(c_4)$	0.0005	0.0002				
$Mean(c_5)$	0.0066	0.0021				
$Mean(c_6)$	0.0012	0.0017				
$Mean(c_7)$	0.0015	0.0004				
$Mean(c_8)$	0.0010	0.0031				
$Mean(c_9)$	0.0093	0.0065				
$Mean(c_{10})$	0.0019	0.0001				
$StdDev(c)$	0.0003	0.0002				
Log-Likelihood	−37,623.5			−22,237.1		
Akaike information criterion	75,739.0			44,514.7		
Bayesian information criterion	75,865.3			44,629.1		

Notes. The unit of observation is a specific consumer's search session. The sample contains 2,303 consumers and 11,540 search sessions. Monetized intercepts are reported in yuan net of price effects. The intercepts for product 1 are normalized for all consumers. Fit statistics are not directly comparable for the two models because their estimation is based on different data sets: the model in columns (1)–(3) uses search data that is not used in the perfect information model in columns (4)–(6).

same product in future sessions. However, it is possible that consumers retain some information from one session to another. To allow for this possibility, we now consider an extension in which we allow for partial correlation in information, which nests our baseline model as a special case when the correlation is set to zero. There are two main reasons why information might be partially correlated over time. First, new information may become available by the time a consumer returns for another search session. This new information can take the form of updated product page content that features new customer ratings and reviews, additional photos, or changes in the product description. Second, partial correlation may arise if consumers remember some product features that they

have learned in previous sessions but forget others, thus partially retaining information acquired during prior search.

To model partial correlation of taste shocks across search sessions, we assume that shocks ε_{ijt} follow an MA(1) process so that

$$\varepsilon_{ijt} = \omega_{ijt} + \rho\omega_{ijt-1},$$

where ω_{ijt} follows a standard normal distribution and is i.i.d. across consumers, products, and search sessions. Upon searching product j in period t, the consumer i observes both the taste shock ε_{ijt} and the transitory shock ω_{ijt}. As before, we assume that consumers have rational expectations consistent with the true distribution of taste shocks. These assumptions imply that a consumer's

Table 5. Alternative Specifications: Common Search Costs Across Products and Persistence in Information

	Main specification			Common search costs across products			Persistence in information		
	Coefficient	Standard error	Monetary parameter	Coefficient	Standard error	Monetary parameter	Coefficient	Standard error	Monetary parameter
$\bar{\xi}_2$	−4.38	1.52	−154.0	−4.48	1.56	−128.3	−3.14	1.33	−117.0
$\bar{\xi}_3$	−5.42	1.08	−190.6	−4.68	0.74	−134.3	−5.69	0.94	−211.5
$\bar{\xi}_4$	−9.54	1.89	−335.6	−9.68	1.95	−277.4	−9.47	4.13	−351.7
$\bar{\xi}_5$	−8.12	2.25	−285.6	−7.95	1.88	−227.9	−7.83	3.12	−291.0
$\bar{\xi}_6$	−7.01	0.75	−246.7	−6.76	0.69	−193.9	−6.35	0.06	−236.0
$\bar{\xi}_7$	−6.25	1.58	−219.8	−6.63	2.22	−190.2	−6.11	2.57	−227.3
$\bar{\xi}_8$	−7.84	0.95	−275.8	−7.47	1.53	−214.2	−6.97	1.47	−259.1
$\bar{\xi}_9$	−7.24	0.84	−254.3	−8.52	0.76	−244.1	−8.90	0.46	−330.8
$\bar{\xi}_{10}$	−4.74	0.67	−166.9	−4.28	0.91	−122.8	−4.58	0.08	−170.4
σ_{ξ_2}	4.56	0.99	162.5	4.97	0.90	144.0	3.73	0.79	138.7
σ_{ξ_3}	4.68	1.44	166.8	3.91	1.26	113.6	3.79	1.49	140.9
σ_{ξ_4}	3.20	1.49	119.1	3.17	1.57	95.1	3.10	1.06	115.2
σ_{ξ_5}	4.08	1.04	147.5	4.08	1.34	119.5	3.67	1.28	136.6
σ_{ξ_6}	5.15	1.96	184.0	5.32	2.53	154.4	4.65	3.51	173.2
σ_{ξ_7}	4.34	0.88	155.6	4.71	0.79	136.8	4.49	1.25	166.8
σ_{ξ_8}	4.57	1.55	164.3	4.45	1.57	129.9	4.89	1.47	181.7
σ_{ξ_9}	4.72	1.31	169.4	5.03	1.39	146.9	4.81	1.65	178.8
$\sigma_{\xi_{10}}$	3.67	0.75	131.0	3.43	0.69	99.4	3.70	0.86	137.3
$Mean(\alpha)$	0.0288	0.0087		0.0352	0.0065		0.0269	0.0109	
$StdDev(\alpha)$	0.0031	0.0015		0.0034	0.0009		0.0020	0.0062	
$Mean(c_1)$	0.0064	0.0044							
$Mean(c_2)$	0.0021	0.0015							
$Mean(c_3)$	0.0010	0.0008							
$Mean(c_4)$	0.0005	0.0002							
$Mean(c_5)$	0.0066	0.0021							
$Mean(c_6)$	0.0012	0.0017							
$Mean(c_7)$	0.0015	0.0004							
$Mean(c_8)$	0.0010	0.0031							
$Mean(c_9)$	0.0093	0.0065							
$Mean(c_{10})$	0.0019	0.0001							
$Mean(c)$				0.0032	0.0020		0.0028	0.0060	
$StdDev(c)$	0.0003	0.0002		0.0003	0.0001		0.0003	0.0001	
ρ							0.3		
Log-Likelihood	−37,623.5			−37,950.9			−37,750.2		
Akaike information criterion	75,739.0			76,281.7			76,023.3		
Bayesian information criterion	75,865.3			76,291.4			76,112.0		

Notes. The unit of observation is a search session of a specific consumer. The sample contains 2,303 consumers and 11,540 search sessions. Monetized intercepts are reported in yuan net of price effects. The intercepts for product 1 are normalized for all consumers.

beliefs at the beginning of search session t can be summarized as follows:[28]

$$\varepsilon_{ijt} \sim N(0, \sigma_\omega^2 \cdot (1 + \rho^2))$$
if product j was NOT searched in $t-1$
$$N(\rho\omega_{ijt-1}, \sigma_\omega^2) \text{ if product } j \text{ was searched in } t-1.$$

Hence, the information obtained when searching a product in a given search session partly carries over to the next search session.[29] The extended model leads to different postsearch taste shock variances across products. However, within a search session, taste shocks are still independently distributed across products; therefore, the optimal search rule can be

characterized using Weitzman's threshold rule described earlier. In Appendix F, we rewrite the likelihood function to accommodate correlated taste shocks and discuss estimation details for this extended model.

The new parameter ρ introduces correlation in taste shocks across sessions, thus potentially increasing persistence in consumers' search and purchase behavior. Because this effect is similar to that of increasing preference heterogeneity, one may ask whether it is possible to identify the correlation parameter ρ separately from the preference heterogeneity parameters σ_{ξ_j}. To see why these parameters can be separately identified, note that the correlation ρ introduces "structural" state

Table 6. Out-of-Sample Fit of the Search Model

	I. Purchase probabilities			II. Search probabilities (based on first searches)	
Product	Data	Predicted	Product	Data	Predicted
1	0.295	0.301	1	0.303	0.232
2	0.099	0.107	2	0.115	0.104
3	0.081	0.104	3	0.095	0.111
4	0.086	0.023	4	0.088	0.031
5	0.057	0.040	5	0.063	0.033
6	0.098	0.083	6	0.099	0.103
7	0.093	0.105	7	0.095	0.119
8	0.051	0.077	8	0.060	0.093
9	0.026	0.047	9	0.033	0.048
10	0.114	0.114	10	0.049	0.126

	III. Persistence		IV. Number of searches	
	Data	Predicted	Data	Predicted
Purchases	0.734	0.666	2.47	2.16
First searches	0.840	0.798		

Notes. For the out-of-sample validation exercise, we split our sample of 2,303 consumers into two parts, allocating 80% of them to the training sample (1,842 consumers) and 20% to the holdout sample (461 consumers). We then estimate the model on the training sample and use the estimated model to predict several key moments. This table compares these predicted moments to their counterparts observed in the holdout sample.

dependence, that is, persistence in choices that is not driven by time-invariant differences in preferences and instead driven by a causal impact of past searches on current decisions. Therefore, if a temporary price discount makes a consumer more likely to search and purchase a given product, this induced choice generates a causal effect on decisions made in the next session. This pattern stands in contrast to a model without correlation ($\rho = 0$) in which decisions in the previous search session have no direct impact on future decisions.

This identification argument is similar in spirit to the arguments developed in the literature on switching costs (Dubé et al. 2010). The main difference from their argument, however, is that in our model, the structural dependence in choices manifests itself in both search and purchase decisions. If a consumer searched and purchased a product in the past, the consumer likely received a favorable draw of shock ε_{ijt} for the product and should be more likely to search and purchase this product in the current session. Instead, if a consumer searched but did not purchase a product, the consumer likely received an unfavorable draw ε_{ijt} and should, therefore, be less likely to search and purchase this product in the following session. Hence, we can identify the correlation ρ from the degree to which consumers' search and purchase decisions are affected by the products they searched and purchased in the previous search session. The results from this extension are reported in the rightmost columns of Table 5. We find estimated

correlation to be relatively low with $\hat{\rho} = 0.3$.[30] Most importantly, the estimates of preference heterogeneity do not significantly change when we allow for correlation in taste shocks. Therefore, although this extension provides a more flexible modeling framework, allowing for correlated information does not change the main conclusions regarding the estimates of preferences heterogeneity.

7. Personalized Pricing
7.1. Price Elasticities and Markups
The bias in heterogeneity estimates may affect our inference about price elasticities and the 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.[31] As a result, we may underestimate demand elasticities and incorrectly conclude that competition among firms is mild.

We illustrate this point in Table 7 in which we report price elasticities from the search and perfect information models. Consistent with the preceding logic, the perfect information model underestimates the absolute values of own-price elasticities for all 10 products. This bias in elasticities, in turn, affects our inference about markups of firms. Assuming Bertrand–Nash pricing, we obtain that the perfect information model implies an average markup of 245.0 yuan, whereas the search model predicts an average markup of only 205.8 yuan. Therefore, by ignoring search, we overestimate markups of firms by 19%. Overall, our results suggest that, to correctly understand the competitive environment, one needs to account for search frictions and the way they affect firms' pricing strategies. If a researcher incorrectly inferred from high choice persistence that products are highly differentiated, she would underestimate how strongly firms compete with each other.

7.2. Personalized Prices
We now consider the implications of our results for personalized marketing. In general, a firm's benefit from personalizing its marketing activities critically depends on its ability to assess the heterogeneity of consumer preferences. If preference heterogeneity is overestimated, a firm overestimate the scope for personalization and engage in suboptimal targeting. To illustrate this point, we consider the example of personalized pricing in which, for a given product, a firm sets different prices based on consumers' search and purchase histories.

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 that follow omit index i to avoid clutter). Suppose the firm observes consumer i's history of searches and

Table 7. Estimated Own-Price Elasticities

Product	Own-price elasticities	
	Search model	Perfect information model
1	−0.511	−0.492
2	−3.156	−2.536
3	−0.737	−0.581
4	−5.862	−2.383
5	−4.816	−3.175
6	−3.996	−3.018
7	−0.673	−0.444
8	−1.825	−1.109
9	−2.359	−1.519
10	−0.724	−0.502
Average	−2.506	−1.576

Notes. The table compares own-price elasticities predicted by two models: the search model and the perfect information model. All elasticities are computed at the current prices observed in the data.

purchases. The history consists of all products searched in previous sessions, the order in which these products were searched, and purchase decisions. Let H denote this history, and let p^H denote the vector of prices the consumer faced in these previous search sessions, in which T is the number of search sessions the consumer conducted in the past. The firm infers consumer type θ 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 θ given history H follows Bayes' rule:[32]

$$f(\theta|H, p^H, \Omega_0) = \frac{L(H|p^H, \theta)f(\theta|\Omega_0)}{\int L(H|p^H, \theta)dF(\theta|\Omega_0)}, \qquad (25)$$

where $L(H|p^H, \theta)$ is the likelihood of observing history H given historical prices p^H and type θ and Ω_0 is a vector of the true parameter values. Given this posterior distribution of types, $f(\theta|H, p^H, \Omega_0)$, the firm chooses the price p_j for product j to maximize its expected total profit

$$\Pi(p_j, H) = \sum_{k\in\mathcal{F}} \int (p_k - mc_k)Pr(y=k|\theta)$$
$$f(\theta|H, p^H, \Omega)g(\Omega)d\theta d\Omega. \qquad (26)$$

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 the type θ, which is captured by the posterior density $f(\theta|H, p^H, \Omega)$, and about hyperparameters Ω, which is reflected in the density $g(\Omega)$. We approximate posterior beliefs about type θ by taking draws from the posterior distribution $f(\theta|H, p^H, \Omega)$. Similarly, we approximate the distribution $g(\Omega)$ using the asymptotic distribution

of the maximum likelihood estimates $\hat{\Omega}$ (see Appendix H for details). In addition, we infer marginal costs mc_j by assuming that 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 profits from different pricing regimes, we use the expected profit function from the estimated search model. By construction, this profit function is maximized when the seller sets personalized prices based on the search model and infers individual preferences from complete histories of past searches and purchases. Here, we derive optimal personalized prices separately for each product, assuming the firm sets prices for product j while holding fixed the (uniform) prices of other products. This simplifying assumption allows us to abstract away from potential equilibrium effects and helps us focus on assessing the scope for targeting.

We first compute personalized prices based on the perfect information model. Specifically, we use estimates from the perfect information model and infer types θ only from purchase histories $H = \{y_t\}_{t=1}^T$. Figure 4 shows the resulting distribution of personalized prices for one product in our sample.[33] To quantify losses from setting prices based on the model that ignores search, we insert obtained personalized prices into the profit function from the search model. We report the resulting change in profits in Table 8. The results show that targeting based on the model that ignores search reduces expected profits on average by 0.6% relative to uniform pricing.

Next, we analyze personalized prices derived from the search model. To isolate the effect of switching to a

Figure 4. Distribution of Personalized Prices
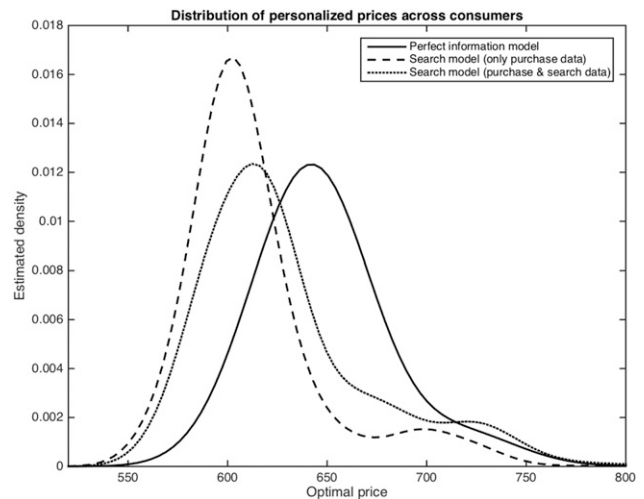
Note. The graph plots the kernel densities of personalized prices under three scenarios: (a) perfect information model (solid line), (b) search model in which types are inferred based only on purchase data (dashed line), and (c) search model in which types are inferred from both search and purchase data (dotted line).

Table 8. Expected Gains from Personalized Pricing Under Different Models and Consumer Histories

	Profit increase from personalization, %		
Product	Perfect information model	Search model (purchase data only)	Search model (purchase and search data)
1	−3.66	3.38	7.24
2	1.31	9.88	15.30
3	0.54	2.23	6.28
4	−4.28	0.25	11.45
5	−1.52	7.40	10.90
6	1.43	2.30	9.10
7	−0.42	0.65	5.90
8	−0.57	5.74	11.91
9	1.06	4.20	9.80
10	0.29	0.64	2.77
Average	−0.58	3.67	9.07

Notes. 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.

different model of consumer behavior, we only allow the firm to set prices based on purchase histories as in the previous exercise (i.e., the firm is not allowed to price-discriminate based on search histories). The results in Figure 4 suggest the distribution of personalized prices has a shape that is similar to that 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 higher absolute values of own-price elasticities, thus suggesting that optimal prices should be set lower than those obtained from the perfect information model. Furthermore, these prices have a lower variance than those generated by the perfect information model. This difference arises because the perfect information model overestimates preference heterogeneity, thus overestimating the scope for targeting. Overall, the results in the second column of Table 8 suggest that setting personalized prices based on the search model and only purchase histories increases profits for the average product by 3.7% relative to uniform pricing.

Finally, we derive optimal personalized prices based on the search model but now allow the firm to set prices based on both purchase and search histories. Adding search data to consumers' histories leads to more precise inference about types θ, thus generating higher profits from personalized pricing. Specifically, our results suggest that, compared with uniform pricing, the expected profits for the average product increase by 9.1%.

In summary, firms gain little by charging personalized prices based on the perfect information model. In fact, setting prices based on this model reduces profits from the average product by 0.6%. Some products lose even more. For example, profits from product 4 decrease by 4.3% from personalization based on the perfect information model. By contrast, firms' profits significantly increase when they explicitly account for costly search and use search data to infer consumers' types. Compared with uniform pricing, profits for the average product increase by 3.7% when firms set prices based on the search model but only use purchase histories for targeting. Profits increase by 9.1% when firms rely on the search model and personalize based on both search and purchase histories.

8. Conclusion

In this paper, we study the estimation of preference heterogeneity in markets in which consumers are imperfectly informed and engage in costly search. We show that, in the typical search model, 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 53%. This bias has important consequences for optimal price setting. When ignoring search frictions, we underestimate own-price elasticities and overestimate firms' markups. Hence, when search is omitted from the analysis, we underestimate the intensity of competition between firms. In addition, personalized prices generated based on a perfect information model are more dispersed than prices computed based on the estimated search model. As a result, personalized prices computed under the incorrect assumption of perfect information reduce expected profits by 0.6%. Personalizing prices based on the search model, by contrast, increases expected profits on average by 9.1%. Thus, ignoring search costs may lead to suboptimal personalized pricing.

Appendix A. Formal Proof: Search Costs Increase Persistence

In this section, we show that choice persistence increases with search costs. We define persistence as the probability of a given consumer purchasing the same product in two consecutive search sessions. To illustrate the generality of this result, we prove it for two models most commonly used in the consumer search literature, namely sequential and simultaneous search models. Specifically, we show

that

$$\frac{\partial Pr(y_t = y_{t-1})}{\partial c} > 0,$$

where y_t and y_{i-1} are purchase decisions of a specific consumer in sessions t and $t-1$ and where we suppress consumer index i to simplify notation.

Throughout this section, we assume that the utility for product j in session t is given by

$$u_{jt} = \xi_j + \varepsilon_{jt}. \tag{A.1}$$

Here, ξ_j and $(\xi_j + \varepsilon_{jt})$ denote presearch and postsearch utilities. We assume that ε_{jt} are independent and identically distributed across products and sessions with cumulative distribution function (cdf) $F(\varepsilon)$. Note that we do not take a stance on which product attributes are included in ξ_j and ε_{jt}. The postsearch part of utility ε_{jt} can be interpreted as a product-specific match of value, price, or some other product attribute that is not known before search. We consider the case in which presearch utilities ξ_j are time invariant. This assumption implies that, in both sequential and simultaneous search models, the order in which the consumer adds products to the search set is deterministic and follows descending presearch utilities ξ_j.[34]

Because presearch utilities are time invariant and postsearch utilities are independent over time, the purchase probability of a given product remains identical across search sessions. Therefore, the probability of choosing the same product in consecutive time periods is given by

$$Pr(y_t = y_{t-1}) = \sum_j Pr(y_t = j)^2. \tag{A.2}$$

In our proof, we first show that choice persistence always increases when we reduce the purchase share of a given product and relocate that probability mass to other products with higher purchase probabilities. In a second step, we show that increasing search costs induces a series of changes that relocate purchase probabilities as described in the previous result; therefore, any increase in search costs necessarily increases choice persistence.

A.1. Relocating Purchase Probabilities Increases Persistence

We start by showing that choice persistence increases when the purchase share of a specific product k, s_k, is relocated to products with higher purchase probabilities. We assume products are ordered in terms of descending purchase probabilities: $s_1 > s_2 > ... > s_J$.

Proposition A.1. *Suppose the purchase share of product k is reduced by $\alpha_k \cdot s_k$, where $\alpha_k \in (0,1]$. At the same time, suppose that the probability mass $\alpha_k \cdot s_k$ is relocated to products $1, ..., k-1$ so that s_1 increases by $\alpha_1 \cdot s_k$, s_2 increases by $\alpha_2 \cdot s_k$, and so on, where $\alpha_1, ..., \alpha_{k-1}$ are positive allocation weights that sum to α_k (i.e., $\alpha_1 \geq 0, ..., \alpha_{k-1} \geq 0$ and $\sum_{j=1}^{j=k-1} \alpha_j = \alpha_k$). If $Pr(y_t = y_{t-1})$ and $\tilde{Pr}(y_t = y_{t-1})$ denote choice persistence before and after such relocation, then we always have $\tilde{Pr}(y_t = y_{t-1}) > Pr(y_t = y_{t-1})$.*

Proof. We compute choice persistence $\tilde{Pr}(y_t = y_{t-1})$ after the relocation of purchase shares by plugging new purchase shares into Equation (A.2):

$$\tilde{Pr}(y_t = y_{t-1}) = \sum_{j=1}^{j=k-1} (s_j + \alpha_j s_k)^2 + ((1-\alpha_k)s_k)^2 + \sum_{j=k+1}^{j=J} s_j^2$$

$$= \sum_{j=1}^{j=k-1} \left(s_j^2 + 2\alpha_j s_j s_k + (\alpha_j s_k)^2 \right) + (1 - 2\alpha_k + \alpha_k^2)s_k^2 + \sum_{j=k+1}^{j=J} s_j^2.$$

$$= Pr(y_t = y_{t-1}) + s_k^2 \left[\sum_{j=1}^{j=k} \alpha_j^2 + 2\left(\sum_{j=1}^{j=k-1} \alpha_j \frac{s_j}{s_k} - \alpha_k \right) \right]$$

The first term in square brackets is a sum of squared allocation shares that is always nonnegative, that is, $\sum_{j=1}^{j=k} \alpha_j^2 \geq 0$. The second term in square brackets is positive because $s_j > s_k$ for $\forall j < k$ by assumption, which implies that, for those products, $\frac{s_j}{s_k} > 1$ and that $\sum_{j=1}^{j=k-1} \alpha_j \frac{s_j}{s_k} - \alpha_k > \sum_{j=1}^{j=k-1} \alpha_j - \alpha_k = 0$. It follows that the complete expression in square brackets is positive, implying that $\tilde{Pr}(y_t = y_{t-1}) > Pr(y_t = y_{t-1})$. $\square$

Intuitively, Proposition A.1 states that, when products are ordered in terms of their purchase probabilities, any relocation in purchase probabilities "upward" (i.e., from any specific product to any combination of products with higher purchase shares) leads to an increase in persistence. For simplicity of exposition, we refer to this type of change in purchase shares as an "upward relocation."

A.2. Sequential Search

To analyze the sequential search case, we rely on the insight that the sequential search model can be equivalently formulated as a discrete choice problem (Armstrong 2017, Choi et al. 2018). Specifically, these papers show that the search model can be equivalently represented as a discrete choice model in which the consumer chooses a product with the highest effective utility w_{kt}, defined as $w_{kt} = min(u_{kt}, z_k)$, where u_{kt} is the postsearch utility of product k and z_k is its reservation utility. Following Choi et al. (2018), we decompose the reservation utility of product k as $z_k = \xi_k + \varepsilon_k^*$, where ε_k^* denotes the threshold value for product k that decreases monotonically in search costs. The effective utility of product k can then be expressed as $w_{kt} = \xi_k + min(\varepsilon_{kt}, \varepsilon_k^*)$.

Using this expression, we can compute the probability that product j is purchased as

$$\begin{aligned} s_j &= Pr(w_{jt} \geq w_{kt} \quad \forall\, k \in J) \\ &= Pr(\xi_j + min(\varepsilon_{jt}, \varepsilon_j^*) \geq \xi_k + min(\varepsilon_{kt}, \varepsilon_k^*) \quad \forall\, k \in J). \end{aligned}$$

Without loss of generality, we let the product index $j \in \{1, ..., J\}$ denote the ordering of products in terms of their presearch utilities so that $\xi_1 > \xi_2 > ... > \xi_J$ and, therefore, $z_1 > z_2 > ... > z_J$.

The following proposition shows that increasing the search cost of one specific product generates an upward relocation of purchase shares.

Proposition A.2. *Consider the case in which product-specific search costs obey the following restrictions: $c_k = c_j$ for products $k \leq j$ and $c_k \geq c_j$ for products $k > j$. Suppose the search cost of product j, c_j, is increased by some amount $\triangle > 0$ small enough such that the order*

of search remains unaffected. It follows that purchase share s_k weakly decreases for product $k = j$, weakly increases for products $k < j$, and remains the same for products $k > j$. Formally, we have

(1) $s_k(c_j + \triangle) - s_k(c_j) \leq 0$ for $k = j$

(2) $s_k(c_j + \triangle) - s_k(c_j) \geq 0$ for $k < j$

(3) $s_k(c_j + \triangle) - s_k(c_j) = 0$ for $k > j$

Proof. Because ε_k^* is decreasing in search costs, a higher value of the search cost c_j implies a lower value of the threshold ε_j^*, and hence, the effective utility of product j weakly decreases

$$w_j(c_j + \triangle) = \xi_j + min(\varepsilon_{jt}, \varepsilon_j^*(c_j + \triangle)) \leq \xi_j + min(\varepsilon_{jt}, \varepsilon_j^*(c_j)) = w_j(c_j).$$

Because w_j is weakly smaller for all realizations of ε_{jt} and effective utilities for other products remain unchanged, the purchase probability of j weakly decreases, and the purchase probability of all other products weakly increases. This proves statements (1) and (2).

Consider products $k > j$. If $\varepsilon_j < \varepsilon_j^*(c_j + \triangle)$, then the effective utility of product j remains the same, that is, $w_j(c_j + \triangle) = w_j(c_j)$. In this case, the comparison between w_j and w_k remains unchanged, so $w_j(c_j + \triangle) \geq w_k$ if and only if $w_j(c_j) \geq w_k$.

Next, consider the other case in which $\varepsilon_j \geq \varepsilon_j^*(c_j + \triangle)$. Because we assume that the order of search is unchanged, the reservation utilities of j and k are still ordered in the same way so that $z_j(c_j + \triangle) > z_k$. Using this observation, we can show that the effective utility of product j is strictly larger than the effective utility of product k:

$$\begin{aligned} w_j(c_j + \triangle) &= \xi_j + \varepsilon_j^*(c_j + \triangle) \\ &= z_j(c_j + \triangle) \\ &> z_k \\ &= \xi_k + \varepsilon_k^* \\ &\geq w_k, \end{aligned}$$

where the first equality follows because $\varepsilon_{jt} \geq \varepsilon_j^*(c_j + \triangle)$, and hence, $min(\varepsilon_{jt}, \varepsilon_j^*(c_j + \triangle)) = \varepsilon_j^*(c_j + \triangle)$. The second and fourth equality follow from the definition of the reservation utility. The final inequality follows because the highest possible value w_k can take on is when the second element in the minimum function of w_k is binding. In summary, we have

$$w_j(c_j) \geq w_j(c_j + \triangle) > w_k,$$

and hence, product j is chosen over product k under search cost c_j and under the new search cost $c_j + \triangle$.

Therefore, $w_j(c_j + \triangle) \geq w_k$ if and only if $w_j(c_j) \geq w_k$ for all realization of ε_j. Because the effective utilities for other products remain unchanged, it follows that $s_k(c_j + \triangle) - s_k(c_j) = 0$ for $k > j$. □

Intuitively, Proposition A.2 states that an increase in product-specific search costs decreases the purchase probability for the focal product, increases purchase probabilities for any product searched earlier, and does not affect purchase probabilities of products searched later. Therefore, this change in purchase shares constitutes an upward relocation. By combining Propositions A.1 and A.2, we can now show that, in the sequential search model, persistence increases when search costs increase.

Theorem A.1. *In the sequential search model, increasing search costs of all products by a fixed amount $\Delta > 0$ leads to an increase in choice persistence for any value of Δ.*

Proof. To prove this theorem, we first note that, for a given level of search costs, products with higher presearch utilities ξ_j also have higher purchase probabilities. This follows because $s_j = Pr(\xi_j + min(\varepsilon_{jt}, \varepsilon_j^*) \geq \xi_k + min(\varepsilon_{kt}, \varepsilon_k^*)$ $\forall k \in J)$,, where in the case of uniform search costs and i.i.d. taste shocks ε_{jt}, the term $min(\varepsilon_{jt}, \varepsilon_j^*)$ is identically distributed across products. Therefore, a higher value of ξ_j leads to a higher purchase share s_j.

We now construct a series of product-specific changes in search costs that generate upward relocations of purchase shares. We start from a given search cost level c_1, which is equal across products, and then increase search costs by a given amount for each product such that new search costs are equal to $c_1 + \Delta$. We first increase search costs in this fashion for the lowest reservation utility product and continue in ascending order of reservation utilities. According to Proposition A.2, increasing search costs for the lowest reservation utility product J leads to an upward relocation of purchase shares from product J, which has the lowest purchase share, to all other products. The search cost increase does not change the search order because product J has the lowest reservation utility before the change in search costs, and the increase in search costs further lowers its reservation utility.

Next, we increase search costs by the same amount for product $J - 1$. Using Proposition A.2 again, we conclude that this change in search costs leads to an upward relocation of purchase shares from product $J - 1$ to all other products $j < J - 1$. Search order does not change because, after the change in search costs for $J - 1$, all products $j < J - 1$ have identical search costs, and all products $j \geq J - 1$ have identical search costs that are higher than those of products $j < J - 1$. Therefore, the order of reservation utilities does not change. We continue this process until we reach product $j = 1$. When search costs are altered for the final product, purchase shares do not change because $j = 1$ continues to be searched first.

Once the sequence of relocations is completed, the search costs of all products have increased to $c_2 = c_1 + \Delta$. Because each individual change in the constructed sequence is an upward relocation, it follows from Proposition A.1 that choice persistence has increased. □

A.3. Simultaneous Search

To prove our result for the simultaneous search model, we rely on several properties of the optimal search strategy. We state these properties without proofs, but an interested reader may consult Chade and Smith (2006) for the relevant derivations. As before, we assume that products are ordered by their presearch utilities ξ_j.

1. The consumer adds products to the search set in order of decreasing presearch utilities ξ_j.

2. For a given level of search costs, the consumer's search set size is deterministic. Search set size does not depend on the realizations of postsearch utilities $(\xi_j + \varepsilon_{jt})$ because the

consumer commits to a search set size before starting the search.

3. Search set size is weakly decreasing in search costs.

4. Purchase shares are ordered in the same way as presearch utilities, so purchase probabilities decrease in j.

In the following proposition, we show that increasing the search cost in the simultaneous search model leads to a change in purchase probabilities from the products that are no longer searched to the products with the highest presearch utilities.

Proposition A.3. *Suppose search costs of all products are increased by some amount $\triangle > 0$ from c_1 to $c_2 = c_1 + \triangle$. If the consumer optimally searched k_1 products before the change, then there exists a product $k_2 \leq k_1$ such that purchase shares s_j weakly increase for products $j \leq k_2$ and weakly decrease for products $j > k_2$. Formally, we have*

$$(1) \quad s_j(c_1 + \triangle) - s_j(c_1) \geq 0 \quad \text{for } j \leq k_2$$
$$(2) \quad s_j(c_1 + \triangle) - s_j(c_1) \leq 0 \quad \text{for } j > k_2$$

Proof. If the optimal search set size is k_1 products under the search cost c_1, then purchase probabilities are given by

$$Pr_j(c_1) = Pr(\xi_j + \varepsilon_{jt} \geq \xi_l + \varepsilon_{lt} \quad \forall l \in \{1, \dots, k_1\}) \text{ if } j \leq k_1$$
$$= 0 \text{ if } j > k_1.$$

Any product $j > k_1$ not included in the search set has a purchase probability of zero. When search costs increase from c_1 to c_2, the search set size weakly decreases, and hence, purchase probabilities are now given by

$$Pr_j(c_2) = Pr(\xi_j + \varepsilon_{jt} \geq \xi_l + \varepsilon_{lt} \quad \forall l \in \{1, \dots, k_2\}) \text{ if } j \leq k_2$$
$$= 0 \text{ if } j > k_2,$$

where $k_2 \leq k_1$.

Comparing the two sets of purchase shares, we conclude that, if search set size remains the same, purchase probabilities also stay the same. If the search set size decreases such that $k_2 < k_1$ with strict inequality, then the purchase probabilities for products $k_2 < j \leq k_1$ decrease to zero and remain zero for any product $j > k_1$ that was not searched at the previous level of search costs. At the same time, the purchase probabilities of products $j \leq k_2$ increase. To see this, note that

$$Pr(\xi_j + \varepsilon_{jt} \geq \xi_l + \varepsilon_{lt} \quad \forall l \in \{1, \dots, k_1\})$$
$$= Pr(\xi_j + \varepsilon_{jt} \geq \xi_l + \varepsilon_{lt} \quad \forall l \in \{1, \dots, k_2\}$$
$$\text{AND} \quad \xi_j + \varepsilon_{jt} \geq \xi_m + \varepsilon_{mt} \quad \forall m \in \{k_2 + 1, \dots, k_1\})$$
$$< Pr(\xi_j + \varepsilon_{jt} \geq \xi_l + \varepsilon_{lt} \quad \forall l \in \{1, \dots, k_2\}),$$

which holds for all products j such that $j \leq k_2$. This last inequality holds because, other things equal, each product j is more likely to be selected when it is being chosen from a smaller set. □

By combining Propositions A.1 and A.3, we can now show that, in the simultaneous search model, persistence increases when search costs increase.

Theorem A.2. *In the simultaneous search model, increasing search costs of all products by a fixed amount $\triangle \geq 0$ leads to an increase in choice persistence for any value of $\triangle$.*

Proof. First, note that products with higher presearch utility ξ_j also have a higher purchase share.[35] Let c_1 denote the initial level of search costs, and let $c_2 = c_1 + \triangle$ denote its value after the change. According to Proposition A.3, there exists a product $k_2 \leq k_1$ such that, when search costs change from c_1 to c_2, purchase shares s_j weakly increase for products $j \leq k_2$ and weakly decrease for products $j > k_2$. Starting from the initial purchase probabilities under c_1, we can iteratively decrease the purchase probability of the product with the lowest nonzero purchase share for all products $k_2 \leq j < k_1$, each time relocating the probability mass to products $j \leq k_2$. Each of these relocations leads to a new distribution of purchase shares in which shares of products $j \leq k_2$ become weakly higher. Hence, increasing search costs from c_1 to c_2 induces changes in purchase shares that can be achieved by a series of upward relocations of this kind. It then follows from Proposition A.1 that choice persistence increases. □

Appendix B. Search Costs Increase Estimates of Preference Heterogeneity

In this section, we show that, in the presence of positive search costs, the bias in the inferred intercept of a given product $(\hat{\xi}_j - \xi_j)$ is increasing in ξ_j. From this statement, it follows that, for any pair of consumers, denoted as consumers 1 and 2, it holds that inferred intercepts are further apart from each other than true intercepts. For example, when $\xi_{1j} > \xi_{2j}$, we infer intercepts are such that $\hat{\xi}_{1j} - \hat{\xi}_{2j} > \xi_{1j} - \xi_{2j}$, and therefore, search costs increase the dispersion in inferred intercepts across consumers. Furthermore, the absolute value of the bias is increasing in search costs, thus leading to a stronger increase in dispersion if search costs are higher.

We show that this result holds through an extensive set of simulations. We first simulate consumer behavior for a market with two products, in which ξ_A is normalized to zero and in which we change the true value of ξ_B. Figure B.1 plots the bias in the inferred intercept as a function of the true intercept value (for the one non-normalized product). We repeat this exercise for two distributions of taste shocks that are commonly used in discrete choice demand models: normal (left graph) and extreme value (right graph). The graphs show that the sign of the bias in inferred intercepts coincides with the sign of the true intercept ξ_B so that positive intercepts are overestimated and negative intercepts are underestimated.[36] Moreover, the absolute value of the bias $|\hat{\xi}_B - \xi_B|$ increases in the absolute value of the intercept $|\xi_B|$. Therefore, the bias is increasing in the value of the true intercept, and for any pair of consumers, the inferred intercepts $\hat{\xi}_{1B}$ and $\hat{\xi}_{2B}$ are further apart from each other than true intercepts ξ_{1B} and ξ_{2B}. Furthermore, the absolute value of the bias increases in search costs, and therefore, higher search costs lead to higher inferred heterogeneity in preferences.

We repeat these simulations for a case with three products in which we fix the intercepts of two products and compute the bias in the intercept of the third product for different intercept values. The two graphs in Figure B.2 show the results from such a simulation for the case in which we set $\xi_A = 0$ and $\xi_B = 0.05$ and vary the intercept of the third product, ξ_C. Similar to the two-product case, we find that intercepts are overestimated (underestimated) if the true intercept is positive (negative) and the bias is increasing in

Figure B.1. Bias in Inferred Intercepts: Two Products
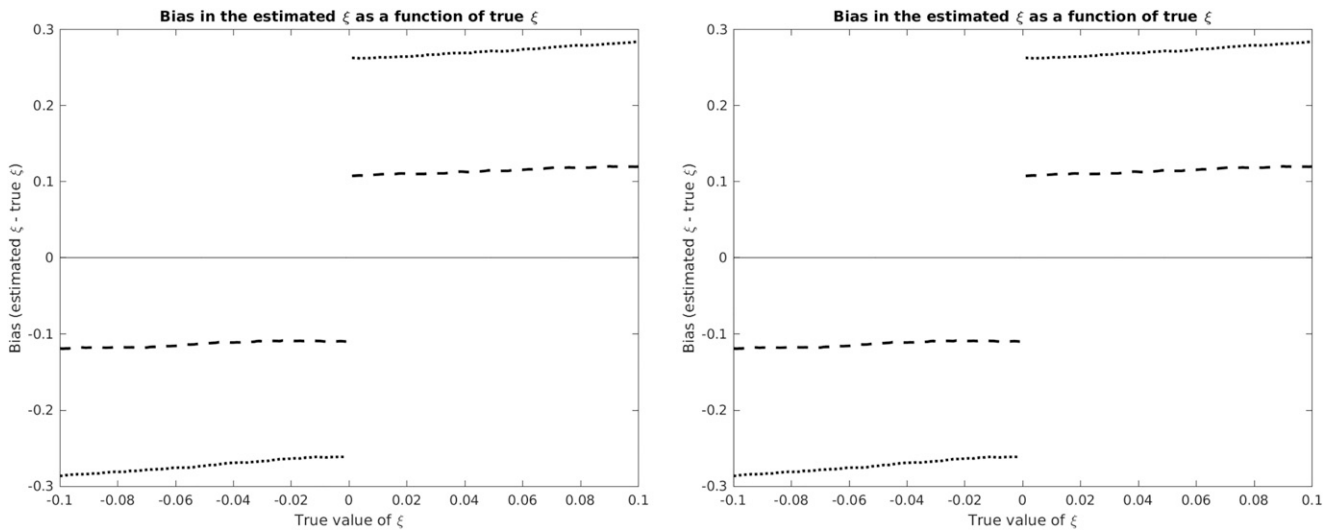
Notes. The two graphs display the difference between inferred and true intercepts under the assumption of a normal (left graph) or extreme value (right graph) taste shocks. We assume $\xi_A = 0$ and plot the bias in ξ_B as a function of the true intercept value. The solid/dashed/dotted lines correspond to the cases in which search costs are equal to $0/0.25/0.4$.

the value of the true intercept. Furthermore, the absolute value of the bias is increasing in search costs.

We repeat similar simulation exercises for other distributions of taste shocks ε (logistic and uniform) as well as for settings with more than three products, and we find that, throughout all simulations, inferred heterogeneity in preferences is increasing in search costs.

Appendix C. Normalization of Taste Shock Variances

When estimating our model, we normalize the variances of both taste shocks to one, setting $\sigma_\mu^2 = 1$ and $\sigma_\varepsilon^2 = 1$. In

this appendix, we discuss why we need both normalizations and explain why they prevent us from monetizing the estimates of search costs.

To illustrate the role of two normalizations, we start by writing consumer utility in its non-normalized form:

$$u_{ijt} = \xi_{ij} - \alpha_i p_{ijt} + \sigma_\mu \mu_{ijt} + \sigma_\varepsilon \varepsilon_{ijt},$$

where taste shocks μ_{ijt} and ε_{ijt} are i.i.d. and follow the standard normal distribution. Next, we divide both utility and search costs by the standard deviation σ_μ of the presearch

Figure B.2. Bias in Inferred Intercepts: Three Products
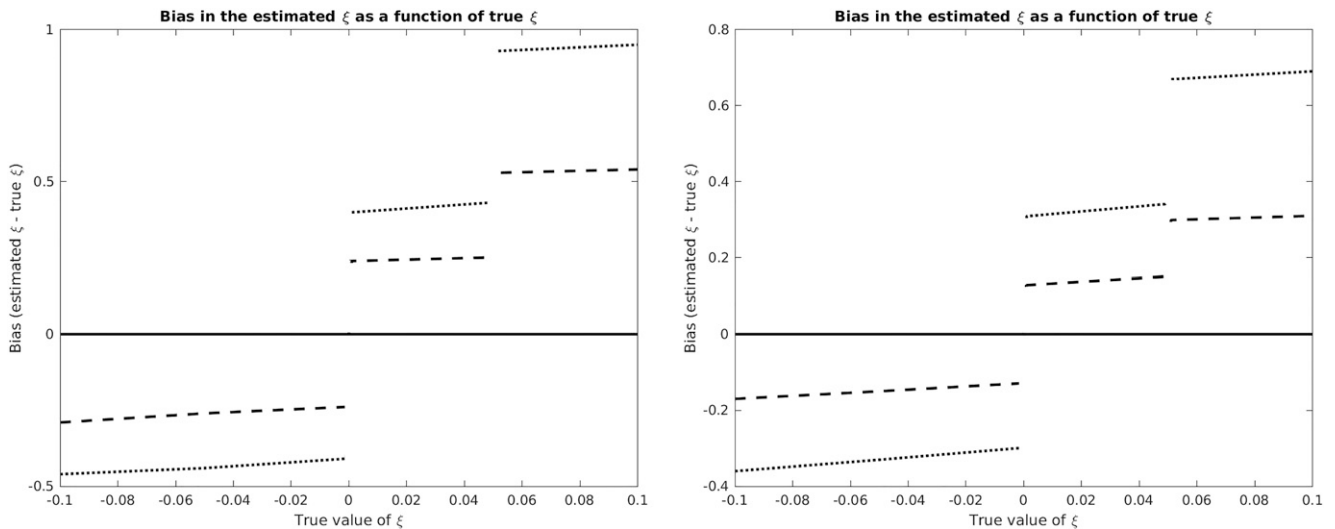
Notes. The two graphs display the difference between inferred and true intercepts under the assumption of a normal (left graph) or extreme value (right graph) taste shocks. We assume $\xi_A = 0$, $\xi_B = 0.05$ and plot the bias in ξ_C as a function of the true intercept value. The solid/dashed/dotted lines correspond to the cases in which search costs are equal to $0/0.25/0.4$.

shocks to obtain:

$$\tilde{u}_{ijt} = \xi_{ij}/\sigma_\mu - (\alpha_i/\sigma_\mu)p_{ijt} + \mu_{ijt} + (\sigma_\varepsilon/\sigma_\mu)\varepsilon_{ijt}$$
$$= \tilde{\xi}_{ij} - \tilde{\alpha}_i p_{ijt} + \mu_{ijt} + \tilde{\sigma}_\varepsilon \varepsilon_{ijt}$$
$$\tilde{c}_{ij} = c_{ij}/\sigma_\mu,$$

where we use the tilde notation for all variables scaled by σ_μ (e.g., $\tilde{u}_{ijt} = u_{ijt}/\sigma_\mu$). Because this transformation rescales utilities, search costs, and reservation utilities in the same way, we have not altered search order, stopping, or purchase decisions of consumers. Thus, we set $\sigma_\mu = 1$ without loss of generality.

Now, consider estimating the standard deviation of postsearch shocks, $\tilde{\sigma}_\varepsilon$. Because increasing this standard deviation $\tilde{\sigma}_\varepsilon$ and decreasing search costs both lead to more search, we can only identify search costs relative to the postsearch shock variance $\tilde{\sigma}_\varepsilon$. Although search costs and $\tilde{\sigma}_\varepsilon$ are not fully colinear, in practice, multiplying the postsearch standard deviation by a factor k roughly scales the estimated search costs by a factor of k while leaving other parameters mostly unchanged. Therefore, there appears to be little variation in the data that would help us estimate $\tilde{\sigma}_\varepsilon$, which is why we set it equal to one in our application.

These normalizations have one important implication for how we interpret estimated parameters. Although we can monetize our estimates of product intercepts, we are unable to monetize search costs. As discussed, we can only identify search costs relative to the standard deviation $\tilde{\sigma}_\varepsilon$; therefore, the estimated search cost in the rescaled model is given by

$$\frac{\tilde{c}_{ij}}{\tilde{\sigma}_\varepsilon} = \frac{c_{ij}/\sigma_\mu}{\sigma_\varepsilon/\sigma_\mu} = c_{ij}/\sigma_\varepsilon.$$

Therefore, when we calculate the ratio of the estimated search costs to the estimated price coefficient, we are computing the ratio of α_i/σ_μ and $c_{ij}/\sigma_\varepsilon$, which does not equal the monetary value of search costs. However, σ_ε affects only estimated search costs but not utility parameters known to the consumer prior to searching, such as the price coefficient and product intercepts. Therefore, we cannot monetize search costs but can monetize product intercepts.

Appendix D. Identification: The Role of Purchase Data

In this section, we illustrate how *conditional* purchase probabilities behave as a function of product intercepts and product-specific search costs. To simplify exposition, we consider a setting with four products—A, B, C, and D—and focus on the behavior of a consumer who searches these products in alphabetical order. We first set $\xi_A = 2.2$, $\xi_B = 1.8$, $\xi_C = 1.6$, $\xi_D = 1.4$ and $c_A = c_B = c_C = c_D = 0.1$. For simplicity, we omit the presearch taste shock in this example. Starting from this baseline, we analyze changes in behavior in response to two changes in model parameters. First, we increase the intercept of product B, and second, we decrease the product-specific search cost for product B. In both cases, we analyze changes that are small enough so that search order remains constant. In Figure D.1, we plot the purchase probability for product B conditional on different sets of searched products when altering either the intercept value or search costs. The graphs show that the change in the intercept leads to higher purchase probabilities for product B conditional on searching products A, B, and C or A, B, C, and D (i.e., sets in which product B is not the last product the consumer searched), whereas the change in the search cost leaves those probabilities unchanged. Therefore, product-specific rates of recall (the probability of purchasing a product searched earlier in the search sequence) can help us identify preferences separately from search costs.

The intuition for the difference in behavior highlighted in Figure D.1 can be gleaned from Equations (6) and (7). The probability of continuing or stopping does not depend on the reservation utilities of products searched previously but does depend on realized utilities of those products. Therefore, the distribution of postsearch taste shocks conditional on searching a specific set of products does not change when altering search costs of any product that is not the last product searched. As a consequence, conditional purchase probabilities remain the

Figure D.1. Conditional Purchase Probabilities When Varying Intercept Value or Search Costs
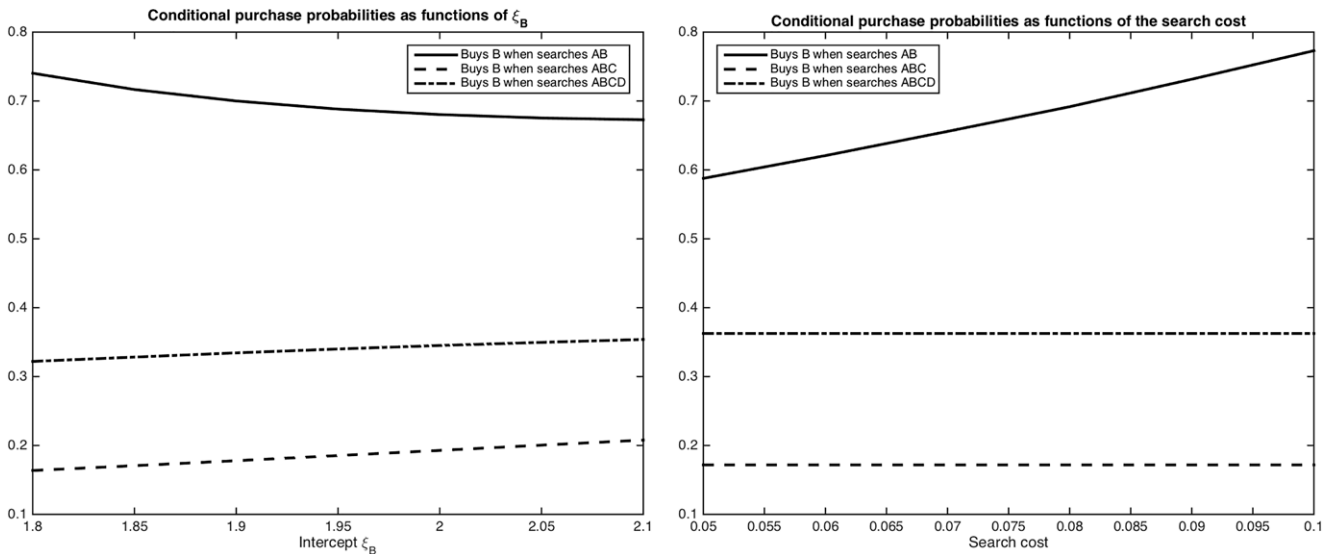
same. Instead, increasing the intercept does change the continuation and stopping decisions at each step, thus leading to a different conditional distribution of post-search taste shocks in different searched sets and, therefore, to different conditional purchase probabilities.

We also note that it is not generally true that conditional purchase probabilities are unaffected by changes in search costs. In the preceding example, the purchase probability of product B conditional on searching A and B does change when altering the search cost of product B (see the right-hand graph in Figure D.1). This change occurs because the realizations of postsearch taste shocks for which set (A,B) is optimal are a function of product B's reservation utility (because product B is the last product searched) and, hence, change when this product's search costs are altered. In this particular example, lowering the search cost of product B makes it more likely that the consumer searches product B (i.e., continues to search after evaluating product A), even if the postsearch taste shock draw for product A is relatively favorable, and hence, the purchase probability of product B conditional on searching A and B decreases.

Appendix E. Estimation Details
In Section 5, we outlined the main features of the importance-sampling estimator. This section provides additional details. With regard to simulation draws, we set $N_S = 1,000$ and $N_M = 100$. That is, for each consumer, we take 100 draws from the proposal distribution of types $g(\boldsymbol{\theta}_i)$. For each of these draws, we take 1,000 draws from the distributions of utility shocks $\boldsymbol{\varepsilon}_{it}$ and $\boldsymbol{\mu}_{it}$ for each consumer–session pair.

To compute individual likelihoods in (23), we need to first calculate reservation utilities z_{ijt} for each set of random draws $\boldsymbol{\theta}_i$ and $\boldsymbol{\mu}_{ijt}$. Calculating reservation utilities is computationally burdensome, so in practice, we use approximations. Recall that z_{ijt} is defined from the following indifference condition, which is equivalent to Equation (2):

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

Under our assumption that ε_{ijt} follows a normal distribution with standard deviation σ_ε, this condition can be written in a more convenient form:

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

where $\Phi(\cdot)$ and $\phi(\cdot)$ represent the Gaussian cdf and probability density function. 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 et al. 2019). To ease computation, we precompute a lookup table that stores the values of reservation utilities z_{ijt} for a large number of combinations of δ_{ijt} and c_{ij}. 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.

Appendix F. Additional Details: Persistence in Information
In this section, we provide additional details on the extension of our main model that allows for correlation in

postsearch taste shocks across search sessions. Recall that, in this extended model, taste shocks ε_{ijt} follow an MA(1) process so that $\varepsilon_{ijt} = \omega_{ijt} + \rho\omega_{ijt-1}$, where ω_{ijt} is drawn from $N(0, 1)$ and is i.i.d. across consumers, products, and search sessions. Upon searching product j in period t, consumer i observes both the value of the postsearch shock ε_{ijt} and the value of the transitory shock ω_{ijt}. These assumptions imply the following information structure. If consumer i did not search product j in period $t - 1$, the consumer expects ε_{ijt} to be distributed according to $\varepsilon_{ijt} \sim N(0, \sigma_\omega^2 \cdot (1 + \rho^2))$. By contrast, if consumer i searched product j in the previous period, the consumer expects ε_{ijt} to be distributed according to $\varepsilon_{ijt}|\omega_{ijt-1} \sim N(\rho\omega_{ijt-1}, \sigma_\omega^2)$.

The likelihood now also depends on the persistence parameter ρ. As before, $D_i = \{\pi_{it}, S_{it}, y_{it}\}_{t=1}^{T_i}$ is the data for all T_i sessions of consumer i, $D = \{D_i\}_{i=1}^N$ is for all consumers, and $\boldsymbol{\Omega}$ is a vector of hyperparameters that determine the distribution of preferences and search costs. We take N_M draws of types from the proposal density $g(\boldsymbol{\theta}_i)$ and compute the simulated likelihood

$$\tilde{L}(D|\boldsymbol{\Omega}, \rho, \boldsymbol{p}) = \prod_{i=1}^N \frac{1}{N_M} \sum_{m=1}^{N_M} \left(\prod_{t=1}^{T_i} \tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \rho, \boldsymbol{p}_{it}) \cdot \frac{f(\boldsymbol{\theta}_i^m|\boldsymbol{\Omega})}{g(\boldsymbol{\theta}_i^m)} \right),$$
(F.1)

where $\boldsymbol{\theta}_i^m$ denotes the mth draw of types for consumer i, and $\tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \rho, \boldsymbol{p}_{it})$ is the simulated consumer and session-specific contribution to the likelihood:

$$\tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \rho, \boldsymbol{p}_{it}) = \frac{1}{N_S} \sum_{s=1}^{N_S} 1\left\{ W_{it}(\boldsymbol{\theta}_i^m, \boldsymbol{\mu}_{it}^s, \boldsymbol{\omega}_{it}^s, \boldsymbol{\omega}_{it-1}^s, S_{it-1}^{(s)}, \boldsymbol{p}_{it}) \geq 0 \right\}.$$
(F.2)

In this expression, $\boldsymbol{\mu}_{it}^s$, $\boldsymbol{\omega}_{it}^s$, and $\boldsymbol{\omega}_{it-1}^s$ are draws of taste shocks and S_{it-1} is the search set in session $(t-1)$. For $(t > 1)$, we observe S_{it-1} and can simply condition on the search set of the previous period. In the case of $(t = 1)$, the previous search set S_{i0} is unobserved, and hence, we simulate realizations of S_{i0}^s.

To simulate draws S_{i0}^s, we simulate decisions of consumer i (for a given type draw $\boldsymbol{\theta}_i^m$ and persistence parameter ρ) for a large number of search sessions, initiating simulation at an arbitrary starting point. Throughout this simulation, we set product prices equal to their average levels observed in our data set. We then discard 1,000 first simulations and consider the remaining simulated sessions to be draws from the stationary distribution $F(S_{i0}|\boldsymbol{\theta}_i, \rho)$.

To estimate the model, we implement a grid search over different values of ρ between -1 and 1. For every candidate value of ρ, we simulate $F(S_{i0}|\boldsymbol{\theta}_i, \rho)$ and then estimate the hyperparameters of the preferences and search cost distributions $\boldsymbol{\Omega}$ using our importance sampling method. Our estimation procedure for this extended model then proceeds in the following steps:

1. Fix the correlation parameter ρ.
2. Fix the vector of hyperparameters $\boldsymbol{\Omega}$.
3. Generate draws of consumer types $\boldsymbol{\theta}_i$ from the proposal distribution $g(\boldsymbol{\theta}_i)$.
4. For each type draw $\boldsymbol{\theta}_i^m$, simulate draws S_{i0}^s as explained.
5. For each type draw $\boldsymbol{\theta}_i^m$ and each draw of the initial condition S_{i0}^s, compute the likelihood $\tilde{L}_{it}(D_{it}|\boldsymbol{\theta}_i^m, \rho, \boldsymbol{p}_{it})$ using

simulation as in the main model. To simplify estimation, we only use one draw of the initial condition S_{i0}^s but 100 draws for transitory shocks ω_{ijt} and presearch shocks μ_{ijt}.

6. Plug the values of the simulated likelihoods $\tilde{L}_{it}(D_{it}|\theta_i^m, \rho, p_{it})$ into the importance sampling formula. Average across proposal types θ_i to compute the integrated individual likelihood and multiply likelihoods of all consumers to obtain total likelihood $\tilde{L}(D|\Omega, \rho, p)$.

7. Maximize $\tilde{L}(D|\Omega, \rho, p)$ with respect to hyperparameters Ω to obtain an estimate $\hat{\Omega}(\rho)$ conditional on the value of the correlation parameter ρ.

8. Repeat steps 4–7 for different candidate values of the correlation parameter ρ.

9. Using grid search, maximize $\tilde{L}(D|\hat{\Omega}(\rho), \rho, p)$ with respect to the correlation parameter ρ.

Appendix G. Importance Sampling vs. Kernel-Smoothed Frequency Estimator

In this section, we compare our approach with a kernel-smoothed frequency estimator. As we show, using a kernel-smoothed frequency estimator in our application is computationally more burdensome and relatively slow. To illustrate this point, we generate a simulated data set that mimics our actual data, and we use this simulated data set to show how the computational burden increases with sample size. Specifically, we estimate a model with 10 products and parameterize the utility function and search cost parameters as in our main specification. For simplicity, we omit the price coefficient and assume that consumers have identical search costs.

Tables G.1 and G.2 show estimation results obtained using two estimation methods for different sample sizes (50, 100, and 500 consumers). Table G.3 summarizes the relative

Table G.1. Kernel-Smoothed Frequency Estimator

	Initial point	True value	Number of consumers		
			$N = 50$	$N = 100$	$N = 500$
$\bar{\xi}_2$	−0.50	−0.10	−0.17	−0.14	−0.12
$\bar{\xi}_3$	−0.50	−0.20	−0.23	−0.20	−0.20
$\bar{\xi}_4$	−0.50	−0.30	−0.32	−0.31	−0.24
$\bar{\xi}_5$	−0.50	−0.40	−0.33	−0.32	−0.31
$\bar{\xi}_6$	−0.50	−0.50	−0.41	−0.41	−0.37
$\bar{\xi}_7$	−0.50	−0.60	−0.47	−0.44	−0.43
$\bar{\xi}_8$	−0.50	−0.70	−0.53	−0.51	−0.48
$\bar{\xi}_9$	−0.50	−0.80	−0.65	−0.58	−0.52
$\bar{\xi}_{10}$	−0.50	−0.90	−1.96	−0.74	−0.60
σ_{ξ_2}	0.20	0.50	0.42	0.59	0.54
σ_{ξ_3}	0.20	0.50	0.43	0.67	0.71
σ_{ξ_4}	0.20	0.50	0.42	0.63	0.45
σ_{ξ_5}	0.20	0.50	0.41	0.57	0.61
σ_{ξ_6}	0.20	0.50	0.45	0.41	0.72
σ_{ξ_7}	0.20	0.50	0.45	0.47	0.47
σ_{ξ_8}	0.20	0.50	0.41	0.46	0.42
σ_{ξ_9}	0.20	0.50	0.42	0.46	0.52
$\sigma_{\xi_{10}}$	0.20	0.50	0.50	0.67	0.48
c	0.20	0.10	0.15	0.13	0.11
CPU time, minutes			99.3	113.9	269.5

Notes. The smoothing parameter is fixed at $h = 5$ as in Honka and Chintagunta (2017). The final three columns show the performance of the estimator across three samples of different size.

Table G.2. Importance Sampling Estimator

	Initial point	Truth	Number of consumers		
			$N = 50$	$N = 100$	$N = 500$
$\bar{\xi}_2$	−0.50	−0.10	−0.11	−0.11	−0.10
$\bar{\xi}_3$	−0.50	−0.20	−0.20	−0.20	−0.20
$\bar{\xi}_4$	−0.50	−0.30	−0.31	−0.29	−0.29
$\bar{\xi}_5$	−0.50	−0.40	−0.40	−0.40	−0.39
$\bar{\xi}_6$	−0.50	−0.50	−0.48	−0.49	−0.51
$\bar{\xi}_7$	−0.50	−0.60	−0.58	−0.61	−0.60
$\bar{\xi}_8$	−0.50	−0.70	−0.70	−0.71	−0.70
$\bar{\xi}_9$	−0.50	−0.80	−0.81	−0.79	−0.80
$\bar{\xi}_{10}$	−0.50	−0.90	−0.89	−0.90	−0.91
σ_{ξ_2}	0.20	0.50	0.51	0.51	0.50
σ_{ξ_3}	0.20	0.50	0.52	0.50	0.50
σ_{ξ_4}	0.20	0.50	0.50	0.50	0.52
σ_{ξ_5}	0.20	0.50	0.51	0.51	0.51
σ_{ξ_6}	0.20	0.50	0.50	0.52	0.52
σ_{ξ_7}	0.20	0.50	0.48	0.49	0.51
σ_{ξ_8}	0.20	0.50	0.51	0.51	0.51
σ_{ξ_9}	0.20	0.50	0.51	0.54	0.53
$\sigma_{\xi_{10}}$	0.20	0.50	0.53	0.50	0.52
c	0.20	0.10	0.10	0.10	0.10
CPU time, minutes			8.0	19.6	33.6

Note. The final three columns show the performance of the estimator across three samples of different size.

performance of both estimators in terms of CPU time. We find that across all sample sizes, the importance sampling procedure is substantially faster. When extrapolating to a sample size of 4,000 consumers, we conclude that it would take us about three days to estimate the model using a kernel-smoothed frequency estimator. By contrast, the importance sampling approach allows us to estimate the model in about three hours. Additionally, the estimates in Tables G.1 and G.2 suggest that the importance sampling approach recovers structural parameters somewhat more accurately. As a whole, our results suggest that importance sampling provides a fast and practical way to estimate our model while, at the same time, retaining precision similar to that of alternative methods.

Table G.3. CPU Time Comparison: Kernel-Smoothed Frequency Estimator vs. Importance Sampling

	CPU time, minutes		
N (number of consumers)	Kernel-smoothed estimation	Importance sampling estimation	Kernel to importance sampling time ratio
50	99.3	8.0	12.4
100	113.9	19.6	5.8
500	269.5	33.6	8.0
Estimated CPU time, minutes			
1,000	561.1	58.3	9.6
2,000	1,467.2	106.9	13.7
4,000	4,573.4	203.9	22.4

Note. Estimated CPU time in the bottom panel is computed by fitting and extrapolating based on a quadratic function.

Appendix H. Computation of Personalized Prices

We compute expected profits as

$$\Pi(p_j, H) = \sum_{k \in \mathcal{F}} \int (p_k - mc_k) Pr(y = k|\boldsymbol{\theta})$$
$$f(\boldsymbol{\theta}|H, \boldsymbol{p}^H, \boldsymbol{\Omega}) g(\boldsymbol{\Omega}) d\boldsymbol{\theta} d\boldsymbol{\Omega}. \tag{H.1}$$

This profit function captures uncertainty regarding both type $\boldsymbol{\theta}$ and structural parameters $\boldsymbol{\Omega}$. To compute personalized prices, we first need to calculate the integrals in (H.1). These integrals do not have closed-form solutions and have to be approximated using simulation. We take bootstrap estimates $\hat{\boldsymbol{\Omega}}^b$ and treat them as draws from the distribution $g(\boldsymbol{\Omega})$.[37] For each of these draws $\hat{\boldsymbol{\Omega}}^b$, we use the Metropolis–Hastings algorithm to take S draws of types $\boldsymbol{\theta}_s^b$ from the posterior distribution $f(\boldsymbol{\theta}|H, p^H, \hat{\boldsymbol{\Omega}}^b)$. 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 (H.1) can then be approximated with the following expression:

$$\Pi(p_j, H) \approx \frac{1}{B \times S} \sum_{k \in \mathcal{F}} \sum_{b=1}^{B} \sum_{s=1}^{S} (p_k - mc_k) Pr(y = k|\boldsymbol{\theta}_s^b). \tag{H.2}$$

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

Endnotes

[1] Switching costs provide another possible explanation for choice persistence (Dubé et al. 2008, 2009). We discuss the role of switching costs in more detail at the end of this section.

[2] This conclusion may change when search costs vary substantially across products, in which case ignoring search may lead to underestimation of preference heterogeneity. In our application, however, we find that preference heterogeneity is overestimated even when we allow for different search costs across products.

[3] Honka (2014) finds that removing search costs from her estimated model has an effect on the persistence in consumers' choices that is 17 times larger than the effect of removing switching costs. We also expect the role of switching costs to be relatively less important in our context of consumer goods than in the car insurance market studied in Honka (2014).

[4] When introducing our empirical model in Section 4, we expand this framework by adding another taste shock known prior to search and adding price to the utility function.

[5] In our empirical application, we define search sessions retrospectively as all search activity within a week prior to a purchase. Because it is more difficult to define search sessions for consumers without purchases, we omit them from the analysis and assume that all consumers have to purchase one of the available options, thus removing the outside option. The search rules derived in this section would be similar in the presence of an outside option. One can simply think of the outside option as an additional product in the choice set that is always available to consumers (i.e., no search cost needs to be incurred). The search order condition would remain the same. The stopping rule maintains the same general structure, but the maximum realized utility of searched options now includes the outside option. Similarly, the purchase decision is based on the maximum utility from all searched options, including the outside option.

[6] We assume that the distribution of ε_{ijt} is continuous and has full support, which ensures that the solution to this equation exists and is unique for any value of c_{ij}.

[7] The reservation utility is time invariant because it only depends on the search costs c_{ij}, product intercepts ξ_{ij}, and the time-invariant distribution of the taste shock ε_{ijt}.

[8] Because match values ε_{ijt} have identical distribution across products, the ranking of reservation utilities in this model coincides with the ranking of product intercepts ξ_{ij}.

[9] Reservation utilities do not have closed-form solutions for most distributions $F_\varepsilon(\varepsilon)$. Using a numerical approximation, we compute $z_{1B} = -1.19$. Under the normality assumption, the probability of searching product B is then equal to $\Phi(-1.19) = 0.12$, where Φ denotes the standard normal cdf.

[10] For this exercise, we treat the consumer-specific purchase probabilities as observable.

[11] Note that we simplify this example by assuming identical search costs across products. One should keep in mind, however, that introducing differences in product-specific search costs may potentially tilt results toward underestimating preference heterogeneity (see the discussion in Section 3.2).

[12] Because the intercepts follow identical distributions for these three products, the distributions of resulting purchase probabilities are also identical.

[13] Although we were unable to formally prove this result, we demonstrate it using simulations. To this end, we simulate consumer behavior and compute inferred product intercepts under a variety of different modeling assumptions.

[14] 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.

[15] 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.

[16] Consumers' stopping decisions depend on the ratio of search costs and σ_ϵ^2, whereas purchase decisions depend only on σ_ϵ^2. Therefore, σ_ϵ^2 could, in principle, be identified by the degree of randomness in purchase decisions (a larger value of σ_ϵ^2 would lead to higher randomness in purchase decisions). In practice, we found that σ_ϵ^2 is hard to estimate, and hence, we normalize it.

[17] The consumer also stops searching if the consumer has exhausted all search opportunities. In this case, the stopping inequality is irrelevant.

[18] Yavorsky et al. (2021, p. 25) state, "To identify unobserved heterogeneity in a continuous mixture model, a researcher must observe consumers more than once, i.e., the data must have a panel structure."

[19] Kim et al. (2010) derive the exact shape of the reservation utility in the appendix of their paper.

[20] In this sense, data on first searches in our setting provide information similar to panel data on purchases in perfect information models.

[21] The function $g(c_{ij})$, which enters the reservation utility z_{ijt} is monotonically decreasing in search costs and ranges from minus infinity to plus infinity as we vary search costs from infinity to zero (Kim et al. 2010). Therefore, we can always change the search cost and the product intercept so that the mean reservation utility $\bar{z}_{ij}$ remains unchanged.

[22] An alternative approach to smoothing the objective function is to use a kernel-smoothed frequency estimator as in Honka (2014) and Honka and Chintagunta (2017); 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 parameters. In Appendix G, we use a simulation study to

compare computational time of a kernel-smoothed frequency estimator and the importance sampling approach.

[23] Our data provider uses cookies to track the browsing and purchase activity of users over time. Each time a user logs in, the website ties browsing history from a given cookie to the user's account. In our data, we observe anonymized identifiers of consumers that correspond to unique accounts.

[24] In principle, one could exclude very short searches because these might reflect navigational searches of consumers who already know which product they wish to purchase. We do not exclude any searches because short search durations are relatively rare in our data (only 15.5% of product page visits in our data are shorter than 15 seconds). Furthermore, the average duration of product page visits in single-product search sessions, which are more likely to be navigational, is similar to that in multiproduct search sessions (the share of visits below 15 seconds is 18.5% in single-product sessions and 14.3% in multiproduct sessions). We also note that the retailer's web page allows consumers to purchase without visiting the product page, for example, by clicking on the button "add to cart" that appears in the list of search results.

[25] More than 80% of searches occur within one week before a purchase. When consumers make two purchases less than seven days apart, we attribute only searches that occurred between the two purchases to the later purchase.

[26] More specifically, all utility parameters are rescaled by $1/\sigma_\mu^2$, whereas search costs are rescaled by $1/\sigma_\varepsilon^2$. We cannot, therefore, monetize the search cost estimate, but we can monetize all other parameters that enter utility.

[27] 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.

[28] If the consumer did not search a given product in the previous time period, the consumer does not know the realization of ω_{ijt-1}; therefore, from the consumer's perspective, $E(\varepsilon_{ijt}) = 0$ and $Var(\varepsilon_{ijt}) = Var(\omega_{ijt}) + \rho^2 Var(\omega_{ijt-1}) = \sigma_\omega^2 \cdot (1 + \rho^2)$. On the other hand, if the consumer does know the realization of ω_{ijt-1}, it follows that $E(\varepsilon_{ijt}|\omega_{ijt-1}) = \rho\omega_{ijt-1}$ and $Var(\varepsilon_{ijt}|\omega_{ijt-1}) = Var(\omega_{ijt}) + \rho^2 \cdot 0 = \sigma_\omega^2$.

[29] We assume that consumers do not anticipate the effects of their current search decisions on their information set in future search sessions. Given that search sessions in our data tend to be relatively spread out in time, we consider this assumption to be reasonable.

[30] We obtain the estimate of $\hat{\rho} = 0.3$ using grid search (see Appendix F for details) but do not compute a standard error for this estimate. We proceed in this way because the model with correlated information is computationally burdensome to estimate, which prevents us from using a bootstrap procedure to calculate standard errors. For computational reasons, we also constrain search costs not to vary across products when estimating this model extension.

[31] Several authors have emphasized that higher dispersion of consumer preferences often leads to less elastic demand (Anderson and Renault 1999, Zhou 2017).

[32] Our estimation approach recovers the parameters that govern the distribution of consumer types but does not explicitly recover the preferences and search costs of individual consumers. Therefore, we require this additional step to infer individual-level parameters based on consumers' search and purchase histories.

[33] 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 $\hat{mc}_5 = 540.2$.

[34] In the sequential search model, assuming that taste shocks ε_{jt} have identical variances across products implies that products'

reservation utilities are ordered in the same way as their presearch utilities ξ_j (Weitzman 1979). Similarly, in the simultaneous search model, the assumption of identical variances implies that the utility distributions of different products are ordered in the sense of first-order stochastic dominance. The consumer then finds it optimal to expand the search set by adding products in the order of descending mean utilities, which coincides with the order of descending presearch utilities ξ_j (Chade and Smith 2006).

[35] This follows from the purchase share expression $Pr_j(c_1) = Pr(\xi_j + \varepsilon_{jt} \geq \xi_l + \varepsilon_{lt} \quad \forall l \in \{1,\dots,k_1\})$ for $j \leq k_1$ and $Pr_j(c_1) = 0$ for $j > k_1$.

[36] A discontinuity in the bias occurs at $\xi_B = 0$ because as ξ_B switches from negative to positive, the consumer changes the search order to searching product B first.

[37] We use the same bootstrap estimates that we obtained earlier to compute standard errors of parameter estimates.

References

Ackerberg DA (2009) A new use of importance sampling to reduce computational burden in simulation estimation. *Quant. Marketing Econom.* 7(4):343–376.

Allenby GM, Rossi PE (1999) Marketing models of consumer heterogeneity. *J. Econometrics* 89(1):57–78.

Anderson SP, Renault R (1999) Pricing, product diversity, and search costs: A Bertrand-Chamberlin-Diamond model. *RAND J. Econom.* 30(4):719–735.

Armstrong M (2017) Ordered consumer search. *J. Eur. Econom. Assoc.* 15(5):989–1024.

Chade H, Smith L (2006) Simultaneous search. *Econometrica* 74(5): 1293–1307.

Chen Y, Yao S (2017) Sequential search with refinement: Model and application with click-stream data. *Management Sci.* 635(12): 4345–4365.

Chintagunta PK, Jain DC, Vilcassim NJ (1991) Investigating heterogeneity in brand preferences in logit models for panel data. *J. Marketing Res.* 28(4):417–428.

Choi M, Dai AY, Kim K (2018) Consumer search and price competition. *Econometrica* 86(4):1257–1281.

De Los Santos BI, Hortacsu A, Wildenbeest M (2012) Testing models of consumer search using data on web browsing and purchasing behavior. *Amer. Econom. Rev.* 102(6):2955–2980.

Dubé J-P, Hitsch GJ, Rossi PE (2009) Do switching costs make markets less competitive? *J. Marketing Res.* 46(4):435–445.

Dubé J-P, Hitsch GJ, Rossi PE (2010) State dependence and alternative explanations for consumer inertia. *RAND J. Econom.* 41(3): 417–445.

Dubé J-P, Hitsch GJ, Rossi PE, Vitorino MA (2008) Category pricing with state-dependent utility. *Marketing Sci.* 27(3):417–429.

Elberg A, Gardete PM, Macera R, Noton C (2019) Dynamic effects of price promotions: Field evidence, consumer search, and supply-side implications. *Quant. Marketing Econom.* 17(1): 1–58.

Giulietti M, Waterson M, Wildenbeest M (2014) Estimation of search frictions in the British electricity market. *J. Indust. Econom.* 62(4): 555–590.

Honka E (2014) Quantifying search and switching costs in the US auto insurance industry. *RAND J. Econom.* 45(4):847–884.

Honka E, Chintagunta PK (2017) Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Sci.* 36(1): 21–42.

Hotz VJ, Miller RA (1993) Conditional choice probabilities and the estimation of dynamic models. *Rev. Econom. Stud.* 60(3): 497–529.

Kim JB, Albuquerque P, Bronnenberg BJ (2010) Online demand under limited consumer search. *Marketing Sci.* 29(6):1001–1023.

Koulayev S (2014) Search for differentiated products: Identification and estimation. *RAND J. Econom.* 45(3):553–575.

Rossi PE, Allenby GM (1993) A Bayesian approach to estimating household parameters. *J. Marketing Res.* 30(2):171–182.

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4): 321–340.

Seiler S (2013) The impact of search costs on consumer behavior: A dynamic approach. *Quant. Marketing Econom.* 11(2):155–203.

Ursu RM (2018) The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Sci.* 37(4):530–552.

Weitzman ML (1979) Optimal search for the best alternative. *Econometrica* 47(3):641–654.

Yavorsky D, Honka E, Chen K (2021) Consumer search in the US auto industry: The role of dealership visits. *Quant. Marketing Econom.* 19(1):1–52.

Zhou J (2017) Competitive bundling. *Econometrica* 85(1):145–172.