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

MANAGEMENT SCIENCE
Vol. 63, No. 11, November 2017, pp. 3911–3929
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The Probit Choice Model Under Sequential Search with an Application to Online Retailing

Jun B. Kim,[a] **Paulo Albuquerque,**[b] **Bart J. Bronnenberg**[c,d]

[a] Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; [b] INSEAD, Fontainebleau 77305, France; [c] Tilburg University, 5037 AB Tilburg, Netherlands; [d] Centre for Economic Policy Research, London EC1V 0DX, United Kingdom

Contact: junkim@ust.hk (JBK); paulo.albuquerque@insead.edu (PA); bart.bronnenberg@uvt.nl (BJB)

Received: July 21, 2014
Revised: April 26, 2015; December 25, 2015
Accepted: February 8, 2016
Published Online in Articles in Advance: October 21, 2016

https://doi.org/10.1287/mnsc.2016.2545

Abstract. We develop a probit choice model under optimal sequential search and apply it to the study of aggregate demand of consumer durable goods. In our joint model of search and choice, we derive an expression for the probability of choice that obeys the full set of restrictions imposed by optimal sequential search. Estimation of our partially analytic model avoids the computation of high-dimensional integrations in the evaluation of choice probabilities, which is of particular benefit when search sets are large. We demonstrate the advantages of our approach in data experiments and apply the model to aggregate search and choice data from the camcorder product category at Amazon.com. We show that the joint use of search and choice data provides better performance in terms of inferences and predictions than using search data alone and leads to realistic estimates of consumer substitution patterns.

History: Accepted by Pradeep Chintagunta, marketing.
Funding: B. Bronnenberg gratefully acknowledges European Union funding from the Marie Curie Program [IRG 230962] and the Netherlands Foundation for Scientific Research [NWO Vici Grant].
Supplemental Material: Data are available at https://doi.org/10.1287/mnsc.2016.2545.

Keywords: optimal sequential search • discrete choice • consumer heterogeneity • aggregate demand models • information economics • market structure

1. Introduction

Online retailers routinely collect and publish data on consumer choice behavior and frequently display additional details about browsing behavior of shoppers at their online stores. Data on product browsing patterns (e.g., in the form of "consumers that viewed *item j* also viewed *items j'*") are available at popular online retailers such as Amazon.com, Target.com, Staples.com, and Kmart.com. Data on purchases in terms of sales rank and purchases conditional on browsing specific products (e.g., in the form of "consumers that viewed *item j* ultimately purchased *items j'*") are shown at Amazon.com and Walmart.com. This availability of online consumer data provides new ways for marketers to better understand consumer decisions in a variety of product categories.[1]

With these publicly available aggregate browsing and choice data in mind, the present study sets out to analyze consumer choice and pre-choice browsing behaviors in a single unified framework. We propose a theory-based empirical model that fully characterizes consumer optimal sequential search and choice decisions in a costly search environment. We apply this model to the study of highly differentiated consumer durable products, in which consumer valuations are often complex.

An important methodological challenge of considering both sequential search and choice decisions is that including optimal sequential search into a model of choice imposes constraints on the utilities for searched products.[2] Under such constraints, the evaluation of choice probabilities is complex and typically calls for numerical simulation to estimate the parameters of interest (e.g., Honka and Chintagunta 2017). However, simulation may become impractical when the number of searched items becomes large and the number of search paths grows. At the core of the proposed model is the choice probability expression that only involves univariate integration. The proposed model thus leads to both statistical and computational gains and broadens the applicability of models that simultaneously consider search and choice.

In addition, by combining aggregate search and choice data, we achieve a better identification of consumer preferences and search costs in the context of aggregate demand models. The empirical identification of consumer heterogeneity has been a challenge with aggregate choice data alone. To better pinpoint preferences, researchers have traditionally augmented such data with information such as second-choice surveys (Berry et al. 2004), consumer awareness data

(Draganska and Klapper 2011), aggregate-level consumer switching rates (Albuquerque and Bronnenberg 2009), advertising expenditures (Goeree 2008), and consumer demographic information (Petrin 2002). The use of extra data is motivated by the notion that variation in the choice set, either over time or across markets, helps model identification (Berry et al. 2004, p. 90). We continue this tradition and use the notion that variation in endogenously formed search sets across consumers is informative about heterogeneous consumer tastes. For example, consumers searching almost exclusively for camcorders with hard drive storage reveals a preference for this attribute. Alternatively, if consumers search for some products with hard drive storage and for other products with DVD storage, that signals more diffuse preferences for storage type. This premise is empirically supported by the existence of a large degree of heterogeneity among consumer search sets (Bronnenberg et al. 2016).

Substantively, this paper seeks to contribute to the empirical search literature in several ways. First, comparing our work to a related paper by Kim et al. (2010), we model both search and choice decisions, whereas the former models search decisions only. Among other advantages, one key improvement resulting from the joint use of search and choice data is that we can explicitly model consumers who search but do not buy from the category. Demand estimates solely based on search data can be biased and may lead to poor predictions and misleading inferences. Moreover, the use of both data facilitates the better identification of search cost parameters, which may be challenging using search data alone. The identification of product search costs in our model is partially based on contrasting search and choice data.

Second, recent papers in a related literature (e.g., Chen and Yao 2016, Ghose et al. 2013) study the identification of search costs in the context of heterogeneous goods and characterize the composition of the optimal search set and, in some cases (Honka and Chintagunta 2017, Koulayev 2014), include both the sequence and composition of consumer search, as we model in this paper. In contrast to these works that use high-dimensional integrations in the model estimation, our modeling approach reduces the multidimensional integration into a unidimensional one and leads to partial-simulation method that is scalable to larger choice sets. The latter makes it applicable to many durable goods categories where large choice sets are the norm and not the exception (Bronnenberg et al. 2016). Many researchers (e.g., Train 2009) advocate for the adoption of such methods whenever possible, as a result of higher accuracy and lower computational cost.

Finally, our empirical model adds to literature on aggregate demand models. Unlike previous approaches in which researchers assume exogenous variations in the choice set (e.g., Goeree 2008), changes in the choice sets in our model are an outcome of an endogenous consumer search process. Moraga-González et al. (2015) also offer a model of endogenous search sets but in a nonsequential search framework, compared with our sequential search approach. In addition, they rely on consumer location and consumer survey on their search behaviors[3] as the basis for modeling consumer search costs, whereas we use the observed consumer search behavior for this purpose directly. After extensive data experiments that demonstrate the statistical and computational advantages of the proposed model, we apply our model to search and choice data from Amazon.com, estimate consumer demand, and study substitution patterns and market structure in the camcorder industry. The wide availability of similar data from many online sources makes our demand model applicable to a variety of product categories in consumer durable goods.

The rest of the paper is organized as follows. In Section 2, we propose the unified model of search, choice, and choice conditional on search, in a setting of optimal sequential search. Section 3 discusses the data used in our empirical application and presents the identification and estimation approaches. In Section 4, we compare the performance of the proposed method with full simulation-based methods and aggregate search models in terms of statistical accuracy and computational cost. Section 5 discusses the estimation results and empirical application. We conclude in Section 6.

2. A Probit Model of Sequential Search and Choice

2.1. Overview

Before presenting the proposed model, we briefly outline the data we have in mind for our empirical application, our overall approach to modeling such data, and how our model components are used to match them. Our application data are aggregate-level data that summarize various aspects of consumer browsing and choice behaviors.[4] In particular, for our empirical application, we have aggregate-level search data, choice data, and choice data conditional on search. Search and choice data are collectively more informative of consumer preferences compared with choice data alone since they reflect multiple decisions of identical consumers in their shopping process. In terms of modeling and estimation, we follow the tradition in choice-based aggregate demand models (e.g., Berry et al. 1995). That is, we model the individual-level optimal search and choice behaviors, aggregate these individual-level predictions over the heterogeneous consumers, and match them to the observed aggregate-level data. In our model, we provide a set of coherent propositions that model various consumer decisions of search and choice all subject to optimal search. These propositions can be independently or jointly used to

model different sets of consumer data. We note that although we derive partially analytic expressions for search and choice probabilities for a consumer, we still need to integrate over heterogeneous consumers to generate the aggregate-level predictions during the estimation.

2.2. Setup

We model consumer search and choice decisions. In our setting, we assume that consumers sequentially keep searching, or browsing online, for information about products as long as the expected marginal benefit of doing so is greater than the marginal cost of search. Upon termination of search, the consumer chooses the highest utility product among the searched products.

The utility of product $l = 1, \ldots, J$ for consumer i is

$$u_{il} = V_{il} + e_{il}, \tag{1}$$

with

$$\begin{aligned} V_{il} &= X_l b_i, \\ b_i &\sim \mathrm{N}(b, B), \\ e_{il} &\sim \mathrm{N}(0, \sigma_{il}^2), \end{aligned}$$

where X_l is a row vector of product characteristics, b_i is a vector that represents individual-specific tastes for product characteristics, and the variance matrix B is assumed to be diagonal.

We interpret the search process as consumer effort to obtain the full match value of option l. Prior to search, we assume that consumers know the expected match value, V_{il}, but that the realization of the exact match value is subject to a shock, e_{il}, drawn from a known distribution. To fully resolve the unknown match value of $V_{il} + e_{il}$, consumers engage in costly search.[5]

In our empirical setting, the values of important attributes, which are captured in V_{il}, are readily accessible by consumers at zero search cost prior to searching for l. Given V_{il}, the goal of search is to resolve the remaining unknown value of e_{il} by incurring a search cost of c_{il}.[6] Upon search, consumers have access to more details about the option and receive the realized value of e_{il}. A large positive value of e_{il} obtained after search means that consumers found a good match between the product and their idiosyncratic preferences.

Finally, the utility of outside good is represented as

$$u_{i0} \sim \mathrm{N}(V_0, \sigma_0^2), \tag{2}$$

and we assume that consumer i knows the value of u_{i0} free of any search cost.[7] The demand primitives in our model are the consumer utility and search cost parameters. Next, we present the three main components from the joint model of optimal sequential search and choice: (1) optimal sequential search, (2) choice under optimal sequential search, and (3) choice conditional on search of a product. The first component is very close to Kim et al. (2010), whereas components 2 and 3 are new and constitute the main model development in this paper. All three model components are used together to model aggregate search and choice data in our empirical application.

2.3. Optimal Sequential Search

For the search part of our model, we use the theoretical framework from Weitzman (1979) and an empirical model of optimal sequential search similar to the one proposed in Kim et al. (2010). We refer to the latter if necessary to avoid repetition here.

Define u^* as the highest utility among the searched products thus far. Conditional on u^*, a consumer's expected marginal benefit from search of product l is[8]

$$\mathscr{B}_l(u^*) = \int_{u^*}^{\infty} (u_l - u^*) f(u_l)\, du_l, \tag{3}$$

where $f(\cdot)$ is the probability density distribution of u_l. Intuitively, $\mathscr{B}_l(u^*)$ captures the expected utility increment from alternative l over the utility u^* in hand.

When the stochastic components of the utility are uncorrelated across alternatives, Weitzman (1979) proves that the optimal sequential search decision of a consumer relies on her *reservation utility* of search. The reservation utility, which we denote by z_l, is the hypothetical utility that makes the consumer indifferent between searching and not searching option l. Mathematically, it is defined by

$$\mathscr{B}_l(z_l) = c_l, \tag{4}$$

where c_l is the search cost for option l. Prior to search, each consumer computes her reservation utilities for all options. Armed with these reservation utilities, the consumer engages in a three-stage search and choice process. First, she searches products in the order of descending reservation utility (selection rule). Second, search stops when the highest utility obtained thus far, u^*, is greater than the highest reservation utility among the items not yet searched (stopping rule). Finally, the product with the highest utility within the searched set is chosen (choice rule).

Because the rank of the reservation utility, $r(l \mid \theta)$, is a one-to-one mapping with product index l, we cast the model using l as the order of reservation utilities.[9] The following result holds for the probability to search.

Proposition 1. *Rank products on reservation utility. The probability that the option with the kth highest reservation utility is searched is equal to*

$$\begin{aligned} \pi_k &= \Pr\Big(\max_{l=1}^{k-1}(V_l + e_l) < z_k\Big) \\ &= \prod_{l=1}^{k-1} \Phi_l(z_k - V_l), \quad k > 1, \end{aligned} \tag{5}$$

where $\Phi_l(\cdot)$ is the cumulative distribution function (CDF) of the random error term e_l.

Proof. This follows immediately from the selection rule in optimal sequential search, in which consumers rank order the alternatives by their reservation utility to decide which option to search next, as well as from Kim et al. (2010), who show that the probability of inclusion for option k is equal to the probability that the first $k-1$ draws of utilities all fall short of z_k. □

2.4. Choice Under Sequential Search Constraints

To model choice, we derive the unconditional probability of choice subject to optimal sequential search. Proposition 2 shows that under optimal sequential search, the choice model does not suffer from the curse of dimensionality. Afterward, Proposition 3 describes the joint probability of search and choice under sequential search.[10]

Proposition 2. *Rank products on reservation utility. The probability that the jth ranked product is chosen equals*

$$\Pr(j) = \sum_{K=j}^{J} \Pr(j, S_K), \tag{6}$$

where $\Pr(j, S_K)$ denotes the joint probability that the jth ranked product is chosen from S_K, and $S_K = [1, \ldots, K]$ is an ordered set such that if $z_k \geq z_l$, then $1 \leq k < l \leq K$.

Proof. This follows directly from the application of the selection rule in which alternatives are sorted in descending order by reservation utilities. If the consumer has a set of unique reservation utilities, then only one sequence exists that is optimal (Weitzman 1979). Furthermore, the selection rule states that consumers search in the order of decreasing reservation utilities. After alternatives are ranked by descending reservation values, the superset of all possible optimal search sets consists of only J member sets, each containing the $K = 1, \ldots, J$ products with the highest reservation values.[11] □

The intuition behind this proposition is that to obtain the probability of j being chosen out of all possible search sets, the aggregation across search sets in Equation (6) is not over all 2^{J-1} possible permutations that contain a particular product.[12] Instead, with the application of selection rule under optimal sequential search, in which consumers initially order all products in descending order of reservation utilities and must follow that optimal "search path," the choice probability is given by the sum over at most J sets, depending on the ranking of chosen option j. This dramatically reduces the number of sets to be evaluated in the unconditional choice probability computation.

Proposition 3. *The joint probability $\Pr(j, S_K)$ that S_K is the ordered set and j is chosen contains two parts, with the second part relevant only if j is the last searched item, i.e., $j = K$, and equals*

$$\Pr(j, S_K) = \int_{z_{K+1}-V_j}^{z_K - V_j} \prod_{l \neq j}^{K} \Phi_l(V_j - V_l + e_j)\phi_j(e_j)\, de_j + I(j = K)(1 - \Phi_j(z_j - V_j))\pi_j, \tag{7}$$

where $I(j = K)$ is an indicator variable that is equal to 1 if $j = K$ and 0 otherwise. For completeness, (1) because consumers cannot purchase products that were not searched, $\Pr(j, S_K) = 0$ when $j > K$, and (2) because the consumer cannot search more than J alternatives, $z_{J+1} = -\infty$.

Proof. To develop the proof, we enumerate all the restrictions that the choice rule, the selection rule, and the stopping rule place on utility of the options in the ordered S_K.

1. The choice rule implies that the chosen option $j = \arg\max_{l=1,\ldots,K}\{u_l\}$; that is, $V_j + e_j > V_l + e_l$, $l \neq j$, $l = 1, \ldots, K$. With independent and normally distributed e_l, $\Pr(e_l < V_j - V_l + e_j \mid e_j) = \Phi_l(V_j - V_l + e_j)$ for $l \neq j$. The conditional probability that j has the highest utility within the ordered set of S_K is the product of these probabilities; $\prod_{l \neq j}^{K} \Phi_l(V_j - V_l + e_j)$.

2. Recall that consumers sort alternatives in descending order by reservation utilities. By the application of selection and stopping rules at K, a decision to search for option K implies $\max\{u_1, \ldots, u_{K-1}\} < z_K$; i.e., the maximum utility in hand after searching $\{1, \ldots, K-1\}$ is less than the highest reservation utility from the unsearched set $\{K, \ldots, J\}$.[13] This means that $V_l + e_l < z_K$ for $\forall\, l \in 1, \ldots, K-1$. However, it does not mean that $V_K + e_K < z_K$, and we need to condition in the derivation below on whether this inequality holds or not.

3. By the stopping rule, if search terminates at K, the utility draw of the chosen alternative denoted by j is greater than the reservation utility of $K+1$. Therefore, $u_j > z_{K+1}$; i.e., $e_j > z_{K+1} - V_j$.

If we ignore the selection and stopping rules in steps 2 and 3, the choice probability of j would be obtained by integrating out e_j from the choice rule for all $l \neq j$,

$$\Pr(j) = \int_{-\infty}^{\infty} \prod_{l \neq j}^{J} \Phi_l(V_j - V_l + e_j)\phi_j(e_j)\, de_j. \tag{8}$$

This model is a formulation of the probit model with independent error terms e.[14] We now impose the restrictions implied by the selection and stopping rules on these choice probabilities, which turn out to be simple integration limits in Equation (8).

Consider as a first case that we observe S_K and the final utility draw u_K is less than its own reservation

value, $u_K < z_K$. Having continued search until K, it must be true that $V_l + e_l < z_K$, $l = 1, \ldots, K$. At the same time, the choice of j implies that $V_l + e_l < V_j + e_j$ for $\forall\, l \neq j$. Note that this yields other $K-1$ restrictions on e_l, $l \neq j$. These two sets of restrictions are true when the most restrictive one is true. Observe that the utility for the chosen alternative $V_j + e_j$ is smaller than z_K. Also, the choice inequalities $V_l + e_l < V_j + e_j$ for $\forall\, l \neq j$ imply that $V_l + e_l < z_K$ for $\forall\, l \neq j$. Put differently, the choice restrictions imply the selection restrictions for all searched options l except for the chosen alternative j.

The only selection restriction that remains is that the utility draw on the chosen option j is low enough to continue search until K, $V_j + e_j < z_K$, or $e_j < z_K - V_j$. Thus, this adds an upper bound on e_j. From the stopping rule at option K, we also have that $e_j > z_{K+1} - V_j$, which implies a lower bound on e_j. In sum, the selection and stopping constraints on utility from sequential search translate only into additional lower and upper bounds on the utility shock of the chosen item e_j. Combining this with Equation (8) gives the following joint probability:

$$\Pr(j, S_K, u_K < z_K) = \int_{z_{K+1}-V_j}^{z_K - V_j} \prod_{l \neq j}^{K} \Phi_l(V_j - V_l + e_j)\phi_j(e_j)\, de_j \tag{9}$$

for $j \in 1, \ldots, K$ and $K \neq 1$ (when $K = 1$, $\Pr(j = 1, S_1) = \Pr(S_1) = 1 - \Phi_1(z_2 - V_1)$).

From a computational standpoint, Equation (9) is very similar to the probit in Equation (8) except with a lower bound on the distribution of unobservables from *termination* of search at K and an upper bound from *continuation* of search until K. To apply this equation to the case of $K = J$, it suffices to set $z_{K+1} = -\infty$.

We now continue with the second case, i.e., observing S_K, but now $u_K \geq z_K$. Combined with the conditions from the selection rule, $\max\{u_1, \ldots, u_{K-1}\} < z_K$, this case can only hold if the choice is K. Thus, when $u_K \geq z_K$ is true and S_K is the optimal search set, the choice of K occurs with probability 1. Consequently, the joint probability $\Pr(j = K, S_K, u_K \geq z_K)$ is equal to the joint probability $\Pr(S_K, u_K \geq z_K)$. This probability can be computed using the search probability in Equation (5). Conditional on $u_K \geq z_K$, search stops at K, and the probability that S_K is the optimal set is the same as the probability that K is included in the search set. This means that $\Pr(S_K \mid u_K \geq z_K) = \pi_K$. Therefore,

$$\Pr(K, S_K \mid u_K \geq z_K) = \pi_K. \tag{10}$$

Note that π_K does not depend on e_K. To obtain the unconditional probability, we use the condition that $u_K \geq z_K$ is equivalent to $e_K \geq z_K - V_K$, and we integrate the conditional probabilities (10) over e_K to obtain

$$\begin{aligned}\Pr(K, S_K, e_K \geq z_K - V_K) &= \int_{z_K - V_K}^{\infty} \pi_K \phi_K(e_K)\, de_K \\ &= \pi_K(1 - \Phi_K(z_K - V_K)).\end{aligned} \tag{11}$$

Combining Equations (9) and (11), we can write for $j \in 1, \ldots, K$

$$\begin{aligned}\Pr(j, S_K) = &\int_{z_{K+1}-V_j}^{z_K - V_j} \prod_{l \neq j}^{K} \Phi_l(V_j - V_l + e_j)\phi_j(e_j)\, de_j \\ &+ I(j = K) \cdot \pi_j(1 - \Phi_j(z_j - V_j)),\end{aligned} \tag{12}$$

which proves the proposition. With both parts of Equation (7), the computation of the joint probability involves only unidimensional integration. □

Proposition 3 develops a parsimonious expression for the summand in Equation (6), making use of both the assumptions of optimal sequential search and consequent ordering of alternatives by reservation utilities. Its major advantage is that even if the optimal ordered search set S_K is large, the joint probability $\Pr(j, S_K)$ requires a univariate expression instead of high-dimensional integration.[15]

2.5. Choice Conditional on Search

To model conditional choice under optimal sequential search, we express the choice probability of option j conditional on searching an option l. To avoid cluttered notation, we write this probability as $\Pr(j \mid l)$, $1 \leq j, l \leq J$.

Proposition 4. *Rank products on reservation utility. The probability that option j is chosen conditional on searching option l is equal to*

$$\Pr(j \mid l) = \frac{\sum_{K=\max(j,l)}^{J} \Pr(j, S_K)}{\pi_l}, \tag{13}$$

where K is the index of the option with smaller reservation utility between j and l,[16] π_l is the probability that lth option is searched (see Equation (5)), and $\Pr(j, S_K)$ is the probability that j is chosen and the optimal set is S_K (see Equation (7)).

Proof. We write the conditional choice probability as

$$\Pr(j \mid l) = \frac{\Pr(j, l)}{\Pr(l)}, \tag{14}$$

where $\Pr(l)$ is the probability that l is searched, and $\Pr(j, l)$ is the joint probability that l is searched and j is chosen. Note that the denominator, $\Pr(l)$, is equal to

the probability that l is in the optimal search set and is given by Equation (5), with

$$\Pr(l) = \pi_l. \tag{15}$$

Our approach for computing the joint probability of $\Pr(j,l)$ is to realize that both l and j must be in the optimal set. Given the optimal search sequence in the order of descending reservation utility, a necessary and sufficient condition for both options to be searched is that the option with the lower reservation utility, i.e., $K = \max(j,l)$, is searched. This means that the joint probability of search and choice is

$$\Pr(j,l) = \sum_{K=\max(j,l)}^{J} \Pr(j,S_K), \tag{16}$$

where the expression for the summand, $\Pr(j,S_K)$, is given in Equation (12). Finally, we obtain the conditional choice probability of $\Pr(j \mid l)$ by substituting Equations (15) and (16) into Equation (14) to obtain Equation (13):

$$\Pr(j \mid l) = \frac{\sum_{K=\max(j,l)}^{J} \Pr(j,S_K)}{\pi_l}, \tag{17}$$

which proves the proposition. □

The result in Proposition 4 relies on individual-level probabilities, which are already computed in Equation (3). Therefore, computing the conditional choice probability from our individual-level model adds no computational burden other than taking a sum over less than J terms. Equation (13) expresses that the probability that j is chosen given search of l is equal to this sum divided by the probability that l is searched.

2.6. Discussion

The four propositions in this section can each be used in isolation in estimation. For instance, Proposition 1 can be used to analyze search data as in Kim et al. (2010). Proposition 3 models search and choice behaviors of consumers, while Proposition 4 models choice decisions conditional on searching an option. The propositions can be used together to jointly model search and choice process as we do in our empirical application. Since we do not observe individual-level search sets in our aggregate-level empirical data, we enumerate all search sets and compute choice probabilities by utilizing additional Proposition 2.

If a researcher wants to model individual-level search sets, one can leverage Proposition 3. Note that Proposition 3 is conditional on presorting of reservation utilities. If researchers observe the search sequence, they can incorporate this information by computing the probability that the observed search sequence satisfies the condition of the descending reservation utilities.[17]

$$\Pr(j,S_K,O_J) = \Pr(j,S_K \mid O_J) \cdot \Pr(O_J),$$

where O_J is a particular realization of J alternatives with reservation utilities sorted in descending order. In our aggregate search data, since we do not observe the actual order of O_J, we invoke the selection rule from the optimal sequential search theory, presort options in the descending reservation values, and operate on the presorted set. That is, we theoretically impose O_J and set $\Pr(O_J) = 1$. However, when a researcher observes the actual individual search order in the data, one needs to separately compute $\Pr(O_J)$ while leveraging our Proposition 3 for $\Pr(j,S_K \mid O_J)$.

3. Data, Estimation, and Identification

3.1. Data

We apply our empirical model by combining three aggregate data sets from Amazon.com in the camcorder industry: view rank data, conditional share data, and sales rank data. The view rank data set is a list of products that are viewed by past consumers, conditional on viewing a focal product in the same browsing session. Consumers "view" a product if they request and visit the web page in which they can find the detailed information of the product. For example, if option B was viewed frequently with option A, option B will appear in A's view rank list. Amazon.com provides this view rank list for all products, and we refer to the set of all view rank lists as view rank data.[18]

The conditional share data consist of the choice shares of products in the category, conditional on viewing a focal product and on category incidence. That is, if option B is frequently chosen among consumers who viewed option A, B will appear on A's conditional share list with its numeric share value. Amazon.com provides these data for all focal products. We refer to the set of all conditional share lists as the conditional share or conditional choice data. Amazon.com's conditional share lists are often truncated; i.e., they only list the top four most popular choices conditional on searching for a focal product. However, their cumulative shares usually add up close to unity and hence provide rich information on conditional consumer choice behaviors.

Finally, Amazon.com publishes the sales rank of its products in the product detail page. The sales rank data are informative about the consumer choice at the online retailer.

For our application, we use camcorder search and choice data from June 2007. We extracted three sets of the aforementioned data and product characteristics for the 200 best-selling camcorders. After removing camcorders in the lowest-price tier and of professional grade, and limiting ourselves to the top six manufacturers and top four media formats, 90 camcorders remain for analysis. The summary statistics for these products are found in Table 1.

Table 1. Description of the Choice Options in the Empirical Data (Occurrence Frequency in Parentheses)

Attributes	Ranges
Brand	Sony (31), Panasonic (20), Canon (14), JVC (14), other (11)
Media type	MiniDV (34), DVD (30), flash memory (9), hard drive (17)
Price	\$532 (mean), \$258 (s.d.)
Form	Compact (8), conventional (82)
High definition	Yes (14), no (76)
Pixel	1.74 M (mean), 1.45 M (s.d.)
Zoom	20.1 (mean), 11.2 (s.d.)
Age (days)	266 (mean), 243 (s.d.)

Regarding the view rank data, products appear on average 26.3 (out of a possible 89) times on other product's view rank lists, with a standard deviation of 20.3. The mean and standard deviation of the conditional share data are 0.24 and 0.23, respectively. This means that conditional on searching a particular option, the market shares of the chosen options are highly concentrated in one or two options.

Next, we discuss the size of the outside option. Its share refers to the fraction of consumers who search but do not buy in the category. Our view rank data capture the search behaviors of all consumers regardless of whether they buy or not. However, the conditional shares and sales ranks reflect consumers who made a choice at Amazon.com. Therefore, we need to account for the outside good share in our empirical model. We use various online reports that indicate a conversion rate at Amazon.com ranging from 9% to 16% (Hancox 2008, Eisenberg 2009) and choose a conservative value of 10% in estimation, with 90% of searchers not buying a camcorder at Amazon.com.

3.2. Estimation

Our estimation approach is to use the equations derived in Section 2 to make individual-level predictions, construct their aggregate-level measures, and match them against the three collected data sets. Given a set of candidate parameters, we simulate individual-level optimal search and choice decisions for a set of heterogeneous pseudo-consumers—i.e., draws—from the assumed distributions, aggregate their search sets and choices, and compute the predictions from these aggregations. We then look for a parameter vector that maximizes the likelihood for the combined data set of view ranks, conditional choice shares, and sales ranks, subject to matching the fraction of consumers who browse but do not purchase.

3.2.1. Predicting View Rank Data.[19] For each product $j = 1, \ldots, J$, Amazon.com lists a set of other products that were viewed together in the form of a commonality index, CI_{jl}, defined as

$$CI_{jl} = \frac{n_{jl}}{\sqrt{n_j} \cdot \sqrt{n_l}}, \tag{18}$$

where n_j and n_l are the numbers of consumers who viewed products j and l, respectively, and n_{jl} denotes the number of consumers who viewed products j and l together. If $CI_{jl} > CI_{jk}$, l appears before k on j's view rank list. More specifically, the view rank indicator variable, $I^V_{j,lk}$, is defined as

$$I^V_{j,lk} = \begin{cases} 1 & \text{if } CI_{jl} > CI_{jk}, \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

where $j \neq l \neq k$. Using our model and candidate parameters, we forecast the commonality index between j and l as

$$CI_{jl} = \widehat{CI}_{jl} + \varepsilon^V_{jl} = \frac{\hat{n}_{jl}}{\sqrt{\hat{n}_j}\sqrt{\hat{n}_l}} + \varepsilon^V_{jl}, \tag{20}$$

where $\widehat{CI}_{jl}$ is the commonality index forecast and $\varepsilon^V_{jl} \overset{\text{iid}}{\sim} \mathrm{N}(0, (\tau_V^2/2))$. Furthermore, $\hat{n}_j = \sum_i \pi_{ij}$, and $\hat{n}_{jl} = \sum_i \min(\pi_{ij}, \pi_{il})$, $j, l = 1, \ldots, J$, and $j \neq l$.[20] The error term captures Amazon.com's potential measurement or aggregate-level prediction errors, similar to Bresnahan (1987) and Bajari et al. (2007).[21] The probability that product j is viewed more often with l than with k in the same search session is

$$\Pr(I^V_{j,lk} = 1) = \Pr(CI_{jk} < CI_{jl}) = \Pr\left(\varepsilon^V_{j,lk} < \widehat{CI}_{jl} - \widehat{CI}_{jk}\right), \tag{21}$$

where $\varepsilon^V_{j,lk} = \varepsilon^V_{jk} - \varepsilon^V_{jl}$ is a random variable with $\varepsilon^V_{j,lk} \sim \mathrm{N}(0, \tau_V^2)$. Hence,

$$\Pr(I^V_{j,lk} = 1) = \Phi\left(\frac{\widehat{CI}_{jl}(\Theta, X) - \widehat{CI}_{jk}(\Theta, X)}{\tau_V}\right), \tag{22}$$

where Θ are model parameters, X are data, and Φ is the CDF for the standard normal distribution. The variable $I^V_{j,lk}$ is directly observed in the view rank data from Amazon.com. For each product j, we observe at most $(1/2) \times (J-1) \times (J-2)$ unique inequalities defined by Equation (19). This leads to a large amount of restrictions on aggregate viewing in the data. In particular, across $J = 90$ products, the total number of observed pairwise ranks is 176,825. The probability of observing the set of all indicator variables, I^V, is

$$\begin{aligned} &\Pr(I^V = 1 \mid \Theta, X) \\ &\quad = \prod_j^J \prod_{l \neq j} \prod_{k \neq l \neq j} \Pr(I^V_{j,lk} = 1 \mid \Theta, X) \\ &\quad = \prod_j^J \prod_{l \neq j} \prod_{k \neq l \neq j} \Phi\left(\frac{\widehat{CI}_{jl}(\Theta, X) - \widehat{CI}_{jk}(\Theta, X)}{\tau_V}\right). \end{aligned} \tag{23}$$

3.2.2. Predicting Sales Rank Data. We represent sales rank data as the aggregate outcome of consumer choices at Amazon.com. We link the (observed) sales ranks to the (unobserved) market shares as follows:

$$I^S_{jl} = \begin{cases} 1 & \text{if } s_j > s_l, \\ 0 & \text{otherwise,} \end{cases} \tag{24}$$

where s_j is j's unobserved true market share. Given X and Θ in our joint model, we predict j's market share by aggregating individual choice probabilities for j (Equation (6)). The market share, s_j, is modeled as the sum of the prediction $\hat{s}_j$ and a measurement error term,

$$s_j = \hat{s}_j + \varepsilon^S_j, \tag{25}$$

where $\varepsilon^S_j \overset{\text{iid}}{\sim} \mathrm{N}(0, (\tau_S^2/2))$, $j \in \{1, \ldots, J\}$. The probability of observing a pairwise sales rank inequality between j and l is computed as

$$\Pr(I^S_{jl} = 1) = \Pr(s_l < s_j) = \Pr(\varepsilon^S_{jl} < \hat{s}_j - \hat{s}_l), \tag{26}$$

where $\varepsilon^S_{jl} = \varepsilon^S_l - \varepsilon^S_j$ is a random variable with $\varepsilon^S_{jl} \sim \mathrm{N}(0, \tau_S^2)$. Thus,

$$\Pr(I^S_{jl} = 1) = \Phi\left(\frac{\hat{s}_j(\Theta, X) - \hat{s}_l(\Theta, X)}{\tau_S}\right), \tag{27}$$

where Φ is the CDF of the standard normal distribution. The joint probability of observing the set of all sales rank indicator variables of I^S is

$$\begin{aligned} \Pr(I^S = 1 \mid \Theta, X) &= \prod_j^J \prod_{l \neq j}^J \Pr(I^S_{jl} = 1 \mid \Theta, X) \\ &= \prod_j^J \prod_{l \neq j}^J \Phi\left(\frac{\hat{s}_j(\Theta, X) - \hat{s}_l(\Theta, X)}{\tau_S}\right). \end{aligned} \tag{28}$$

3.2.3. Predicting Conditional Choice Share Data. Finally, we discuss the prediction of conditional choice shares. The observed choice share of j conditional on searching l, $s_{j|l}$, is modeled as

$$s_{j|l} = \hat{s}_{j|l}(\Theta, X) + \varepsilon^C_{jl}, \tag{29}$$

where $\hat{s}_{j|l}(\Theta, X)$ is the model prediction of conditional choice shares aggregated across consumers and $\varepsilon^C_{jl} \sim \mathrm{N}(0, \tau_C^2)$ is the measurement error. Note that the conditional share prediction is obtained by aggregating individual conditional choice probabilities in Equation (13) across draws. The probability of $\Pr(s_{j|l} \mid \Theta, X)$ is

$$\begin{aligned} \Pr(s_{j|l} \mid \Theta, X) &= \phi\left(\frac{\hat{s}_{j|l}(\Theta, X) - s_{j|l}}{\tau_C}\right) \\ &= \frac{1}{\tau_C \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{(\hat{s}_{j|l}(\Theta, X) - s_{j|l})^2}{2 \cdot \tau_C^2}\right), \end{aligned} \tag{30}$$

where ϕ is a probability density function for a standard normal distribution. Then the joint probability of observing the set of all conditional share values of $s = \{s_{j|l}\}$ is

$$\Pr(s \mid \Theta, X) = \prod_l^J \prod_j \Pr(s_{j|l} \mid \Theta, X), \tag{31}$$

where j indexes the options that appear in l's conditional share list.

3.2.4. Likelihood Function. Given the set of data of $Y = \{I^V, I^S, s\}$ and the model parameter vector of Θ, our likelihood function is

$$\mathcal{L}(\Theta \mid Y, X) = \Pr(Y \mid \Theta, X).$$

Assuming that the error terms are independent within and across sets, we decompose the likelihood function into the contributions by the view data, sales rank data, and conditional share data:

$$\begin{aligned} \Pr(Y \mid \Theta, X) = \Pr(I^V = 1 \mid \Theta, X) \cdot \Pr(I^S = 1 \mid \Theta, X) \\ \cdot \Pr(s \mid \Theta, X). \end{aligned} \tag{32}$$

The corresponding log-likelihood function is

$$\begin{aligned} \mathcal{LL}(\Theta \mid Y, X) = &\sum_j \sum_{l \neq j} \sum_{k \neq l} \log(\Pr(I^V_{j,lk} = 1 \mid \Theta, X)) \\ &+ \sum_j \sum_{l \neq j} \log(\Pr(I^S_{jl} = 1 \mid \Theta, X)) \\ &+ \sum_l \sum_j \log(\Pr(s_{j|l} \mid \Theta, X)), \end{aligned} \tag{33}$$

with the probabilities in the three summations defined in Equations (22), (27), and (30), respectively.[22]

3.2.5. Outside Goods. Let s_0 denote the observed share of consumers who search but do not buy in the category. We interpret the outside good share as an independent observation at the population level and hence impose its share as a constraint during the estimation. Given a set of parameters Θ, we can forecast the share of the outside goods $\hat{s}_0(\Theta, X)$. We impose that $\hat{s}_0(\Theta, X) = s_0$ in estimation. This makes our approach a constrained maximum likelihood approach. We implement this using the penalty method. First, we introduce squared difference between $\hat{s}_0(\Theta, X)$ and s_0 as a penalty:

$$g(\hat{s}_0 \mid \Theta, X, s_0) = -(\hat{s}_0(\Theta, X) - s_0)^2. \tag{34}$$

The penalty function g is continuous in Θ and reaches its maximum at the value of Θ that makes $\hat{s}_0(\Theta, X)$ equal to s_0. Next, we augment the likelihood function (33) with this penalty term using a weight w. Thus, the objective function maximized is $\mathcal{LL}(\Theta \mid Y, X) + w \cdot g(\hat{s}_0 \mid \Theta, X, s_0)$. In the application, we have used $w = 4 \times 10\text{E}7$, which leads to a very accurate fit with the observed outside good s_0.

To obtain standard errors of the parameter estimates, we use the bootstrap resampling method proposed in Efron and Tibshirani (1994).

3.3. Identification

In this section we discuss model identification. Our parameter set includes the mean utility and consumer heterogeneity parameters and the mean and product-specific search cost parameters. The presearch uncertainty variance σ_{ij}^2 and the outside goods utility variance σ_0^2 are fixed to 1 for identification purposes, common in choice models. Since our model is estimated using search and choice data, we condition our discussion on the availability of such data.

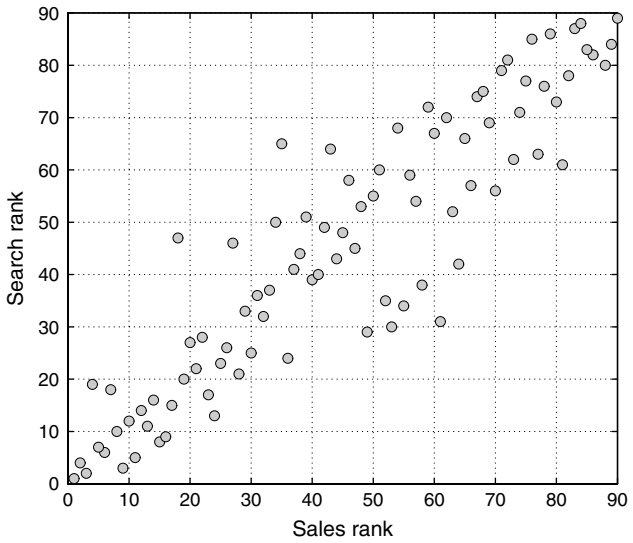
The average search and choice popularity of products—measured by view ranks and sales ranks—identifies the mean utility parameters. Under our joint model, the mean utility parameters are identified by the correlation between the variation in product popularity and the variation in product characteristics. Choice or search measures alone can identify mean utility parameters, but their joint use does so more efficiently.

For search cost-related parameters, the mean search cost is identified by the lengths of the view rank lists for each product. That is, if the search cost is low, the view rank lists will be long, and vice versa. Product-specific search costs are identified by the discrepancies between search and choice popularity. The intuition is that the product utility enters both search and choice decisions, while search costs enter search decisions only (see also, e.g., Ghose et al. 2013). For example, a high value product with average search cost will be popular during both search and choice stages, putatively resulting in similar search and choice shares. By contrast, a low value product with low search costs is popular during search but not in choice because the search costs are sunk at choice stage and value matters only. Conversely, options with high search cost will be less popular during search than during choice, compared with other products. This way, contrasting search and choice shares is informative about product-specific search costs.

As an illustration, Figure 1 displays search and choice ranks in our empirical data: each circle represents a product's sales rank on the x-axis and its search rank on the y axis.[23] In the figure, if an observation lies on the 45° line, its search and sales ranks are identical. If the observation is above (below) the 45° line, its sales rank is higher (lower) than the search rank, implying that the product is more (less) popular during choice than search.

Figure 1 shows a strong positive correlation between search and sales ranks, providing support to our premise that search and choice share common demand primitives and their joint use will lead to a more efficient parameter estimation. However, it also shows large differences between search and sales ranks for some products. One Panasonic DVD camcorder is ranked 61st in terms of sales but 31st in terms of search.

Figure 1. Scatter Plot of Sales Rank and Search Rank of Products



Under our model, this product is likely to have a low search cost that makes it more popular during search than choice. Not only do the view and sales ranks register differences between search and choice, the view rank and conditional share data do also. That is, if product l is popular in j's view rank list but is not so in j's conditional share list, we infer that l was searched more with j because of its lower search cost but was not chosen as a result of lower utility. Therefore, these gaps are informative about product-specific search costs.

The identification of consumer heterogeneity mainly comes from observed view rank lists being more homogeneous than the universal choice set. That is, the identification comes from the correlation between characteristics of a focal product and those of related options in its view rank list. Imagine options k_1 and k_2 to be low-priced, flash memory camcorders and options l_1 and l_2 high-priced, hard disk camcorders. If each consumer searches more within price tier and media format type than across, we infer that preferences for price and media format type must be different across consumers. The conditional share data help identify consumer heterogeneity further for the same reason.

4. Data Experiments

4.1. Overview

In this section, we first show that the proposed model can recover the model parameters well. Next, using numerous data experiments, we demonstrate the multiple advantages of the proposed model. First, we show that the proposed method has statistical and computational advantages compared with a frequency-based simulator. Second, we demonstrate the benefit of the proposed joint model of search and choice over aggregate search models; see, e.g., Kim et al. (2010).

4.2. Comparison with Frequency-Based Estimator

4.2.1. Setup. The utility of consumer i for product j is defined as

$$u_{ij} = V_{ij} + e_{ij} = X_j\beta_i - \alpha_i P_j + e_{ij}, \tag{35}$$

where X_j is a row vector of product j's characteristics and P_j is j's price; β_i is a column vector of individual-specific sensitivities for product characteristics, α_i is the price coefficient, and $\text{var}(e_{ij}) = \sigma^2$. We set $\sigma^2 = 1$ for identification purposes. We create $J = 32$ product options with five binary attributes X_j—arbitrarily named as brand (Sony or not), pixel (> 1 MB or not), zoom (> 10× or not), media type (MiniDV or not), and form (compact or not)—and a continuous attribute, price. Furthermore, we impose a theory-driven restriction on the price coefficient, assuming a log-normal distribution. Mathematically,

$$\begin{aligned} \log(\alpha_i) &\sim \text{N}(\beta_p, \sigma_p^2), \\ \beta_i &\sim \text{N}(\beta_0, \Sigma_\beta), \end{aligned} \tag{36}$$

where Σ_β is a diagonal matrix with each entry representing consumer heterogeneity for attribute j. We also generate product-specific search costs as a function of attributes:

$$c_j = \exp(\gamma_0 + \gamma_1 X_j^c), \tag{37}$$

where γ_0 is the base search cost parameter and γ_1 is search cost sensitivity to X_j^c. For our data experiment, we set $X_j^c = [ad_j, X_{j2}]$, where ad_j is the "advertising" for product j, which is modeled as an exogenous search cost shifter, and X_{j2} is the binary variable of pixel > 1 MB.[24] Therefore, we assume that this variable affects both consumer utility and search cost, and we test whether our joint model can identify the separate effects of a variable entering the utility and search cost.

To generate synthetic data, we assign values to the model parameters and draw 50,000 "pseudo-households" from the joint distribution of parameters. For each i, we compute V_{ij}, z_{ij}, and other relevant quantities and obtain individual optimal search set and choice. We then aggregate individual-level decisions across consumers according to Amazon.com's recipe and generate the commonality indices, conditional choice shares, and market shares. Finally, we add measurement errors to these quantities using Equations (20), (25), and (29), and we generate view rank, sales rank, and conditional share data.

4.2.2. Parameter Recovery. In estimation, we maximize the log-likelihood function (33) that includes the model's forecasts of sales ranks, view ranks, and conditional share (with a penalty term for the outside option share when applicable). To compute these forecasts, we fix $I = 500$ draws from the standard normal and log-normal distributions to draw from Equation (36). At a given set of parameter values, we use these draws to compute the expected utilities V_{ij}, reservation utilities z_{ij}, and search probabilities π_{ij} for each pseudo-household i and product j. The model's forecast of the view ranks is computed from aggregating the 500 values of $\{\pi_{ij}\}$, $i = 1, \ldots, I$, into the estimated commonality indices. The model's forecast of sales ranks and conditional shares is constructed by first computing i's choice probabilities using Equations (6) and (7). For the conditional choice probabilities, we use Equations (7) and (13). We next aggregate the individual choice probabilities across pseudo-households to forecast market shares, sales ranks, and conditional shares.

The parameter estimates and standard errors are shown in the column labeled "Proposed method" in Table 2. The standard errors are computed from repeated estimations over 12 different simulated data sets obtained from redrawing the sampling errors in Equations (20), (25), and (29). The table shows that the recovered parameters are close to their actual values and within sampling errors. We conclude that data similar to that used in our empirical analysis can identify the model parameters.[25]

4.2.3. Frequency-Based Simulator. In recent empirical models, researchers have numerically simulated the set of complex search restrictions on the utility event space to evaluate the joint probability of search and choice (e.g., Chen and Yao 2016, Ghose et al. 2013, Honka 2014, Honka and Chintagunta 2017). Although the past work with relatively smaller search sets is encouraging, it is less clear how the frequency simulator performs with larger search sets requiring high-dimensional integrations. Here, we demonstrate how our partial simulation-based method performs against full simulation-based methods.

For the frequency simulator, we adopt the kernel-smoothed frequency simulator (McFadden 1989) and follow the implementation recipes detailed in Honka (2014) and Honka and Chintagunta (2017), who used it in an empirical model of search and choice. The frequency estimator approximates the joint probability of $\Pr(j, S_K)$ in Equation (7), which can also be defined as

$$\begin{aligned} \Pr(j, S_K) = \Pr[&(u_{ij} > \max\{u_{i1}, \ldots, u_{ij-1}, u_{ij+1} \ldots, u_{iK}\}) \\ &\cap (z_{in+1} > \max\{u_{i1}, \ldots, u_{in}\}) \\ &\cap (u_{ij} > z_{iK+1})], \quad n = 1, \ldots, K-1. \end{aligned} \tag{38}$$

Once we simulate this joint probability, we can compute the choice probability in Equation (7) as well as the conditional choice probability in Equation (17). Since the original frequency-based estimator is discontinuous, we smooth the above probability using

$$\Pr(\omega; \lambda) = \frac{1}{1 + \exp(\sum_m - \lambda\omega_m)}, \tag{39}$$

with the scale factor $\lambda = 5$ and with the vector of ω defined below.

Table 2. Parameter Estimates from the Proposed Method and Kernel-Smoothed Frequency Simulator

			Estimated value (s.e.)		
				Frequency-based simulator	
Parameters		True value	Proposed method	$Q = 20$	$Q = 40$
Mean utility	Sony	0.5	0.56 (0.08)	0.58 (0.08)	0.60 (0.14)
	Pixel < 1 MB	−0.5	−0.53 (0.13)	−0.56 (0.15)	−0.55 (0.12)
	Zoom > 10×	0.5	0.53 (0.09)	0.56 (0.10)	0.56 (0.13)
	Media type: MiniDV	−0.5	−0.53 (0.16)	−0.65 (0.20)	−0.66 (0.19)
	Form: Compact	1	1.00 (0.11)	1.07 (0.20)	1.07 (0.23)
	Price	−1	−0.96 (0.20)	−0.83 (0.18)	−0.85 (0.13)
Heterogeneity	Sony	0.5	0.55 (0.23)	0.48 (0.28)	0.58 (0.35)
	Pixel <1 MB	1	1.11 (0.21)	1.22 (0.37)	1.12 (0.28)
	Zoom >10×	0.5	0.46 (0.16)	0.40 (0.24)	0.50 (0.28)
	Media type: MiniDV	1	1.03 (0.19)	1.15 (0.26)	1.17 (0.34)
	Form: Compact	0.5	0.55 (0.16)	0.51 (0.20)	0.49 (0.25)
	Price	0.5	0.61 (0.21)	0.77 (0.38)	0.63 (0.37)
Search cost	Base cost	−3	−3.17 (0.27)	−3.15 (0.32)	−3.13 (0.32)
	Advertising	−2	−1.84 (0.22)	−1.99 (0.33)	−2.02 (0.37)
	Pixel <1 MB	−1	−0.82 (0.19)	−1.18 (0.38)	−1.16 (0.34)
Aggregation error (s.d.)	τ_V	0.2	0.293 (0.04)	0.30 (0.03)	0.30 (0.02)
	τ_S	0.005	3E−4 (4E−3)	1E−3 (3E−3)	1E−3 (4E−3)
	τ_C	0.05	0.07 (0.01)	0.08 (0.01)	0.08 (5E−3)
Statistical performance	MAD		0.067	0.102	0.085
	Mean s.e.		0.147	0.206	0.213
Average computational metrics	Time for one objective function evaluation (sec.)		18	53	108
	Number of objective function evaluations		454	592	626
	Gross time for one estimation (min.)		138	528	1,122

- For each i, take $q = 1, \ldots, Q$ draws from the vector $\{e_i\}$.
- For each draw of e_i, compute the elements of the vector ω_i^q as

—(choice rule) $\omega_{i,k}^q = V_{ij} + e_{ij} - V_{ik} - e_{ik}$, $k = 1, \ldots, K$, $k \neq j$;

—(selection rule) $\omega_{i,n+K-1}^q = z_{in+1} - \max\{V_{i1} + e_{i1}, \ldots, V_{in} + e_{in}\}$, $n = 1, \ldots, K-1$; and

—(stopping rule) $\omega_{i,2K-1}^q = V_{ij} + e_{ij} - z_{iK+1}$.

- Evaluate the joint probability of search and choice for i using Equation (39):

$$\widehat{\Pr}_q(j, S_K) = \frac{1}{1 + \exp(\sum_{m=1}^{2K-1} -\lambda \cdot \omega_{im}^q)}.$$

- Integrate out the e_i by averaging over the Q draws of the search and choice probability:

$$\widehat{\Pr}(j, S_K) = \frac{1}{Q} \sum_{q=1}^{Q} \widehat{\Pr}_q(j, S_K).$$

The results for the kernel-smoothed frequency-based estimations are shown in the two right columns of Table 2, with varying numbers of draws Q used to integrate the joint probability in the integration. We compute the mean (across parameters) absolute distance (MAD) between the true vector of Θ^{true} and its estimate of $\hat{\Theta}$, as well as the average (across-parameters) standard error for $\hat{\Theta}$. While the MAD measures how well the estimates recover the location of true parameters, the average standard error tracks their efficiency, with better methods having lower values on both metrics. Finally, we compare computational costs in detail: we report the average time taken for one objective function evaluation, the number of objective function evaluations needed in one estimation, and the total time taken per estimation. Given that any optimization algorithm must evaluate the objective function repeatedly during the course of estimation, these metrics provide a comprehensive summary of computational costs.

From the table, we report that our partial simulation-based method outperforms the kernel-smoothed frequency simulator, as both MAD and average mean standard error are significantly smaller for different values of Q. At the same time, its computational cost is far lower than that of the kernel-smoothed frequency simulator.[26] This implies that the proposed method improves the accuracy and accelerates the estimation of large-scale joint models of consumer search and choice,[27] when compared with kernel-smoothed frequency simulators.

4.3. Comparison with Aggregate Search Models

In a related paper, Kim et al. (2010) use search data only and study consumer demand. In principle, the combined use of both search and choice data in our joint model helps achieve better model identification and inferences compared with an aggregate search model that does not accommodate choice data. To test this conjecture, we conduct additional data experiments and compare the proposed model's performance to those of aggregate search models in a setting similar to our empirical context. To that end, we use two versions of aggregate search models. The first version is the search model of Kim et al. (2010) augmented with outside goods share data, while our second version is the direct application of aggregate search model in Kim et al. (2010) without such augmentation. For the implementation of the former, we still need to estimate the joint model of search and choice. However, during the course of estimation, we use the moments from search data and outside goods share data only while ignoring the choice moments in the objective function.[28] The comparison of these two versions helps to better understand the incremental gain from modeling the sales ranks and conditional share data in our empirical setting.

For this set of experiments, we set the outside good utility as in Equation (2) and set the value of V_0 such that outside goods share is 89%. We also set σ_0^2 to 1 for identification purposes.[29] The overall setup, synthetic data generation, and parameter recovery part for the experiments are identical to those in Section 4.2, except the following. First, for synthetic data generation, we draw 100,000 pseudo-households from the joint distribution of parameters. Second, for parameter recovery, we take $I = 1{,}400$ draws from the distributions of random coefficients.[30]

Table 3 shows the parameter estimates from three different models. First, the column labeled "Proposed model" shows the parameter estimates of the proposed joint model. In particular, this column shows that all parameters, including the mean outside good utility V_0, are well recovered and are all statistically close to their actual values. Second, the other two columns show the parameter estimates of two aggregate search models based on Kim et al. (2010). The column "w/outside goods" shows the parameter estimates from the model of Kim et al. (2010) with outside share data in the estimation, while the last column shows the parameter estimates from applying the original model of Kim et al. (2010) to the data generated from a process with an outside good. Comparing the three models, our proposed model outperforms both versions of aggregate search models in terms of parameter recovery (MAD) and estimator efficiency (the average standard error). From these exercises, we conclude that the joint use of search and choice data helps achieve better parameter recovery and the incorporation of the outside goods share data substantially

Table 3. Comparison of Parameter Estimates from Different Models

			Estimated values (s.e.)		
				Aggregate search model	
Parameter		True values	Proposed model	w/outside goods	
Mean utility	Sony	0.50	0.46 (0.03)	0.42 (0.06)	0.59 (0.13)
	Pixel <1 MB	−0.50	−0.51 (0.05)	−0.53 (0.08)	−0.93 (0.35)
	Zoom >10×	0.50	0.50 (0.03)	0.38 (0.07)	0.59 (0.09)
	Media type: MiniDV	−0.50	−0.50 (0.06)	−0.42 (0.07)	0.06 (0.11)
	Form: Compact	1.00	0.97 (0.05)	0.86 (0.05)	1.13 (0.18)
	Price	−1.00	−1.03 (0.07)	−1.04 (0.05)	−1.73 (0.41)
Heterogeneity	Sony	0.50	0.51 (0.06)	0.43 (0.06)	0.50 (0.11)
	Pixel <1 MB	1.00	0.96 (0.04)	1.01 (0.06)	1.02 (0.15)
	Zoom > 10×	0.50	0.49 (0.02)	0.45 (0.05)	0.48 (0.06)
	Media type: MiniDV	1.00	1.07 (0.03)	1.08 (0.06)	1.14 (0.16)
	Form: Compact	0.50	0.53 (0.05)	0.53 (0.05)	0.47 (0.07)
	Price	0.50	0.44 (0.08)	0.49 (0.08)	0.59 (0.48)
Search cost	Base cost	−4.00	−4.14 (0.09)	−4.24 (0.16)	−2.95 (0.40)
	Advertising	−3.00	−2.96 (0.08)	−3.05 (0.14)	−2.41 (0.41)
	Pixel <1 MB	−1.00	−0.98 (0.09)	−1.10 (0.18)	−2.70 (0.47)
	Outside good V_0	4.10	4.15 (0.05)	4.00 (0.07)	NA
Aggregation	τ_V	0.05	0.07 (0.01)	0.08 (0.01)	0.08 (0.01)
Error	τ_S	0.002	4E−4 (5E−4)	NA	NA
s.d.	τ_C	0.002	0.03 (0.01)	NA	NA
Statistical	MAD		0.033	0.074	0.356
Performance	Mean s.e.		0.047	0.077	0.225

Note. Aggregate search model refers to Kim et al. (2010).

improves performance of aggregate search models in our empirical setting.

5. Empirical Illustration

5.1. Specification

In our empirical application, we represent a product as a bundle of characteristics and use the identical utility specification of Equations (35) and (36) in Section 4.2. We include eight product characteristics: brand name, media format, form factor, high definition, zoom, number of pixels, and price (in thousands of dollars).[31] We also assume that the stochastic utility term is a normally distributed random error, with mean 0 and variance σ^2, and that this error term is identical and independent across i and j.[32] For reasons of parsimony, we use one common heterogeneity parameter for all brands, as well as one for all media formats. We set the outside good utility as in Equation (2) and fix one of the brand intercepts to zero and estimate the outside good mean utility, in part to measure changes in the intercept of the outside good with different assumptions of the outside good size. We also normalize all variance terms of σ^2 and σ_0^2 to 1 for identification purpose.

In addition, we specify j's search cost as in Equation (37) in our data experiment. In our empirical application, the row vector of X_j^c is defined as

$$X_j^c = [I_j^{\mathrm{Hi}}, \log(A_j)],$$

where I_j^{Hi} is an indicator variable if product j is a high-definition alternative, and A_j is the age of product j.[33] Newer and more advanced products, such as high-definition products, are featured more prominently by retailers, potentially leading to lower search costs.[34] Conversely, older products are no longer featured and become more difficult to find.

Finally, with respect to the possibility of price endogeneity in our utility specification, we allow for very flexible product fixed effects, which in principle reduces the potential for unobservables that might be related to prices.[35] However, we acknowledge that our model may be still susceptible to remaining correlation between unobservable demand shocks and price, and that our price coefficient should be interpreted with this taken into account.

For estimation, we use 2,000 draws of pseudo-consumers from the joint distribution of random coefficients, and for each consumer draw i, we compute reservation utilities $\{z_{ij}\}$ and search, choice, and conditional choice probabilities using our propositions in Section 2. We then aggregate the computed quantities across consumers following the recipe in Section 3.2 and estimate model parameters.

5.2. Model Fit and Estimates

For internal validation, we investigate how well the proposed model predicts the search and sales data patterns. We report that the hit rates of pairwise rank inequalities, in which we compare the relative positions of two options in the actual and predicted rank data, are 77% for sales rank and 86% for view rank data. As an external validation, we use out-of-sample data from September 2007 and compute similar measures. Across $J = 86$ products, the pairwise hit rates are 73% for sales rank and 82% for the view rank data. We conclude that our model matches the search and sales patterns in- and out-of-sample well.

Next, we compare our model fit against that obtained when using view rank data plus the outside good constraint alone. This model is a version of Kim et al. (2010) that accounts for the outside good share. Without the sales rank and conditional share data, the in-sample sales rank hit rate is 72%, which is marginally lower than our model predictions of 77%—evidence that using more data improves the sales prediction with respect to the aggregate search approach of Kim et al. (2010). The view rank hit rate is comparable at 85%.[36] As a robustness check, we reestimated our empirical model with the outside goods share values of 86% and 92%.[37] We report that, besides a shift in the mean utility of outside option of V_0, the remaining parameter estimates are quite similar and show very high correlations (0.999). Therefore, our parameter estimates seem robust to small changes in values in outside good shares.

5.2.1. Parameter Estimates. We present the parameter estimates in Table 4. We find that the brand intercepts have face validity: Sony, one of the best-known brands during our data collection period, exhibits the highest mean brand coefficient of 1.96, while Panasonic, another popular brand, has the second-highest mean value of 1.69. The estimates show significant heterogeneity in brand preferences with an estimate of 0.58 for the standard deviation of its distribution. In terms of other product characteristics, the hard drive media option is the most preferred, with a coefficient normalized at zero compared with the negative coefficients of other media options: MiniDV is the next preferred option (−0.69), and flash memory (−1.63) is the least preferred. Heterogeneity for media formats is also high, with its estimate of standard deviation at 0.71. Compactness negatively influences the mean utility with a large heterogeneity, and consumers prefer higher number of pixels (0.21). Finally, we report that our price coefficient estimates imply an average own-price elasticity of −1.88.

At the time of our analysis, high-definition (HD) products were still at an early stage of their product life cycle, which may explain its negative mean utility coefficient. In addition, high prices of HD TVs may have also played a role because a HD TV is required to take a full advantage of the high-definition feature of camcorders. In December of 2007, the same year of our

Table 4. Estimates of the Model Parameters for the Camcorder Category

	Variable	Mean (s.e.)		Heterogeneity (s.e.)	
Utility	*Sony*	1.96	(0.10)	0.58	(0.02)[a]
	Panasonic	1.69	(0.11)	0.58	(0.02)
	Canon	1.37	(0.13)	0.58	(0.02)
	JVC	1.25	(0.10)	0.58	(0.02)
	Samsung	0.95	(0.07)	0.58	(0.02)
	Media type: MiniDV	−0.69	(0.08)	0.71	(0.02)[b]
	Media type: DVD	−1.26	(0.06)	0.71	(0.02)
	Media type: Flash memory	−1.63	(0.22)	0.71	(0.02)
	Compact	−0.82	(0.20)	1.20	(0.14)
	High definition	−0.72	(0.14)	1.05	(0.07)
	Zoom	−0.01	(2E−3)	0.02	(2E−5)
	Pixel	0.21	(0.03)	0.21	(0.03)
	log(*Price*)	1.12	(0.08)	0.89	(0.04)
	V_0 (*outside good*)	4.81	(0.18)		
Search cost	*Base search cost* (γ_0)	−9.22	(0.32)		
	High definition	−5.06	(0.37)		
	Age	0.96	(0.06)		
Aggregation Error s.d.	*View rank*	0.13	(0.004)		
	Sales rank	0.002	(0.01)		
	Conditional share	0.40	(0.04)		
	Log-likelihood	−63,475			

[a]Random effects variance is common across brands.
[b]Random effects variance is common across media formats.

data, the average price of a 40-inch LCD HD TV was about \$1,500 (Magid 2007).[38] Heterogeneity in preferences for high definition is especially high compared with other attributes, which is justified by the innovative nature of this attribute.

5.2.2. Inferences. In terms of the search cost, high-definition products have a lower search cost, which is likely driven by its innovative nature, leading these products to stand out more than traditional products. Across products and consumers, our estimates lead to a mean and median for search costs of \$1.30 and \$0.25, respectively.[39]

Using these parameters, we estimate a median and modal search set size conditional on choice of 17 and 10, respectively. This suggests that the average search set includes a small fraction of the $J = 90$ products in our data. In addition, this relatively large consumer search set size implies that our parsimonious joint model is a more feasible estimation framework for similar categories, while a full-simulation-based estimation may be challenging to implement.

Finally, we compute the consumer's willingness to pay for various product attributes and compare their values with Kim et al. (2010). First, we report that the median brand premium of Sony over Panasonic across consumers from our model is \$89, while the corresponding value from Kim et al. (2010) is \$598. In addition, the consumers' median willingness to pay for additional 1 MB of pixel is \$67 in the proposed model, while its counterpart in Kim et al. (2010) is \$321. Given the price range of camcorders in Table 1, we conclude that the proposed model that uses both search and choice data and allows for an outside option leads to more realistic inferences compared with aggregate search model.

5.3. Prediction Exercises

We predict how consumers substitute to different products when manufacturers (1) increase prices and (2) withdraw products from their product lines. Companies can use the first simulation to identify competing products and the second simulation as an impetus to product line management. As an illustrative case, we use the proposed model to potentially streamline Sony's product portfolio, because Sony has recently announced its decision to scale down its operations in many consumer electronics categories, including digital cameras (Hofilena 2014).

We start by predicting how consumers substitute selected products when their prices increase. To this end, we increase the price of each product by 10% and compute cross-price elasticities of other products.[40] Table 5 shows the substitutes with the highest and second-highest cross elasticities for selected models. We find that the substitute products share some or all attribute features or have similar attribute values with the focal product. For instance, the two best substitutes for a Sony DVD product with a retail price of \$351 are other products with same media type, in a similar price range. As another example, the best substitutes for a compact Sharp product with flash memory (FM) media selling at \$594 are other compact camcorders

Table 5. Price-Induced Substitution Patterns of Selected Camcorders

Focal product	Price (\$)	Own elasticity	Best substitute	Price (\$)	Cross elasticity	Second-best substitute	Price (\$)	Cross elasticity	% to outside goods
Sony, DVD	351	−1.53	Samsung, DVD	294	0.12	Sony, DVD	437	0.10	3.9
JVC, Hard drive	409	−1.52	JVC, Hard drive	510	0.12	JVC, Hard drive	452	0.09	4.6
Samsung, FM, Compact	399	−1.06	Sharp, FM, Compact	449	0.05	Sharp, FM, Compact	436	0.02	5.1
Sony, DVD, HD	754	−2.32	Panasonic, DVD, HD	834	0.11	Sony, DVD, HD	973	0.09	3.8
Sharp, HD, FM	594	−1.41	Sharp, HD, FM	449	0.48	Sharp, HD, FM	718	0.16	4.0

Table 6. Top Market Share Gainers from Sony's Smaller Product Portfolio

Sony's share changes before and after the decision					
			Old share (%)	New share (%)	Share gain (%)
Top 24	Sony		40.72	42.64	1.92
Bottom 7	Sony		3.59	0.00	−3.59
All	Sony		44.31	42.64	−1.67
Products with the largest share changes					
Brand	Media type	Price ($)	Old share (%)	New share (%)	Share gain (%)
Sony	MiniDV	529	5.32	5.55	0.23
JVC	Hard drive	677	8.76	8.95	0.19
Sony	Hard drive	778	4.77	4.96	0.19
Sony	Hard drive	644	3.75	3.89	0.14
Sony	MiniDV	254	2.81	2.94	0.13

from Sharp in similar price ranges. We believe these implied consumer substitution patterns are realistic in the camcorder market. The model is also able to quantify the percentage of consumers who choose not to buy any options because of the price change (last column). Overall, manufacturers can use the proposed model to identify close competitors and quantify the impact of attribute changes on the market structure.

For the second exercise, we predict market share changes if Sony withdraws some of its least popular products and streamlines its portfolio. In the past, Sony has experienced product proliferation in the camcorder category offering 31 products, or about one-third of all the options in our empirical data. Given its poor performance in several consumer electronics categories (Hofilena 2014), Sony could rationalize its product portfolio by discontinuing its least popular products.

As an illustration, we simultaneously withdraw the seven least popular products from Sony's product line and study the redistribution of its market shares between Sony and other manufacturers. Table 6 shows the simulation results. Before withdrawal, Sony's total market share is 44.31%, with 40.72% coming from the top 24 products and 3.59% from the bottom 7 options. After the decision to drop the bottom seven products, Sony's new total market share is 42.64%. Therefore, instead of losing 3.59% of the market share from its bottom seven products, because of internal substitution, Sony is predicted to lose only 1.67%. The list of products that is predicted to gain market shares is also shown in the same table. For instance, a Sony MiniDV camcorder with price of $529 gains an additional market share of 0.23%. Sony can replicate this analysis with other products to find the best set to maintain.

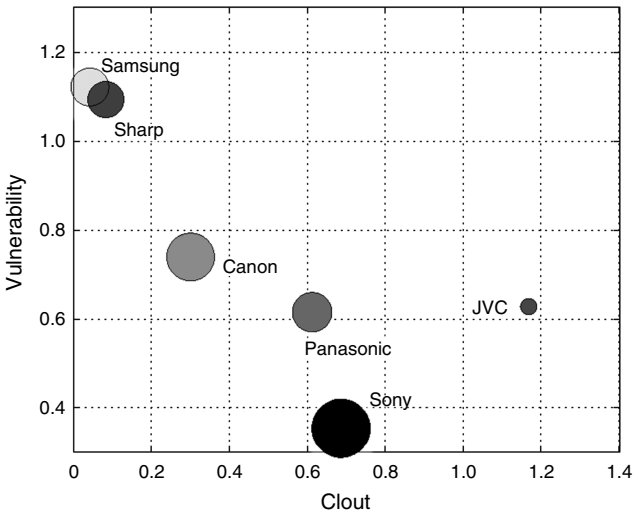
5.4. Market Structure Analysis

Analyzing the market structure is informative for product line managers concerned with the positioning of their brands (Van Heerde et al. 2004). We illustrate how our model provides insights about market structure in a durable goods category, in which research is far less common compared with consumer packaged goods categories (e.g., Elrod 1988, Erdem 1996).

We use the framework of clout and vulnerability to represent the competitive positions of brands (Kamakura and Russell 1989, Van Heerde et al. 2004, Sonnier et al. 2011).[41] Figure 2 compares the clout and vulnerability for the six major camcorder manufacturers. Each circle represents a manufacturer, and the x and y coordinates represent clout and vulnerability, respectively. The size and color of the circle represent the average sales rank and selling price: a larger circle represents more sales and a darker color represents higher price.

We first observe an overall negative association between clout and vulnerability, which implies

Figure 2. Clout and Vulnerability for Camcorder Manufacturers (Larger Circles Represent Higher Sales and Darker Circles Represent Higher Selling Prices)



asymmetric competitive positions among the brands (Kamakura and Russell 1989). JVC and Sony show the highest and second-highest clout levels among the brands. Sony, with an extensive product line, shows the high level of clout and lowest level of vulnerability, which may in part be attributable to its highest brand recognition and reputation for quality. JVC offers 14 products, 8 of which are based on hard drive media type: given its small circle size, we infer that JVC mainly serves a small consumer segment with strong preferences for the hard drive media type, which helps justify the high clout and low vulnerability. By contrast, Samsung and Sharp are the weakest brands, with the lowest clout and a high vulnerability.

In summary, the clout and vulnerability graph—constructed from search and choice data and estimates from the proposed model—helps managers understand market structure and obtain essential insights regarding the price competitiveness of each brand in the category.

6. Conclusion

In this paper, we propose a theory-based joint model of optimal sequential search and choice. On the basis of the premise that consumers look for information to fully resolve match values about products in a costly search environment, we conceptualize search sets as the outcome of an optimal sequential search process, with consumers making their choices from the resulting set of alternatives. Our model fully characterizes the costly search and choice decisions driven by the same demand primitives in a tractable way. Through extensive data experiments, we demonstrate that the partial simulation-based joint model of search and choice outperforms both frequency-based simulation methods and aggregate search models. For our empirical analysis, we use aggregate consumer search and choice data on the digital camcorder category from Amazon.com and study consumer substitution patterns and market structure.

We make the following contributions to the literature on consumer information search and aggregate demand models. Methodologically, our model explains search and choice decisions subject to conditions induced by optimal sequential search with a parsimonious expression for choice probability. By doing so, the model leads to a partial-simulation estimation framework and avoids full simulation-based estimation. This feature is particularly attractive in cases when consumer search sets may be large. In addition, given that consumer search data capture rich substitution patterns and because the identification of consumer heterogeneity has been one of the main challenges in aggregate demand models, the joint use of search and choice data helps researchers in the study of aggregate demand models. Substantively, we apply the proposed model to aggregate-level browsing and sales data from Amazon.com. Using the estimated demand primitives, we study consumer substitution patterns in the presence of price changes, the market outcome from the product withdrawal, and the market structure at the manufacturer level. We believe that the proposed model helps product managers identify substitute products and understand the competitive positions of their brands using public data available at many retailers.

Acknowledgments

The authors thank seminar participants at the 2012 Marketing Science Conference in Boston, Massachusetts; the 9th Korea University Business School International Marketing Symposium in Seoul, Korea; and the 2013 Invitational Choice Symposium in Noordwijk, the Netherlands.

Appendix A. Public Data on Consumer Search and Choice

There are many online retailers that provide data that summarize the aggregate-level consumer search and purchase. Such data can be used to study aggregate demand in a vast array of product categories. Table A.1 summarizes the availability of such data for some of the largest online retailers as of April 2014.

Table A.1. Summary of Online Retailers and Data Availability

Online store	Search data	Choice data	Choice conditional on search
Amazon.com, Amazon.co.uk	Customers who viewed this item also viewed . . .	Sales rank	Customers who viewed this item bought . . .
Bestbuy.com	Customers who viewed this item also viewed . . .	Best seller (sort)	NA
Walmart.com	NA	Best seller (sort)	Customers who viewed this item bought . . .
Target.com	Guests who viewed this item also viewed	Best seller (sort)	NA
Overstock.com	Customers who viewed this item also . . .	Best seller (sort)	NA
Kmart.com	Customers who viewed this item also . . .	Best seller (sort)	NA
QVC.com	Customers who viewed this item also . . .	Best seller (sort)	NA

Appendix B. An Example on Joint Probability of Optimal Sequential Search and Choice

Suppose we model the joint probability that, out of $J = 4$ options, a consumer searches for a set of three options in the order of $S_K = [1,2,3]$ and makes a choice of $j = 2$. Let option l's utility be $u_l = V_l + e_l$, where V_l and e_l are expected and random utility components for l, respectively. Next, l's reservation utility is denoted by z_l, which is a measure of attractiveness to search for l and equals the hypothetical, in-hand utility that makes the consumer indifferent about searching option l (see Weitzman 1979 or Section 2.3 in this paper).

From the observed search sequence in S_K, option 1 was the most attractive to search, while option 4 was the least attractive; i.e., $z_1 > z_2 > z_3 > z_4$. Collectively, optimal sequential search and choice generate a set of restrictions on e_l. First, the choice of $j = 2$ generates the following set of inequalities on the utilities of the searched options:

$$\begin{aligned} V_2 + e_2 &> V_1 + e_1, \\ V_2 + e_2 &> V_3 + e_3. \end{aligned} \tag{B.1}$$

Second, the sequence and composition of optimal search set, S_K, imposes the following set of inequalities on the random utility components of e_1 and e_2:

$$\begin{aligned} e_1 &< z_2 - V_1, \\ e_2 &< z_3 - V_2, \\ e_2 &> z_4 - V_2. \end{aligned} \tag{B.2}$$

Note that the original form of second inequality is $\max(V_1 + e_1, V_2 + e_2) < z_3$. However, given $V_2 + e_2 > V_1 + e_1$, we can simplify it involving option 2 only. The same logic holds for the third inequality. Intuitively speaking, the decision to search for an extra product after searching option 1 implies that the utility draw for option 1 was not attractive enough to forgo the expected attractiveness of the unsearched set. For instance, in Equation (B.2), the second inequality of $e_2 < z_3 - V_2$ means that the option 3 is attractive to search since its reservation utility of z_3 is greater than the realized utility value of the best alternative so far, $u_2 = V_2 + e_2$.

If we ignore search, accounting solely for choice conditions in Equation (B.1) leads to a conditional choice probability of

$$\Pr(j = 2 \mid S_K) = \Pr(e_1 < V_2 - V_1 + e_2,\ e_3 < V_2 - V_3 + e_2),$$

which, assuming that e_l are iid random variables that follow a normal distribution, gives a probit conditional choice probability expressed as

$$\Pr(j = 2 \mid S_K) = \int_{-\infty}^{\infty} \prod_{l \neq 2}^{3} \Phi_l(V_2 - V_l + e_2)\phi_2(e_2)\, de_2. \tag{B.3}$$

When considering both optimal search and choice, we simultaneously account for inequalities in (B.1) and (B.2). Under both sets of conditions, the joint probability of $\Pr([j = 2] \cap S_K)$ is

$$\begin{aligned} \Pr\big([e_1 < V_2 - V_1 + e_2, e_3 < V_2 - V_3 + e_2] \\ \cap [e_1 < z_2 - V_1, e_2 < z_3 - V_2, e_2 > z_4 - V_2]\big). \end{aligned} \tag{B.4}$$

Endnotes

[1] For a more elaborate list of available online data sources on consumer browsing and purchase, please refer to Appendix A. In addition, Amazon.com itself represents a large volume of sales in a number of categories: for example, more than 60% of consumers are willing to buy from consumer electronics goods categories at this online retailer (Walker Sands 2014).

[2] We provide an example in Appendix A.

[3] As part of the survey, the respondents report the dealers they visited prior to their purchase.

[4] We offer examples of Amazon.com's store-level browsing and choice data at the beginning of Section 1. We also provide a list of public data sources in Appendix A.

[5] Note that our interpretation of e_{il} is similar to those found in Anderson and Renault (1999) and Kim et al. (2010).

[6] Note that some recent papers model consumer search for attributes and assume that the idiosyncratic match values are unobserved to the researcher but observed by the consumer. For instance, in Honka and Chintagunta (2017), consumers search for prices instead of match values. Our definition of search is broadly applicable to product markets for which consumers may search for more than product attribute values. This distinction impacts the model, because in Honka and Chintagunta (2017), choice is still stochastic from the point of view of the researcher, as the idiosyncratic match values are still unobserved; in our case, choice given the match value is deterministic but is measured with error (see Section 3.2).

[7] This means that the outside good is present in all search sets. Additionally, and common to empirical work on search (e.g., Kim et al. 2010, Honka and Chintagunta 2017), we assume that the first alternative is also free to search, thereby remaining consistent with observing consumers who search at least one product.

[8] We omit the individual index i for clarity.

[9] Hence, product of $l = 1$ is the product with the highest reservation utility, and product $l = J$ has the lowest reservation utility.

[10] We do not observe individual sequences of search in our data and therefore do not make use of the individual-level information implied by the selection rule. However, we represent our data as generated from aggregations of individual optimal search sequences. Thus, we use the selection rule from the optimal search theory in predicting individual search sequence in our empirical model development.

[11] If there are ties in the reservation values, the number of possible sets increases.

[12] This proposition makes the proposed approach computationally feasible in aggregate models. Without optimal search sequence, i.e., without the rule in which the only search path considered is the optimal one with descending reservation utilities, the number of all feasible search sets that contain j is approximately 6×10^{26} with $J = 90$. Evaluating a sum over this many terms is impossible.

[13] During the process of search up to alternative K, the following set of inequalities must have been successively true to continue searching:

$$\max\{u_1, \ldots, u_l\} < z_{l+1}, \quad l = 1, \ldots, K-1.$$

However, all these conditions up to $l = 1, \ldots, K-2$ are summarized by the last inequality at $l = K-1$, since

$$\max\{u_1, \ldots, u_l\} \leq \max\{u_1, \ldots, u_{K-1}\} < z_K < z_{l+1}, \quad l = 1, \ldots, K-2.$$

As a consequence, we only need the selection condition at K.

[14] For a similar formulation, see Train (2009) for the expression of the probit model under complete search.

[15] Our derivations in this section are general and can be used under different distributional assumption for the utility function as long as the corresponding CDF is used in Equation (12).

[16] Given the presort of reservation utilities in descending order, $K = \arg\min\{z_j, z_l\} = \max(j, l)$.

[17] Mathematically, the joint probability of observing a choice j, search set S_K, and search order O_l is

[18] Kim et al. (2010) provides an example of the view rank list (see p. 1002).

[19] Our discussion in this subsection on search aggregation is similar to Kim et al. (2010), and we direct readers to that paper for details.

[20] Note that if $z_j > z_k$, the probability that j and k occur together in a set is equal to the probability that k is in the set. That is, $\pi_{i,\{j \text{ and } k\}} = \pi_{ik} = \min(\pi_{ij}, \pi_{ik})$.

[21] We assume that the measurement errors, ε, for view ranks, sales ranks, and conditional shares in Equations (20), (25), and (29) are independent and identically distributed (iid) within each data set and are independent across data sets.

[22] Our likelihood function may be nonsmooth because the ranking of alternatives for search in our sequential search model can change discretely in response to small variations of the parameters (see Section 2.3). To maximize this function, we follow a two-stage estimation strategy where we first use a stochastic algorithm, called the differential evolution algorithm, to extensively explore the parameter space, and we next use the best estimates as the starting point in a gradient-based search algorithm in the second stage. Details of this procedure are outlined in Kim et al. (2010).

[23] In computing k's search rank from the view rank data, we take into account both the incidence and position of k's appearance on the view rank lists of focal products $j \neq k$ as well as j's popularity. We approximate k's search popularity as

$$f_k = \sum_{j=1, k \neq j}^{J} r_j^S \cdot r^V(k \mid j),$$

where r_j^S is the sales rank of the focal option j and $r^V(k \mid j)$ is k's rank position on j's view list. We have also tried other approaches, including $r_j^S = 1$, with similar implications.

[24] Note that this variable is the second element in X_j vector in the utility.

[25] In this exercise, we estimated the model without an outside option to accommodate the comparison study with frequency-based estimator. However, in the next subsection of 4.3, we estimate the model with the outside option and confirm that the model recovers the outside mean utility value of V_0 well.

[26] This exercise was conducted under identical computing conditions for all methods. For the hardware, we used a desktop computer with Intel Core i7-2600 CPU with a clock speed of 3.4 GHz and a RAM size of 16 GB. For computational reasons, we did not attempt higher values than $Q = 60$ in this exercise.

[27] The gains in computational time increase with the number of products in the choice set: for $J = 32$, one objective function evaluation took 18 seconds and 53 seconds for the proposed method and frequency-based estimator, respectively. For $J = 64$, one objective function evaluation took 29 seconds and 102 seconds.

[28] Therefore, it is a departure of the original study by Kim et al. (2010).

[29] The outside option is implemented in the following manner in our estimation. A consumer already has the known value of outside option, $u_{i0} = V_0 + e_{i0}$, which is not known to the analysts. Then the first inner option 1 is free to search; i.e., we impose $Z_{i1} = \infty$ in the implementation. Given that it is always $u_{i0} < Z_{i1}$, consumers will always search for first inner option of 1. After this, the outside option is treated just as another choice option once it is in the search set.

[30] A higher number of draws is needed to account for a small fraction of consumers who buy from the category upon search.

[31] Consistent with our empirical setting, this set of attributes and their values are mostly available at the product links. Therefore, consumers have access to these attribute values prior to their decision to search. In addition, our selection of attributes is similar to those in Gowrisankaran and Rysman (2012).

[32] The error term, e_{ij}, captures consumer's idiosyncratic, remaining match value such as consumer's experience or usage scenarios for the option. For instance, consumer reviews that are typically available in the product detail page provide such information (Chen and Xie 2008).

[33] If additional information is available on how products are presented to consumers, our model is flexible enough to handle the effects of other components through the search cost function.

[34] In principle, we could have common variables in the utility function and search cost—because we observe both search and choice data—and let the data decide which variables are significant in the two components. However, we opted to limit the variables that enter the search costs for the following reasons. First, to the best of our knowledge, prior literature does not provide any theory that guides us on how certain product attributes affect consumer search costs. This means that even if a full-scale estimation is implemented, we may not be able to interpret why some attributes lead to higher or lower search costs, limiting its applicability. Second, our approach for utility and search cost specifications is consistent with recent empirical papers in which researchers broadly base search costs on information environments at the online stores. For instance, there is minimal or no overlap between search cost and utility specifications in Ghose et al. (2013) and Chen and Yao (2016). Third, a fully parametrized search cost specification would lead to multiple estimation issues. A full search cost specification would add 12 more parameters, increasing the total number of parameters by over 40% in our empirical application. Given the aggregate nature of our data, we think an empirical model of this size would be very demanding on the data and its estimation require a much higher computational time.

[35] Similar modeling assumptions are made in empirical models of optimal sequential search—for example, in Kim et al. (2010) and Chen and Yao (2016).

[36] The in-sample sales rank prediction improvement may seem marginal between the two models. A short discussion is in order. First, we believe this finding is applicable to our empirical context only in which we observe a far higher number of observations in view rank data than in sales rank data. Hence, the likelihood contribution from the sales rank is limited in our empirical setting. Second, in the "Inferences" section, we show that the implications of our model are more reasonable.

[37] Our choice of the values in this robustness study depends mainly on Amazon.com's conversion rates reported at various online media sources in which we find values ranging from 16.4% to 9.6%. These conversion rates translate into outside option shares of 83.6% and 90.4% and are close to the two values used in the robustness tests.

[38] By contrast, an informal inspection for the most popular 42-inch LCD HD TV sold at Amazon.com in summer 2014 yields prices between \$350 and \$450.

[39] Monetary search costs can be computed by dividing the search cost by the price coefficient (Honka and Chintagunta 2017).

[40] To study substitution patterns based on consumer preferences only, we set search costs identical for all options in this exercise, thereby focusing on substitutions net of any effects from navigational features at Amazon.com. Our primary goal is to understand consumer substitutions in the general product market.

[41] For the operationalization of clout and vulnerability, we follow Kamakura and Russell (1989).

References

Albuquerque P, Bronnenberg BJ (2009) Estimating demand heterogeneity using aggregated data: An application to the frozen pizza category. *Marketing Sci.* 28(2):356–372.

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

Bajari P, Fox JT, Ryan SP (2007) Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients. *Amer. Econom. Rev.* 97(2):459–463.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Berry S, Levinsohn J, Pakes A (2004) Differentiated products demand systems from a combination of micro and macro data: The new car market. *J. Political Econom.* 112(1):68–105.

Bresnahan TF (1987) Competition and collusion in the American automobile industry: The 1955 price war. *J. Indust. Econom.* 35(4):457–482.

Bronnenberg BJ, Kim JB, Mela CF (2016) Zooming in on choice: How do consumers search for cameras online? *Marketing Sci.* 35(5):693–712.

Chen Y, Xie J (2008) Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Sci.* 54(3):477–491.

Chen Y, Yao S (2016) Sequential search with refinement: Model and application with click-stream data. *Management Sci.*, ePub ahead of print September 28, http://dx.doi.org/10.1287/mnsc.2016.2557.

Draganska M, Klapper D (2011) Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *J. Marketing Res.* 48(4):653–669.

Efron B, Tibshirani RJ (1994) *An Introduction to the Bootstrap* (CRC Press, Boca Raton, FL).

Eisenberg B (2009) Top 10 online retailers by conversion rate: August 2009. Accessed September 18, 2016, http://www.bryaneisenberg.com/top-10-online-retailers-by-conversion-rate-august-2009/.

Elrod T (1988) Choice map: Inferring a product-market map from panel data. *Marketing Sci.* 7(1):21–40.

Erdem T (1996) A dynamic analysis of market structure based on panel data. *Marketing Sci.* 15(4):359–378.

Ghose A, Ipeirotis PG, Li B (2013) Surviving social media overload: Predicting consumer footprints on product search engines. Working paper, University of New York, New York.

Goeree MS (2008) Limited information and advertising in the US personal computer industry. *Econometrica* 76(5):1017–1074.

Gowrisankaran G, Rysman M (2012) Dynamics of consumer demand for new durable goods. *J. Political Econom.* 120(6):1173–1219.

Hancox P (2008) Is Amazon's 9.6% conversion rate low? Here's why I think so.... (February 3), http://www.conversionblogger.com/is-amazons-96-conversion-rate-low-heres-why-i-think-so/.

Hofilena J (2014) Sony to cut more jobs at failing electronic equipment division. *Japan Daily Press* (January 3), http://japandailypress.com/sony-to-cut-more-jobs-at-failing-electronic-equipment-division-0341844/.

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 U.S. auto insurance industry. *Marketing Sci.* 36(1):21–42.

Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.

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.

Magid L (2007) Today's HDTV, or next year's? *New York Times* (December 20), http://www.nytimes.com/2007/12/20/technology/personaltech/20basics.html.

McFadden D (1989) A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57(5):995–1026.

Moraga-González JL, Sándor Z, Wildenbeest MR (2015) Consumer search and prices in the automobile market. Working paper, Vrije Universiteit Amsterdam, Amsterdam.

Petrin AK (2002) Quantifying the benefits of new products: The case of the minivan. *J. Political Econom.* 110(4):705–729.

Sonnier GP, McAlister L, Rutz OJ (2011) A dynamic model of the effect of online communications on firm sales. *Marketing Sci.* 30(4):702–716.

Train K (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, Cambridge, UK).

Van Heerde HJ, Mela CF, Manchanda P (2004) The dynamic effect of innovation on market structure. *J. Marketing Res.* 41(2):166–183.

Walker Sands (2014) Reinventing retail: What businesses need to know for 2014. White paper, Walker Sands, Chicago.

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