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

Jun B. Kim Paulo Albuquerque Bart J. Bronnenberg*

May 9, 2015

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 a semi-closed form expression for the probability of choice that obeys the full set of restrictions imposed by optimal sequential search. Our joint model leads to a partial simulation-based estimation that avoids high-dimensional integrations in the evaluation of choice probabilities and that is particularly attractive when search sets are large. We illustrate the advantages of our approach using 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.

Keywords: Optimal sequential search, discrete choice, consumer heterogeneity, aggregate demand models, information economics, market structure

*Jun B. Kim is an Assistant Professor of Marketing at Hong Kong University of Science and Technology. Paulo Albuquerque is an Associate Professor of Marketing at INSEAD. Bart J. Bronnenberg is a Professor of Marketing, and CentER Research Fellow at Tilburg University, and a Research Fellow at CEPR, London. We thank seminar participants at the 2012 Marketing Science Conference in Boston, MA, the 9th KUBS International Marketing Symposium in Seoul, South Korea, and the 2013 Invitational Choice Symposium in Noordwijk, the Netherlands. Bart Bronnenberg gratefully acknowledges EU funding from the Marie Curie Program (IRG 230962) and the Netherlands Foundation for Scientific Research (NWO Vici Grant).

1 Introduction

Online retailers routinely collect and publish data on consumer choice behavior. Recently, they have also started displaying 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 unified manner. 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 search into a model of choice imposes constraints on the utilities for searched products.[2] In these cases, the evaluation of choice probabilities is complex and typically calls for numerical simulation to estimate the parameters of interest while accounting for such constraints. However, simulation is impractical when the number of searched items becomes large. In contrast, we develop an expression involving only univariate integration, which broadens the applicability of models that simultaneously consider search and choice in categories with larger choice sets.

By combining aggregate search and choice data, we achieve 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. To better pinpoint preferences, researchers have traditionally augmented such data with information such as second choice surveys (Berry,

[1] For a more elaborate list of available data on consumer browsing and purchase, please refer to Appendix A. 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 (WalkerSands, 2014).

[2] We provide an example in Appendix B.

Levinsohn, and Pakes 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 identification (Berry, Levinsohn, and Pakes, 2004, page 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 strong 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, Kim, and Mela 2014).

Substantively, this paper seeks to contribute to the empirical search literature in several ways. First, comparing our work to a related paper by Kim, Albuquerque, and Bronnenberg (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. In addition, the use of both data facilitates the 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 related literature (e.g., Ghose, Ipeirotis, and Li (2012) and Chen and Yao (2012)) 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 (2014) and Koulayev (2014)), include both the sequence and composition of consumer search, as we do in this paper. In contrast to these works, our approach leads to a semi-closed form, partial-simulation method that does not require the high dimensional numerical integration and 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, Kim, and Mela, 2014). Many researchers (e.g., Train, 2009) advocate for the adoption of such methods whenever possible, due to higher accuracy and lower computational cost.

Lastly, 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, Sándor, and Wildenbeest (2012) also offer a model of endogenous search sets, but in a non-sequential search framework, compared to our sequential search approach. In addition, they rely on consumer location as the basis for modeling consumer search costs, whereas we use the consumer search behavior for this purpose directly. We apply the proposed approach 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 durables.

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 addition, we compare the performance of the proposed method with full simulation-based methods, in terms of accuracy and computation cost. Section 4 discusses the estimation results and empirical application. We conclude in section 5.

2 A Probit Model of Sequential Search and Choice

2.1 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 do 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 $\ell = 1,...,J$ for consumer i is

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

with

$$\begin{aligned} V_{i\ell} &= X_\ell b_i, \\ b_i &\sim N(b, B), \\ e_{i\ell} &\sim N\left(0, \sigma_{i\ell}^2\right), \end{aligned}$$

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 consumer search process as a costly consumer effort to obtain the full match value of option ℓ. Prior to search, we assume that consumers know the expected match value, $V_{i\ell}$, but that the realization of the exact match value is subject to a shock, $e_{i\ell}$, drawn from a known distribution. To fully resolve the unknown match value of $V_{i\ell} + e_{i\ell}$, consumers engage in costly search.[3]

In our empirical setting, the values of important attributes, which are captured in $V_{i\ell}$, are readily accessible by consumers at zero search cost prior to searching for ℓ. Given $V_{i\ell}$, the goal of search is to resolve the unknown value of $e_{i\ell}$ by incurring a search cost of $c_{i\ell}$. Upon search, consumers have access to more details about the option and receive the realized value of $e_{i\ell}$. A large positive value of $e_{i\ell}$ 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

$$u_{i0} \sim N\left(V_0, \sigma_0^2\right),$$

and we assume that consumer i knows the value of u_{i0} free of any search cost.[4] The demand primitives in

[3] Note that our interpretation of $e_{i\ell}$ is similar to those found in Anderson and Renault (1999) and Kim, Albuquerque, and Bronnenberg (2010).

[4] This means that the outside good is present in all search sets. Additionally, and common to empirical work on search (e.g., Kim, Albuquerque, and Bronnenberg 2010; Honka and Chintagunta 2014), we assume that the first alternative is also free to search, thereby remaining consistent with observing consumers who search at least one product.

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, Albuquerque, and Bronnenberg (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.

2.2 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, Albuquerque, and Bronnenberg (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[5]

$$\mathscr{B}_\ell(u^*) = \int_{u^*}^{\infty} (u_\ell - u^*) f(u_\ell) \, du_\ell, \tag{2}$$

where $f(\cdot)$ is the probability density distribution of u_ℓ. Intuitively, $\mathscr{B}_\ell(u^*)$ 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

$$\mathscr{B}_\ell(z_\ell) = c_\ell, \tag{3}$$

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

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

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, $r(\ell|\theta)$, is a one-to-one mapping with product index ℓ, we cast the model using ℓ as the order of reservation utilities.[6] The following result holds for the probability to search.

Proposition 1. *Rank products on reservation utility. The probability that the option with the k^{th} highest reservation utility is searched is equal to*

$$\begin{aligned}\pi_k &= \Pr\left(\max_{\ell=1}^{k-1}(V_\ell + e_\ell) < z_k\right) \\ &= \prod_{\ell=1}^{k-1}\Phi_\ell(z_k - V_\ell),\ k > 1, \end{aligned} \tag{4}$$

where $\Phi_\ell(\cdot)$ 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, Albuquerque, and Bronnenberg (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. □

2.3 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. Afterwards, proposition 3 describes the joint probability of search and choice under sequential search.[7]

[6] Hence, product $\ell = 1$ is the product with the highest reservation utility and product $\ell = J$ has the lowest reservation utility.

[7] 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 modeling individual search sequence in our empirical model development.

Proposition 2. *Rank products on reservation utility. The probability that the j^{th} ranked product is chosen equals*

$$\Pr(j) = \sum_{K=j}^{J} \Pr(j, S_K), \tag{5}$$

where $\Pr(j, S_K)$ *denotes the joint probability that the j^{th} ranked product is chosen from S_K, and $S_K = [1, \ldots, K]$ is an ordered set such that if $z_k \geq z_\ell$ then $1 \leq k < \ell \leq K$.*

Proof. This follows directly from the application of 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). Further, the selection rule states that consumers search in the order of decreasing reservation utilities. After alternatives are ranked by descending reservation values, the super-set of all possible optimal search sets consists of only J member sets, each containing the $K = 1, ..., J$ products with the highest reservation values.[8] □

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 (5) is not over all 2^{J-1} possible permutations that contain a particular product.[9] Instead, with the application of selection rule under optimal sequential search, 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 optimal 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*

$$\begin{aligned}\Pr(j, S_K) &= \int_{z_{K+1}-V_j}^{z_K - V_j} \prod_{\ell \neq j}^{K} \Phi_\ell (V_j - V_\ell + e_j)\, \phi_j(e_j)\, de_j + \\ &\qquad\qquad I(j = K)\,(1 - \Phi_j(z_j - V_j))\,\pi_j, \end{aligned} \tag{6}$$

where $I(j = K)$ is an indicator variable that is 1 if $j = K$ and 0 otherwise. For completeness, (1) because

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

[9] This proposition makes the proposed approach computationally feasible in aggregate models. With $J = 90$, the number of sets that contain j is approximately 6×10^{26}. Evaluating a sum over this many terms is impossible.

consumers can not purchase products that were not searched, $\Pr(j,S_K)=0$ *when* $j>K$*, and (2) because the consumer can not 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 $j=\arg\max_{\ell=1,\ldots,K}\{u_\ell\}$, that is, $V_j+e_j>V_\ell+e_\ell,\ \ell\neq j,\ell=1,\ldots,K$. With independent and normally distributed e_ℓ, $\Pr(e_\ell<V_j-V_\ell+e_j|e_j)=\Phi_\ell(V_j-V_\ell+e_j)$, for $\ell\neq j$. The conditional probability that j has the highest utility within the ordered set of S_K is the product of these probabilities, $\prod_{\ell\neq j}^{K}\Phi_\ell(V_j-V_\ell+e_j)$.

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 $\max\{u_1,\cdots,u_{K-1}\}<z_K$, i.e., the maximum utility in hand after searching $\{1,...,K-1\}$ is less than the highest reservation utility from the un-searched set $\{K,...,J\}$.[10] This means that $V_\ell+e_\ell<z_K$, for $\forall\ell\in 1,...,K-1$. 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 $\ell\neq j$,

$$\Pr(j)=\int_{-\infty}^{\infty}\prod_{\ell\neq j}^{J}\Phi_\ell(V_j-V_\ell+e_j)\,\phi_j(e_j)de_j. \tag{7}$$

This model is a formulation of the probit model with independent error terms e.[11] We now impose the

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

$$\max\{u_1,\cdots,u_\ell\}<z_{\ell+1},\ell=1,\cdots,K-1.$$

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

$$\max\{u_1,\cdots,u_\ell\}\leq\max\{u_1,\cdots,u_{K-1}\}<z_K<z_{\ell+1},\ell=1,\cdots,K-2.$$

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

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

restrictions implied by the selection and stopping rules on these choice probabilities, which turn out to be simple integration limits in Equation (7).

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 $V_\ell + e_\ell < z_K$, $\ell = 1, ..., K$. At the same time, the choice of j implies that $V_\ell + e_\ell < V_j + e_j$ for $\forall \ell \neq j$. Note that this yields another $K-1$ restrictions on e_ℓ, $\ell \neq j$. Both 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 $V_\ell + e_\ell < V_j + e_j$ for $\forall \ell \neq j$ imply that $V_\ell + e_\ell < z_K$ for $\forall \ell \neq j$. 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 (7), gives the following joint probability.

$$\Pr(j, S_K, u_K < z_K) = \int_{z_{K+1}-V_j}^{z_K - V_j} \prod_{\ell \neq j}^{K} \Phi_\ell \left(V_j - V_\ell + e_j\right) \phi_j(e_j) de_j, \tag{8}$$

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

Computationally, Equation (8) is very similar to the probit in Equation (7) 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, $\max\{u_1, \cdots, u_{K-1}\} < z_K$, 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 (4). 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 $\Pr(S_K|u_K \geq z_K) = \pi_K$. Therefore,

$$\Pr(K, S_K|u_K \geq z_K) = \pi_K. \tag{9}$$

Note that π_K does not depend on e_K. To obtain the unconditional probability, we use that the condition $u_K \geq z_K$ is equivalent to $e_K \geq z_K - V_K$, and integrate the conditional probabilities (9) over e_K to obtain

$$\begin{aligned} \Pr(K, S_K, e_K \geq z_K - V_K) &= \int_{z_K - V_K}^{\infty} \pi_K \phi_K(e_K) de_K \\ &= \pi_K \left(1 - \Phi_K(z_K - V_K)\right). \end{aligned} \tag{10}$$

Combining Equations (8) and (10), we can write for $j \in 1, ..., K$

$$\begin{aligned} \Pr(j, S_K) = \textstyle\int_{z_{K+1}-V_j}^{z_K - V_j} \prod_{\ell \neq j}^{K} \Phi_\ell(V_j - V_\ell + e_j)\, \phi_j(e_j) de_j + \\ I(j = K) \cdot \pi_j \left(1 - \Phi_j(z_j - V_j)\right) \end{aligned} \tag{11}$$

which proves the proposition. With both parts of Equation (6), the computation of the joint probability involves only uni-dimensional integration. □

Proposition 3 develops a parsimonious expression for the summand in Equation (5), 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.[12]

2.4 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 $\Pr(j|\ell)$, $1 \leq j, \ell \leq J$.

[12] 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 11.

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

$$\Pr(j|\ell) = \frac{\sum_{K=\max(j,\ell)}^{J} \Pr(j, S_K)}{\pi_\ell} \tag{12}$$

where π_ℓ is the probability that ℓ^{th} option is searched (see Equation 4) and $\Pr(j, S_K)$ *is the probability that j is chosen and the optimal set is S_K (see Equation 6).*

Proof. We write the conditional choice probability as,

$$\Pr(j|\ell) = \frac{\Pr(j,\ell)}{\Pr(\ell)}, \tag{13}$$

where $\Pr(\ell)$ is the probability that ℓ is searched and $\Pr(j,\ell)$ is the joint probability that ℓ is searched and j is chosen. Note that the denominator, $\Pr(\ell)$ is equal to the probability that ℓ is in the optimal search set and is given by Equation (4), with

$$\Pr(\ell) = \pi_\ell. \tag{14}$$

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

$$\Pr(j,\ell) = \sum_{K=\max(j,\ell)}^{J} \Pr(j, S_K) \tag{15}$$

where the expression for the summand, $\Pr(j, S_K)$, is given in Equation (11). Finally, we obtain the conditional choice probability of $\Pr(j|\ell)$ by substituting Equations (14) and (15) into Equation (13) to obtain Equation (12).

$$\Pr(j|\ell) = \frac{\sum_{K=\max(j,\ell)}^{J} \Pr(j, S_K)}{\pi_\ell}, \tag{16}$$

which proves the proposition. $\square$

The result in proposition 4 relies on individual level probabilities, which are already computed in Equation (3). Therefore computing the conditional choice shares from our individual level model adds no computational burden other than taking a sum over less than J terms. Equation (12) 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.

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, Albuquerque, and Bronnenberg (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 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 pre-sorting 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. [13]

3 Data, Estimation, and Identification

3.1 Data

We apply our 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 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

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

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

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 O_J, we invoke the selection rule from the optimal sequential search theory, pre-sort options in the descending reservation values given a set of parameters, and operate on the pre-sorted 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 $\Pr(j, S_K | O_J)$.

Attributes	Ranges
Brand	Sony (31), Panasonic (20), Canon (14), JVC (14), other (11)
Media Types	MiniDV (34), DVD (30), Flash Memory(FM)(9), Hard Drive (17)
Price	\$532 (mean), \$258 (std. dev.)
Form	Compact (8), Conventional (82)
High-Definition	Yes (14), No (76)
Pixel	1.74M (mean), 1.45M (std. dev.)
Zoom	20.1 (mean), 11.2 (std. dev.)
Number of Links	2.96 (mean), 6.00 (std. dev.)
Age (days)	266 (mean), 243 (std. dev.)

Table 1: Description of the choice options in the empirical data (occurrence frequency in parenthesis)

lists as view rank data.[14]

The conditional share data consists 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 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. Lastly 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.

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.

[14] Kim, Albuquerque, and Bronnenberg (2010) provides an example of the view-rank list (see page 1002).

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.

Predicting view rank data[15] For each product $j = 1, ..., J$, Amazon.com lists a set of other products that were viewed together in the form of a commonality index, $CI_{j\ell}$, defined as,

$$CI_{j\ell} = \frac{n_{j\ell}}{\sqrt{n_j} \cdot \sqrt{n_\ell}}, \tag{17}$$

where n_j and n_ℓ are the numbers of consumers who viewed products j and ℓ, respectively, and $n_{j\ell}$ denotes the number of consumers who viewed products j and ℓ together. If $CI_{j\ell} > CI_{jk}$, ℓ appears before k on j's

[15]Our discussion in this subsection on search aggregation is similar to Kim, Albuquerque, and Bronnenberg (2010) and we direct readers to that paper for details although our estimation framework is different.

view rank list. More specifically, the view rank indicator variable, $I^V_{j,\ell k}$, is defined as,

$$I^V_{j,\ell k} = \begin{cases} 1 & \text{if } CI_{j\ell} > CI_{jk} \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

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

$$CI_{j\ell} = \widehat{CI}_{j\ell} + \varepsilon^V_{j\ell} = \frac{\hat{n}_{j\ell}}{\sqrt{\hat{n}_j}\sqrt{\hat{n}_\ell}} + \varepsilon^V_{j\ell}, \tag{19}$$

where $\widehat{CI}_{j\ell}$ is the commonality index forecast and $\varepsilon^V_{j\ell} \overset{iid}{\sim} N(0, \frac{\tau_V^2}{2})$. Further, $\hat{n}_j = \sum_i \pi_{ij}$ and $\hat{n}_{j\ell} = \sum_i \min\left(\pi_{ij}, \pi_{i\ell}\right)$, $j, \ell = 1, ..., J$ and $j \neq \ell$.[16] The error term captures Amazon.com's potential measurement or aggregate-level prediction errors, similar to Bresnahan (1987) and Bajari, Fox, and Ryan (2007). The probability that product j is viewed more often with ℓ than with k in the same search session is

$$\Pr(I^V_{j,\ell k} = 1) = \Pr(CI_{jk} < CI_{j\ell}) = \Pr(\varepsilon^V_{j,\ell k} < \widehat{CI}_{j\ell} - \widehat{CI}_{jk}), \tag{20}$$

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

$$\Pr(I^V_{j,\ell k} = 1) = \Phi\left(\frac{\widehat{CI}_{j\ell}(\Theta, X) - \widehat{CI}_{jk}(\Theta, X)}{\tau_V}\right), \tag{21}$$

where Θ are model parameters, X are data, and Φ is the CDF for the standard normal distribution. The variable $I^V_{j,\ell k}$ is directly observed in the view rank data from Amazon.com. For each product j, we observe at most $\frac{1}{2} \times (J-1) \times (J-2)$ unique inequalities defined by equation (18). 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,

$$\Pr(I^V = 1|\Theta, X) = \prod_j^J \prod_{\ell \neq j} \prod_{k \neq \ell} \Pr(I^V_{j,\ell k} = 1|\Theta, X) = \prod_j^J \prod_{\ell \neq j} \prod_{k \neq \ell \neq j} \Phi\left(\frac{\widehat{CI}_{j\ell}(\Theta, X) - \widehat{CI}_{jk}(\Theta, X)}{\tau_V}\right). \tag{22}$$

[16] 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, $\pi_{i,\{j \text{ and } k\}} = \pi_{ik} = \min\left(\pi_{ij}, \pi_{ik}\right)$.

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,

$$I^S_{j\ell} = \begin{cases} 1 & \text{if } s_j > s_\ell \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

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 5). The market share, s_j, is modeled as the sum of the prediction $\hat{s}_j$ and a measurement error term,

$$s_j = \hat{s}_j + \varepsilon^S_j, \tag{24}$$

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

$$\Pr(I^S_{j\ell} = 1) = \Pr(s_\ell < s_j) = \Pr(\varepsilon^S_{j\ell} < \hat{s}_j - \hat{s}_\ell), \tag{25}$$

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

$$\Pr(I^S_{j\ell} = 1) = \Phi\left(\frac{\hat{s}_j(\Theta, X) - \hat{s}_\ell(\Theta, X)}{\tau_S}\right), \tag{26}$$

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

$$\Pr(I^S = 1|\Theta, X) = \prod_j^J \prod_{\ell \neq j}^J \Pr(I^S_{j\ell} = 1|\Theta, X) = \prod_j^J \prod_{\ell \neq j}^J \Phi\left(\frac{\hat{s}_j(\Theta, X) - \hat{s}_\ell(\Theta, X)}{\tau_S}\right). \tag{27}$$

Predicting conditional choice share data Lastly, we discuss the prediction of conditional choice shares. The observed choice share of j conditional on searching ℓ, $s_{j|\ell}$, is modeled as,

$$s_{j|\ell} = \hat{s}_{j|\ell}(\Theta, X) + \varepsilon^C_{j\ell}, \tag{28}$$

where $\hat{s}_{j|\ell}(\Theta, X)$ is the model prediction of conditional choice shares aggregated across consumers and $\varepsilon^C_{j\ell} \sim N(0, \tau_C^2)$ is the measurement error. Note that the conditional share prediction is obtained by aggregating individual conditional choice probabilities in equation (12) across draws. The probability of $\Pr(s_{j|\ell}|\Theta, X)$ is,

$$\Pr(s_{j|\ell}|\Theta, X) = \phi\left(\frac{\hat{s}_{j|\ell}(\Theta, X) - s_{j|\ell}}{\tau_C}\right) = \frac{1}{\tau_C \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{\left(\hat{s}_{j|\ell}(\Theta, X) - s_{j|\ell}\right)^2}{2 \cdot \tau_C^2}\right), \tag{29}$$

where ϕ is a PDF for a standard normal distribution. Then the joint probability of observing the set of all conditional share values of s is

$$\Pr(s|\Theta, X) = \prod_{\ell}^{J}\prod_{j} \Pr(s_{j|\ell}|\Theta, X), \tag{30}$$

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

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

$$\mathscr{L}(\Theta|Y) = \Pr(Y|\Theta).$$

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 respectively.

$$\Pr(Y|\Theta) = \Pr(I^V = 1|\Theta) \cdot \Pr(I^S = 1|\Theta) \cdot \Pr(s|\Theta) \tag{31}$$

The corresponding log-likelihood function is,

$$\mathscr{L}\mathscr{L}(\Theta|Y) = \sum_j \sum_{\ell \neq j} \sum_{k \neq \ell} \log\left(\Pr(I^V_{j,\ell k} = 1|\Theta)\right) + \sum_j \sum_{\ell \neq j} \log\left(\Pr(I^S_{j\ell} = 1|\Theta)\right) + \sum_{\ell} \sum_j \log\left(\Pr(s_{j|\ell}|\Theta)\right). \tag{32}$$

with the probabilities defined in equations (21), (26), and (29) respectively.

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,

$$g(\hat{s}_0|\Theta, X, s_0) = -\left(\hat{s}_0(\Theta, X) - s_0\right)^2. \tag{33}$$

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 (32) with this penalty term using a weight w. Thus, the objective function maximized is $\mathscr{L}\mathscr{L}(\Theta|Y) + w \cdot g(\hat{s}_0|\Theta, X, s_0)$. In the application, we have used $w = 4 \times 10E7$, 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 pre-search 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, Ipeirotis, and Li, 2012). For example, a high value product with average search costs will be popular during both search and choice stages, putatively resulting in similar search and choice shares. In 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 choice, compared to other products. This way, contrasting search and choice shares is informative about product-specific search costs.

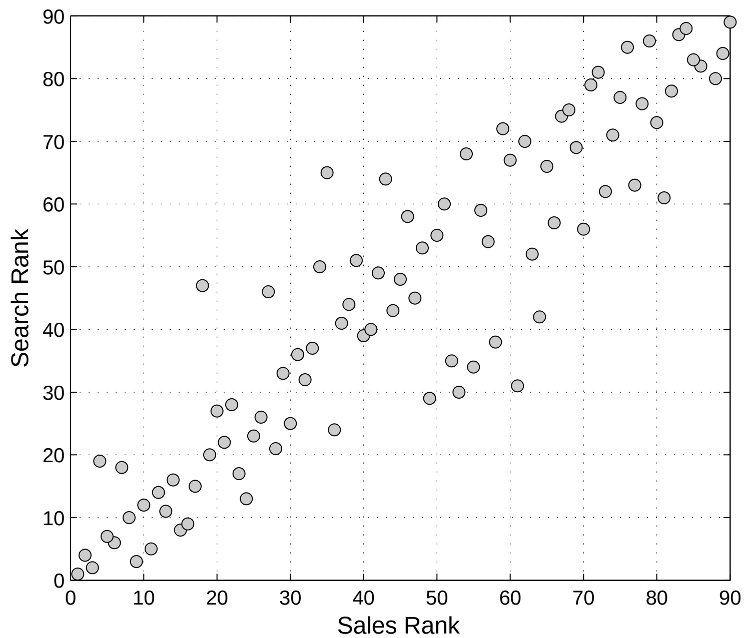

Figure 1: Scatter plot of sales-rank and view-rank of products

As an illustration, figure 1 displays search and choice ranks: each circle represents a product's sales-rank on the x-axis and its search-rank on the y-axis.[17] In the figure, if an observation lies on the 45° line, its

[17] 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.

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 even ranked 61^{st} in terms of sales but 31^{st} in terms of search. Under our model, this product is likely to have a low search cost that makes it more popular during search than choice. Not only 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 due to its lower search cost but was not chosen due to 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 homogenous 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 price, flash-memory camcorders and options ℓ_1 and ℓ_2 high price, hard-disk camcorders. If each consumer searches more within media format type than across types, and within a price level, we infer that their 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.

3.4 Performance of the Proposed Model

In this section, we conduct a numerical experiment to demonstrate the ability of the proposed model to recover parameter values and its advantages compared to frequency-based simulators.

Set-up The utility of consumer i for product j is,

$$u_{ij} = V_{ij} + e_{ij} = X_j\beta_i - \alpha_i P_j + e_{ij}, \tag{34}$$

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 with individual-specific sensitivities for product characteristics, and $\text{var}(e_{ij}) = \sigma^2$. We set $\sigma^2 = 1$ for identification purpose. We create $J = 32$ product options with 5 binary attributes X_j - arbitrarily named as brand, "Sony" or not; pixel, "> 1 MB" or "< 1 MB"; zoom, "$> 10\times$" or "$< 10\times$"; media, "mini DV" or not; and form, "compact" or not - and a continuous attribute, price. Further, we impose a theory-driven restriction on the price coefficient, assuming a log-normal distribution. Mathematically,

$$\begin{aligned} \log(\alpha_i) &\sim N(\beta_p, \sigma_p^2) \\ \beta_i &\sim N(\beta_0, \Sigma_\beta), \end{aligned} \tag{35}$$

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,

$$c_j = \exp\left(\gamma_0 + \gamma_1 X_j^c\right), \tag{36}$$

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 $X_j^c = [L_j]$ where L_j is the "number of links" to product j.

To generate 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 optimal search sets and choices. 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 (19), (24), and (28) and generate view rank, sales rank, and conditional share data.

Parameter Recovery To estimate the model, we take $I = 300$ draws from the distributions of random coefficients. For each individual draw of i, we compute the expected utilities V_{ij}, reservation utilities z_{ij}, and search probabilities π_{ij} for each j. The 300 values of $\{\pi_{ij}\}$, $i = 1, \ldots, I$, are aggregated and used to forecast the commonality indices, which are then used to forecast the view ranks. Given i's optimal search set, we compute i's choice probabilities using equations (5) and (6). For the conditional choice probabilities, we use equations (6) and (12). Finally, by aggregating the individual choice probabilities across consumers, we forecast market shares, sales ranks, and conditional shares. During the estimation, the log-likelihood in equation (32) with a penalty term for outside option share is maximized.

The parameter estimates and standard errors are shown in the column of "Proposed method" in Table 2. The standard errors were computed from repeated estimations over 10 different simulated data sets obtained from re-drawing the sampling errors in equations (19), (24), and (28). 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.[18]

Parameters		True value	Estimated value (s.e.) Proposed method	Frequency-based simulator $Q=40$	$Q=60$
Mean utility	Sony	0.5	0.55 (0.10)	0.64 (0.21)	0.53 (0.26)
	Pixel < 1 MB	−0.5	0.46 (0.13)	−0.45 (0.09)	−0.36 (0.17)
	Zoom > 10×	0.5	−0.51 (0.13)	0.50 (0.24)	0.55 (0.08)
	Media: mini DV	−0.5	−0.54 (0.12)	−0.70 (0.20)	−0.63 (0.26)
	Form: Compact	1	1.07 (0.14)	1.03 (0.14)	1.04 (0.17)
	Price	−1	−1.01 (0.21)	−0.73 (0.18)	−0.85 (0.26)
Heterogeneity	Sony	0.5	0.69 (0.17)	0.61 (0.34)	0.54 (0.40)
	Pixel < 1 MB	1	1.24 (0.11)	1.12 (0.34)	1.20 (0.38)
	Zoom > 10×	0.5	0.55 (0.11)	0.45 (0.25)	0.43 (0.23)
	Media: mini DV	1	1.18 (0.20)	1.22 (0.36)	1.22 (0.31)
	Form: Compact	0.5	0.46 (0.29)	0.34 (0.29)	0.35 (0.35)
	Price	0.5	0.50 (0.23)	0.58 (0.29)	0.44 (0.33)
Search cost	Base cost	−3	−3.05 (0.19)	−3.28 (0.37)	−3.49 (0.62)
	Effect of links	−3.5	−3.23 (0.16)	−3.54 (0.38)	−3.42 (0.62)
Aggregation error(s.d.)	τ_V	0.15	0.33 (0.03)	0.35 (0.04)	0.35 (0.039)
	τ_S	0.01	0.004 (0.003)	0.06 (0.09)	0.004 (0.002)
	τ_C	0.1	0.07 (0.02)	0.03 (0.08)	0.06 (0.05)
	MAD		0.09	0.12	0.12
	Mean s.e.		0.14	0.23	0.27
	Computation time (sec)		23.5	134.7	201.4

Table 2: Parameter estimates from the proposed method and kernel-smoothed frequency simulator

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

[18] In this exercise, we estimated the model without outside option to accommodate the comparison study with simulation-based methods. However, we also estimated the model with the outside option and confirmed that the model recovers the outside mean utility value of V_0 well. Please see Appendix C. For our exercise with outside option, we used a far higher number of draws, $I = 2,000$ to account for a very large fraction of consumers who search but do not buy.

(e.g., Chen and Yao, 2012; Ghose, Ipeirotis, and Li, 2012; Honka, 2014; Honka and Chintagunta, 2014). 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 integrals. 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 closely follow the implementation recipes detailed in Honka (2014) and Honka and Chintagunta (2014), who used it in an empirical model of sequential search and choice. The frequency estimator approximates the joint probability of $\Pr(j, S_K)$ in equation (6), which can also be defined as

$$\Pr(j,S_K) = \Pr\left[(u_{ij} > \max\{u_{i1},\cdots,u_{iK}\}) \cap (z_{in+1} > \max\{u_{i1},\cdots,u_{in}\}) \cap (u_{ij} > z_{iK+1})\right], \quad n = 1,\ldots,K-1. \tag{37}$$

Once we simulate this joint probability, we can compute the choice probability in equation (6), as well as the conditional choice probability in equation (16). Since the original frequency-based estimator is discontinuous, we smooth the above probability using,

$$\Pr(\omega;\lambda) = \frac{1}{1+\exp(\sum_m -\lambda\,\omega_m)}, \tag{38}$$

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

- 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) $\omega_{1i}^q = V_{ij} + e_{ij} - \max\{V_{i1} - e_{i1},\cdots,V_{iK} - e_{iK}\}$;
  - (selection rule) $\omega_{n+1,i}^q = z_{in+1} - \max\{V_{i1} - e_{i1},\cdots,V_{in} - e_{in}\}$, $n = 1,\ldots,K-1$;
  - (stopping rule) $\omega_{K+1,i}^q = V_{ij} + e_{ij} - z_{iK+1}$.

- Evaluate the joint probability of search and choice for i using equation (38),

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

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

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

Comparison with Frequency-Based Methods The results for the kernel-smoothed frequency-based estimations are shown in the two right columns of Table 2, with a varying number of draws Q used to integrate the joint probability in the simulation. We use several measures of comparison. We compute the mean (across parameters) absolute distance (MAD) between true vector of Θ^{true} and its estimate of $\hat{\Theta}$. In addition, we compute and report the average (across parameters) standard error for $\hat{\Theta}$. While the MAD computes the location of estimated parameters, the mean standard error tracks their efficiency. Better methods have lower values on both metrics. Lastly, we compare computation costs per evaluation of the goal function.

From the table, we note that our method outperforms kernel-smoothed frequency simulator because both MAD and average mean standard error are smaller for different values of Q. At the same time, its computation time is far lower than that from the kernel-smoothed frequency simulator.[19] We conclude that our proposed method achieves both higher accuracy and lower computational cost compared to kernel-smoothed frequency simulator in our empirical setting.

Comparison with Search-Only Models A related paper, Kim, Albuquerque, and Bronnenberg (2010) estimates search data only and study the consumer demand. That paper is also able to make choice predictions, but because the authors do not observe choice data, their estimation does not allow for the incorporation of an outside good. In addition, the combination of search and choice data in our model helps better identification and inferences. In theory, the model proposed here should achieve better market predictions that take

[19]This exercise was conducted on a desktop computer with Intel Core(TM) i7-2600 CPU with a clock speed of 3.4GHz and a RAM size of 16GB. For computational reasons, we did not attempt higher values than $Q = 60$ in this exercise.

into account both stages of the consumer buying process. We compare predictions and inferences between the two models in the section 4.2.

4 Empirical Illustration

4.1 Specification

In our empirical application, we represent a product as a bundle of characteristics and use the identical utility specification of Equations (34) and (35) in section (3.4). We include eight product characteristics: brand name, media format, form factor, high definition, zoom, screen size, number of pixels, and price (in thousand of dollars).[20] 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 independent across i and j.[21] 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

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

We 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 (36) in our data experiment. In our empirical application, the row vector of X_j^c is defined as,

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

where L_j is the number of times that product j appears on other products' pages in the category, I_j^{Hi} is an indicator variable if product j is a high-definition alternative, and A_j is the age of product j.[22] Links to a

[20] Consistent with our empirical setting, the attributes and their values are 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).

[21] The error term, e_{ij}, captures consumer's idiosyncratic 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).

[22] If additional information is available on how products are presented to consumers, our model is flexible enough to handle the

product page within the retailer's domain make that product more visible and are likely to lower search costs (Kim, Albuquerque, and Bronnenberg, 2010). Newer and more advanced products, such as a high-definition products, are featured more prominently by retailers, potentially leading to lower search costs.[23] Conversely, older products are no longer featured and become more difficult to find. To limit possible endogeneity of L_j that is constructed based on consumer browsing and firm choices, we lag this variable by a significant time (a month in our application).

Lastly, with respect to the possibility of price endogeneity in our utility specification, we allow for very flexible product fixed effects, which reduces the potential for large unobserved demand shocks which might be related to prices.[24]

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.

4.2 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 correlation between predicted and actual sales ranks for all $J = 90$ options is 0.703. The hit-rates of pairwise rank inequalities, in which we compare the relative positions of two options in the actual and predicted data are 75.4% for sales ranks and 87% for view ranks. As an external validation, we use out-of-sample data from September 2007 and compute similar measures. Across $J = 86$ products, the correlation between the predicted and actual sales ranks is 0.62 and the pair-wise hit rates are 72% for sales rank and 83% 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 previous literature that used search data only by estimating

effects of other non-utility components through the search cost function.

[23] In principle, we could have included all 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. This however would complicate the identification of search costs, as described earlier in the identification section. Instead, we opted to limit the variables that should enter the search cost based on past literature and navigational features at Amazon.com.

[24] Similar modeling assumptions are made in empirical models of optimal sequential search, for example in Kim, Albuquerque, and Bronnenberg (2010) and Chen and Yao (2012).

the model parameters following Kim, Albuquerque, and Bronnenberg (2010) using view-rank data alone. Their in-sample sales rank correlation and sales rank hit rates are 0.676 and 74.0%, respectively, which are marginally lower than our model predictions of 0.703 and 75.4%, evidence that using more data improves with respect to the search-only approach of Kim, Albuquerque, and Bronnenberg (2010). Their hit rate for the view-rank data is 89%, which is better than that from our model. This is understood since the proposed model uses both search and choice data while Kim, Albuquerque, and Bronnenberg (2010) uses search data only.

As a robustness check, we re-estimate the empirical model with the outside good share values of 85% and 92%. 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 (0.998). Therefore, our parameter estimates seem robust to small changes in outside good shares.

Parameter Estimates We present the parameter estimates in Table 3. 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 2.24, while Panasonic, another popular brand, has the second highest mean value of 1.94.[25] The estimates show significant heterogeneity in brand preferences with an estimate of 0.62 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 to the negative coefficients of other media options: MiniDV is the next preferred option (-0.81) and Flash Memory (-1.55) is the least preferred. Heterogeneity for media formats is also high, with its estimate of standard deviation at 0.76. Compactness negatively influences the mean utility and consumers prefer higher number of pixels (0.17). Finally, we report that our price coefficient estimates imply an average own price elasticity of -2.11.

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 full advantage of the high-definition feature of camcorders. In December of 2007, the same year of our data, the average price of a 40-inch LCD HD TV was about $\$1,500$ (Magid 2007).[26] Heterogeneity in preferences for high-definition is especially high

[25] We normalize the coefficient for brand Sharp at zero and estimate the mean utility of the outside good V_0.

[26] In contrast, an informal inspection for the most popular 42 inch LCD HD TV sold at Amazon.com in summer 2014, yields

Primitive	Variable	mean ($s.e.$)	heterogeneity($s.e.$)
Utility	Sony	2.24 (0.15)	0.62 (0.03)[a]
	Panasonic	1.94 (0.15)	0.62 (0.03)
	Canon	1.85 (0.14)	0.62 (0.03)
	JVC	1.64 (0.15)	0.62 (0.03)
	Samsung	1.03 (0.19)	0.62 (0.03)
	Media: MiniDV	−0.81 (0.09)	0.76 (0.05)[b]
	Media: DVD	−1.25 (0.13)	0.76 (0.05)
	Media: Flash Memory	−1.55 (0.41)	0.76 (0.05)
	Compact	−0.78 (0.29)	1.37 (0.19)
	High-Definition	−0.60 (0.17)	1.07 (0.09)
	Zoom	−0.03 (0.01)	0.03 (0.004)
	Pixel	0.17 (0.04)	0.17 (0.02)
	log (Price)	1.32 (0.09)	0.90 (0.04)
	V_0(outside good)	4.71 (0.22)	
Search cost	Base search cost (γ_0)	−6.22 (0.62)	
	Number of links	−0.92 (0.11)	
	High-Definition	−3.60 (0.34)	
	Age	0.55 (0.10)	
Aggregation	View-rank	0.13 (0.01)	
error	Sales-rank	0.001 (0.009)	
std. dev.	Conditional share	0.20 (0.12)	
	Log-likelihood	−57,241	

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

Table 3: Estimates of the model parameters for the Camcorder category

compared to other attributes, which is justified by the innovative nature of this attribute.

Inferences In terms of the search cost, a higher number of links to a product page reduces its search cost (-0.92), as a page may be easier to access. High-definition products have a lower search cost as well, which is likely driven by its innovative nature, leading this products to stand out among more traditional products. Across products and consumers, our estimated parameters lead to a mean and median for search costs of \$2.52 and \$0.46 respectively.[27]

Using these parameters, we estimate a median and modal search set size conditional on choice of 14 and 7, respectively. This suggests that the average search set includes a small fraction of the $J = 90$ products in our data. In addition, a relatively large consumer search set size implies that our parsimonious joint model is

prices between \$350 and \$450.

[27]Monetary search costs can be computed by dividing the search cost by the price coefficient (Honka and Chintagunta 2014).

a more feasible estimation framework for similar categories, while a full-simulation based estimation may be challenging to implement.

Lastly, we compute consumer's willingness to pay for various product attributes and compare their values from Kim, Albuquerque, and Bronnenberg (2010). First, we report the median brand premium of Sony over Panasonic across consumers from our model is \$81 while the corresponding value from Kim, Albuquerque, and Bronnenberg (2010) is \$613. In addition, the consumers' median willingness to pay for additional 1 MB of pixel is \$46 in the proposed model, while its counterpart in Kim, Albuquerque, and Bronnenberg (2010) is \$608. Given the price range of camcorders in Table 1, we conclude that the proposed model using both search and choice data and allowing for an outside option leads to more realistic inferences compared to a search-only model.

4.3 Prediction Exercises

In this subsection, we conduct two 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 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.[28] Table 4 shows the substitutes with the highest and second highest cross-elasticities for selected models. We find that the substitute products share some or all attributes or have similar attribute values with the focal product. For instance, the two best substitutes for a Sony Hard Drive product with a retail price of \$601 are other products either with same brand or media type, in a similar price range. As another example, the best substitutes for a compact Sharp product with FM media selling at \$594 are other compact camcorders from Sharp and Samsung. We believe these implied consumer substitution patterns are realistic in the camcorder

[28] In order to study substitution patterns based on consumer preferences only, we set search cost 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.

Focal Product	Price	Own Elasticity	Best Substitute	Price	Cross Elasticity	Second Best Substitute	Price	Cross Elasticity	% to outside goods
Sony, Hard Drive	\$601	−3.50	JVC, Hard Drive	\$510	0.61	Sony, DVD	\$497	0.39	3.3
Canon, MiniDV	\$289	−2.61	JVC, MiniDV	\$309	0.82	Canon, MiniDV	\$230	0.32	3.4
Canon, DVD	\$300	−2.51	Sony, DVD	\$351	0.18	Canon, DVD	\$568	0.05	3.4
Sony, MiniDV Hi-def	\$1,397	−4.00	Panasonic, DVD Hi-def	\$1,011	0.41	Sony, Hard-Drive Hi-def	\$1,178	0.14	0.8
Sharp, FM Compact,Hi-def	\$594	−1.79	Sharp, FM Compact	\$436	0.55	Samsung, FM Compact	\$308	0.29	3.0

Table 4: Price-induced substitution patterns of selected camcorders

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 decides to withdraw some of its least popular products and streamline its portfolio. In the past, Sony seems to have 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 brands. Table 5 shows the results. Before withdrawal, Sony's total market share is 43.76%, with 39.25% coming from the top 24 products and 4.51% from the bottom seven options. After the decision to drop the bottom seven products, Sony's new market share is 41.66%. Therefore, instead of losing 4.51% of the market share from its bottom seven products, because of internal substitution Sony is predicted to lose only 2.10%. The list of products that gain market share is also shown in the same table. For instance, a Sony miniDV camcorder with price of \$529 gains a market share of 0.29%. Sony can replicate this analysis with other products to find the best set to discontinue.

Sony's share changes before and after the decision					
			Old share	New share	Share gain
Top 24	Sony		39.25%	41.66%	2.41%
Bottom 7	Sony		4.51%	0%	−4.51%
All	Sony		43.76%	41.66%	−2.10%

Products with the largest share changes					
Brand	Media type	Price	Old share	New share	Share gain
Sony	miniDV	\$529	4.46%	4.74%	0.29%
JVC	Hard Drive	\$677	7.72%	7.94%	0.22%
Sony	Hard Drive	\$644	4.49%	4.70%	0.21%
Sony	Hard Drive	\$778	4.44%	4.64%	0.20%
Canon	miniDV	\$250	4.53%	4.73%	0.20%

Table 5: Top market share gainers from Sony's short product portfolio

4.4 Market Structure Analysis

Analyzing the market structure is informative for product line managers concerned with the positioning of their brands in the marketplace (Van Heerde, Mela, and Manchanda 2004). We illustrate next how our proposed model provides insights about market structure in a durable goods category, in which research is far less common compared to 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, Mela, and Manchanda, 2004; Sonnier, McAlister, and Rutz, 2011).[29] 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 asymmetric competitive positions among the brands (Kamakura and Russell 1989). Sony shows the highest level of clout and lowest level of vulnerability, which may in part be attributable to its highest brand recognition and reputation for quality, although clouts of Canon, JVC, and Panasonic are not far. JVC offers 14 products, eight 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

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

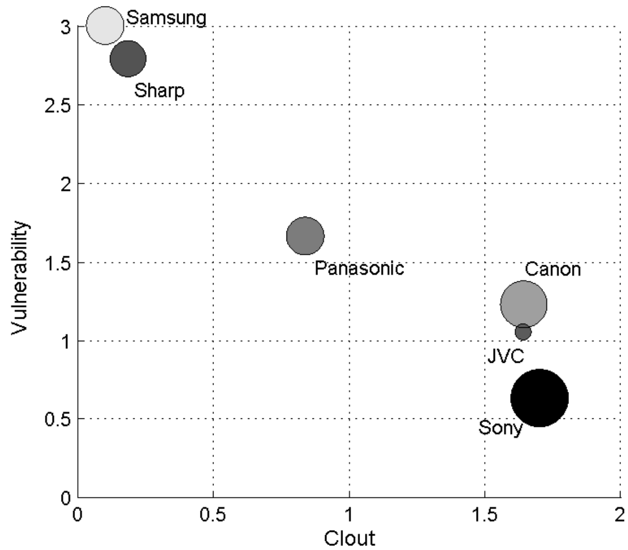

Figure 2: Clout and vulnerability for camcorder manufacturers (larger circles represent higher sales and darker circles represent higher prices).

the high clout and low vulnerability. In 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.

5 Conclusion

In this paper, we propose a theory-based, joint model of optimal sequential search and choice. Based on 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. 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, under the withdrawal of a few products, and the market structure at the brand 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.

References

Albuquerque, Paulo and Bart J. Bronnenberg. 2009. "Estimating demand heterogeneity using aggregated data: An application to the frozen pizza category." *Marketing Science* 28 (2):356–372.

Anderson, Simon P and Regis Renault. 1999. "Pricing, product diversity, and search costs: A Bertrand-Chamberlin-Diamond model." *The RAND Journal of Economics* :719–735.

Bajari, Patrick, Jeremy T Fox, and Stephen P Ryan. 2007. "Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients." *The American economic review* :459–463.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated products demand systems from a combination of micro and macro data: The new car market." *Journal of Political Economy* 112 (1):68–105.

Bresnahan, Timothy F. 1987. "Competition and collusion in the American automobile industry: The 1955 price war." *The Journal of Industrial Economics* :457–482.

Bronnenberg, Bart J., Jun B. Kim, and Carl F. Mela. 2014. "Zooming in on choice: How do consumers search for cameras online?"

Chen, Yubo and Jinhong Xie. 2008. "Online consumer review: Word-of-mouth as a new element of marketing communication mix." *Management Science* 54 (3):477–491.

Chen, Yuxin and Song Yao. 2012. "Search with refinement." Working paper.

Draganska, Michaela and Daniel Klapper. 2011. "Choice set heterogeneity and the role of advertising: An analysis with micro and macro data." *Journal of Marketing Research* 48 (4):653–669.

Efron, Bradley and Robert J Tibshirani. 1994. *An introduction to the bootstrap*, vol. 57. CRC press.

Eisenberg, Bryan. 2009. "Top 10 online retailers by conversion rate: June 2009." URL `http://www.grokdotcom.com/2009/08/03/top-10-online-retailers-by-conversion-rate-june-2009/`.

Elrod, Terry. 1988. "Choice map: Inferring a product-market map from panel data." *Marketing Science* 7 (1):21–40.

Erdem, Tülin. 1996. "A dynamic analysis of market structure based on panel data." *Marketing Science* 15 (4):359–378.

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li. 2012. "Surviving social media overload: Predicting consumer footprints on product search engines."

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

Gowrisankaran, G. and M Rysman. 2012. "Dynamics of consumer demand for new durable goods." *Journal of Political Economy* 120:1173–1219.

Hancox, Paul. 2008. “Is Amazon’s 9.6 % conversion rate low? Here’s why I think so...” URL `http://www.conversionblogger.com/is-amazons-96-conversion-rate-low-heres-why-i-think-so/`.

Hofilena, John. 2014. “Sony to cut more jobs at failing electronic equipment division.” URL `http://japandailypress.com/sony-to-cut-more-jobs-at-failing-electronic-equipment-division-0341844/`.

Honka, Elisabeth. 2014. “Quantifying search and switching costs in the US auto insurance industry.” *The Rand Journal of Economics* forthcoming. Available at SSRN 2023446.

Honka, Elisabeth and Pradeep K. Chintagunta. 2014. “Simultaneous or sequential? search strategies in the US auto insurance industry.” Working paper.

Kamakura, Wagner A and Gary J Russell. 1989. “A probabilistic choice model for market segmentation and elasticity structure.” *Journal of Marketing Research* :379–390.

Kim, Jun B, Paulo Albuquerque, and Bart J Bronnenberg. 2010. “Online demand under limited consumer search.” *Marketing Science* 29 (6):1001–1023.

Koulayev, Sergei. 2014. “Estimating demand in online search markets, with application to hotel bookings.” Forthcoming.

Magid, Larry. 2007. “Today’s HDTV, or Next Year’s?” URL `http://www.nytimes.com/2007/12/20/technology/personaltech/20basics.html`.

McFadden, Daniel. 1989. “A method of simulated moments for estimation of discrete response models without numerical integration.” *Econometrica: Journal of the Econometric Society* :995–1026.

Moraga-González, José Luis, Zsolt Sándor, and Matthijs R Wildenbeest. 2012. “Consumer search and prices in the automobile market.”

Petrin, Amil K. 2002. “Quantifying the benefits of new products: The case of the minivan.” *Journal of Political Economy* 110 (4):705–729.

Sonnier, Garrett P, Leigh McAlister, and Oliver J Rutz. 2011. “A dynamic model of the effect of online communications on firm sales.” *Marketing Science* 30 (4):702–716.

Train, Kenneth. 2009. *Discrete choice methods with simulation*. Cambridge, UK: Cambridge university press.

Van Heerde, Harald J, Carl F Mela, and Puneet Manchanda. 2004. “The dynamic effect of innovation on market structure.” *Journal of Marketing Research* 41 (2):166–183.

WalkerSands. 2014. “Reinventing retail: What businesses need to know for 2014.”

Weitzman, Martin L. 1979. “Optimal search for the best alternative.” *Econometrica: Journal of the Econometric Society* :641–654.

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. The following table summarizes the availability of such data for some of the largest online retailers as of April 2014.

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

Table A.1: Summary of online retailers and data availability

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 $u_\ell = V_\ell + e_\ell$, 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 2.2 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 forego 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

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

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

$$\Pr(j=2|\,S_K) = \int_{-\infty}^{\infty} \prod_{\ell \neq 2}^{3} \Phi_\ell\left(V_2 - V_\ell + e_2\right) \phi_2(e_2) de_2. \tag{B.3}$$

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

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

C Additional Data Experiment

In this section, we report the results of data experiment with an outside option. We set the outside good utility as

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

where we set the value of V_0 such that outside goods share is 89% and set σ_0^2 to 1 for identification purpose. The overall set-up, data generation and parameter recovery for the experiment is identical to those in Section 3.4 except the following. First, for data generation we draw $250{,}000$ "pseudo-households" from the joint distribution of parameters. Second, for parameter recovery, we take $I = 1{,}500$ draws from the distributions of random coefficients. A higher number of draws are needed to account for a small fraction of consumers

Parameter		True values	Estimated values (s.e.)
Mean utility	Sony	0.5	0.48 (0.05)
	Pixel $<$ 1 MB	-0.5	-0.65 (0.08)
	Zoom $>$ 10$\times$	0.5	0.50 (0.06)
	Media: mini DV	-0.5	-0.44 (0.04)
	Form: Compact	1	0.93 (0.05)
	Price	-1	-1.12 (0.06)
Heterogeneity	Sony	0.5	0.45 (0.06)
	Pixel $<$ 1 MB	1	1.01 (0.05)
	Zoom $>$ 10$\times$	0.5	0.47 (0.05)
	Media: mini DV	1	0.98 (0.06)
	Form: Compact	0.5	0.51 (0.05)
	Price	0.5	0.36 (0.11)
Search cost	Base cost	-3	-3.14 (0.12)
	Effect of links	-3.5	-3.48 (0.10)
	Outside good V_0	4.3	4.32 (0.06)
Aggregation	τ_V	0.15	0.19 (0.01)
error	τ_S	0.01	0.001 (0.0003)
std. dev.	τ_C	0.10	0.16 (0.03)

Table C.2: Estimation results from the numerical experiment with an outside option.

who buy from the category upon search. Table C.2 shows that the parameters including mean outside utility of V_0 are well recovered, all close to their actual values and within sampling errors.