Large-scale Demand Estimation with Search Data

Tomomichi Amano
Andrew Rhodes
Stephan Seiler

Working Paper 19-022

Large-scale Demand Estimation with Search Data

Tomomichi Amano
Harvard Business School

Andrew Rhodes
Toulouse School of Economics

Stephan Seiler
Stanford University

Working Paper 19-022

Large-scale Demand Estimation with Search Data

Tomomichi Amano
Columbia University

Andrew Rhodes
Toulouse School of Economics

Stephan Seiler
Stanford University

This draft: May 16, 2018

Preliminary and incomplete, please do not circulate without the authors' permission.

Many online markets are characterized by sellers that stock large numbers of products and sell each product infrequently. At the same time, consumer browsing information is typically tracked by online retailers and is much more abundant than purchase data. We propose a demand model that caters to this type of setting. Our approach, which is based on search and purchase data, is computationally light and allows for flexible substitution patterns. We apply the model to a data set containing browsing and purchase information from a retailer stocking over 500 products, recover the elasticity matrix, and solve for optimal prices for the entire assortment.

Keywords:

1 Introduction

Demand estimation is a cornerstone of empirical IO and quantitative marketing and the basis for optimal price setting, merger simulations, and many other applications. However, many demand models are computationally burdensome to estimate, especially when the number of products is large, a situation that is increasingly common in online markets. The main contribution of this paper is to propose a model of consumer search and choice that is simple to estimate and can flexibly capture substitution patterns in markets with large numbers of products. We propose leveraging rich information on consumers' browsing behavior prior to a purchase, which is typically recorded in online markets. This novel source of information reveals to the researcher the set of products the consumer searched and considered for purchase and hence provides additional information on consumer preferences and substitution patterns.

Our proposed model allows us to achieve a high degree of flexibility with regards to substitution patterns while remaining computationally light. In terms of computational burden, our model is comparable to a (random coefficient) logit model for a market with small choice sets. In contrast, most structural search papers which have used browsing data (or similar data on considered / searched products) tend to be computationally burdensome and cannot be easily scaled to large choice sets. A second advantage of our approach lies in the fact that we are able to remain agnostic with regards to assumptions on information sets, search protocol, etc. that are typically necessary to estimate structural models of consumer search. Our approach nests many modeling approaches in the search and consideration set literature and can be understood as a reduced-form approximation to others.

The strength of our model lies in estimating demand for markets with a large number of products, but sparse purchases. The additional information gleaned from the search data together with the computationally light approach makes the model easy to estimate and yet able to flexibly recover substitution patterns. These features make it a powerful tool to inform price setting by online retailers. In our empirical application, we use browsing and purchase data for a retailer carrying over 500 products and estimate the elasticity matrix for the entire assortment. We then use our estimates to solve for the optimal price vector for the entire assortment.

We set up our model of search and choice as follows. We assume that the consumer forms a consideration set based on "some process" and then chooses the utility maximizing product from this set. Crucially, we avoid explicitly modeling the process by which consideration sets are formed, and instead treat consideration set probabilities directly as objects to be estimated. The two-stage setting does not preclude us from allowing for dependencies between consideration probabilities and purchase decisions (which are a feature of many search models). Absent any data restrictions, we can allow preference parameters (that drive choice conditional on consideration) to be specific to each possible consideration set. In practice, we allow for a more limited amount of flexibility and estimate different preference parameters for certain "types" of consideration sets (rather than different parameters for each consideration set). Furthermore, to make the model tractable, we assume that price enters the choice process, but does not drive consideration. The exclusion of price from the consideration stage is a common assumption in many (but not all) consideration set and search models.

We jointly estimate consideration set probabilities and the parameters of the choice process conditional on consideration. We then infer price elasticities based on these model primitives. Crucially, price elasticities are simple to compute because consideration set probabilities remain unchanged in response to price changes due the exclusion of price from the consideration process. Our estimated consideration set probabilities are simply given by the empirical frequencies of the various consideration sets occurring in the data. Therefore, because our estimation sample exhibits (by construction) the estimated consideration frequencies, we do not have to store the consideration set probability vector during estimation. We simply condition on the observed consideration sets in our sample when counterfactually changing prices to compute elasticities. This "computational trick" allows us work with a high-dimensional vector of estimated consideration set probabilities without adding any computational burden.

The way that consideration sets inform substitution patterns in our model is most easily understood with the following example. Consider an illustrative market containing 10 products as displayed in Figure 1. Assume that focal product A is either searched together with products B,C,D and E in the top row (illustrated by the solid rectangle) or product F (illustrated by the dashed rectangle). However, product A is never searched together with any other product. Therefore a price increase for any product outside of the set of co-searched products (products G, H, I, J) will not affect demand for A, whereas price changes for products B,C,D, E and F will. Search patterns such as the ones illustrated in Figure 1 are likely driven by similarity along some characteristic dimension. For instance one could imagine that our illustrative markets contains automobiles with high horsepower cars in the top row (the bottom row contains low horsepower cars) and the first column contains red cars (other columns correspond to different colors). A major advantage over perfect information demand models based on purchase data is that the groupings of substitute products are directly observed in the data and do not need to be inferred indirectly from consumers' purchases. Furthermore, there is no need to define the relevant set of characteristics that might drive the substitutability of products. Instead, the data directly reveals any relevant product groupings. In particular when the number of products is large, the search data can help to uncover groups of substitute products in a flexible way without the need to define the (potentially large number of) characteristics driving product groupings.

We apply our demand model to data on browsing (i.e. search) and purchase behavior in one large product category of an online retailer. The category contains almost 600 products. We obtained data for a 3 month period which contains 470,000 product searches by 186,000 users and 13,000 purchases. Apart from search and purchase data we also obtained data on prices and other (time-invariant) product characteristics which we allow to enter the conditional utility function. We estimate three models with varying degrees of flexibility in the specification of the conditional utility function. To assess whether the additional flexibility matters, we assess how the optimal price vector changes when enriching the model and find the changes minor between

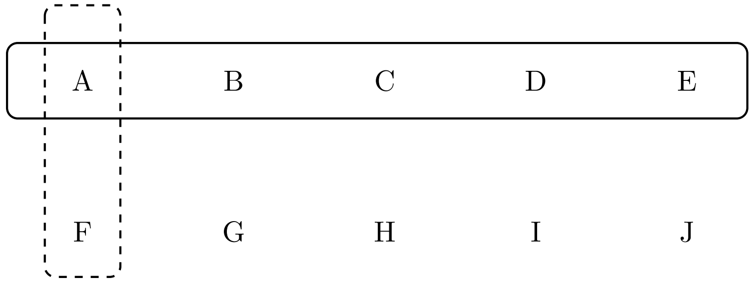

Figure 1: **Illustrative Search Patterns.**

the two most flexible specifications. Interestingly, optimal prices derived from the model estimates are significantly higher than current prices of the retailer. This seems to at least partly derive from the fact that the firm did not previously take cross-product substitution effects into account when setting prices. We are currently in the process of exploring ways to validate our estimates by running selective price experiments and to implement price changes based on our estimates.

Our paper relates to several distinct streams of literature. First, it relates to a literature that employs non-structural approaches based on various sources of data in order to uncover and visualize substitution patterns among products. Netzer, Feldman, Goldenberg, and Fresko (2012) use the co-occurrence of products mentioned in online discussion forums. Lee and Bradlow (2011) employ a related approach based on customer reviews and the products mentioned in them. The two papers in this realm that are closest to our approach are Ringel and Skiera (2016) and Kim, Albuquerque, and Bronnenberg (2011), which both use online search / browsing data in order to to analyze competitive market structure and product substitution. Neither of the papers estimates an elasticity matrix despite the focus on substitution patterns, nor are the measures of "relatedness" of products derived from an underlying model of optimal consumer behavior. Instead, these papers provide a visualization of closeness in product space with an implicit understanding (but no formal derivation of such a relationship) that these help us learn about substitution patterns and hence elasticities of demand.

A second relevant stream of literature comprises papers modeling consideration sets as well as structural models of consumer search. To the best of our knowledge, there has been no clear taxonomy to categorize consideration and search models. Broadly speaking, one can think of consideration set models as containing two separate stages, consideration and choice, whereas search models are based on a unified utility maximization framework that underpins both search and purchase decisions. The first stage in a consideration model is typically either characterized as a more passive stage in which consumer become aware of products due to external factors such as advertising, product displays etc. (Bronnenberg and Vanhonacker (1996), Mehta, Rajiv, and Srinivasan (2003), Pancras (2010)) or can be understood as a reduced-form of a structural search model. To organize the literature and the relation of different papers to our approach, we adopt a "two stages versus unified optimization framework" way of categorizing the two approaches.

The literature on consideration sets typically considers cases in which consideration sets are not observed[1] and hence they are treated as a latent process. Typically identification stems from exclusion restrictions.[2] For instance Goeree (2008) assumes that consideration is driven by advertising which does not enter the consumer's utility in the purchase stage. Andrews and Srinivasan (1995), Bronnenberg and Vanhonacker (1996), and Mehta, Rajiv, and Srinivasan (2003), assume that product displays and feature advertising only affect consideration, but not utility. In general, consideration is modeled as a function of observable product characteristics (Andrews and Srinivasan (1995), Bronnenberg and Vanhonacker (1996), Goeree (2008), Barroso and Llobet (2012), Gaynor, Propper, and Seiler (2016)), which implies that any co-occurrence of products in consumers' consideration set has to be rationalized by observables. For instance high horse-power cars being considered together can be captured by allowing for horse-power (possibly interacted with demographics) to affect the consideration probability. It is therefore observed characteristics that determine how often products are considered together and hence whether they are close substitutes or not. Despite a different non-linear functional form, substitution patterns ultimately trace back to observable characteristics just as in the case of perfect information demand models.

Structural models of consumer search (such as Kim, Albuquerque, and Bronnenberg (2010), De Los Santos, Hortacsu, and Wildenbeest (2012), Honka (2014) and Chen and Yao (2016))[3] instead present a unified framework of consumers' utility maximization that rationalizes observed search and purchase patterns. Typically, these models are relatively computationally burdensome and most of the models above are therefore estimated for markets with a relatively small number of products. Our approach is computationally lighter, but this comes at the cost of not modeling the search process explicitly. Furthermore, in search models, both search and choice are driven by the specified utility function. Typically utility is defined in characteristics space and hence, similar to the two-stage approach with independent consideration probabilities, cross-elasticities are determined by the observable characteristics that enter utility. Our approach instead allows us to remain agnostic about the process that drives consideration and this in turn allows more flexibility with regards to substitution patterns.

Finally, there is an emerging literature on large-scale demand estimation that is based on purchase data and does not make use of additional data sources such as search / browsing information. Smith, Rossi, and Allenby (2016) use a Bayesian approach to flexibly estimate market partitions using supermarket scanner data. Chiong and Shum (forthcoming) use sparse random projection to reduce the dimensionality of the estimation problem.

The remainder of the paper is structured as follows. We first outline our model of search and

[1] There are a few exceptions (Roberts and Lattin (1991), Draganska and Klapper (2011)) where consideration sets are observed, and usually these are stated consideration sets obtained via a survey among consumers.

[2] In a recent paper, Abaluck and Adams (2016) base identification of consideration sets on asymmetries in cross-derivatives. Kawaguchi, Uetake, and Watanabe (2016) use variation in product availability to identify consideration sets.

[3] Other papers within the broader literature of consumer search include Seiler (2013), Koulayev (2014), Honka and Chintagunta (2016), Pires (2016) and Haviv (2016).

describe estimation, the derivation of elasticities and calculation of optimal prices. In Section 3 we describe the data. In Section 4 we apply the model to data from an online retailer, estimate the elasticity matrix for an assortment of almost 600 products, and solve for profit-maximizing prices. Section 5 compares our approach to other models of consideration set formation and consumer search as well as perfect information demand models. Finally, we provide some concluding remarks.

2 Model Framework

We start by outlining a general model of consideration and choice that nests approaches taken in the consideration set and consumer search literatures. We then introduce the relevant assumptions needed to make the model tractable and describe estimation. Finally, we show how price elasticities are computed based on parameter estimates and how we can use the estimates to solve for optimal prices.

2.1 General Model

Consider a setting with J differentiated products. Each consumer wishes to buy at most one product, and makes her choice according to the following two-step procedure. In the first stage, a consumer narrows down her consideration set to a subset of the products. We are agnostic about how the consumer's consideration set is formed. Since the consumer can choose not to buy, the outside option always appears in her consideration set. Let S denote the set of all possible subsets of the products that contain the outside option, and let s be an element from that set. We assume that in the population consideration sets follow a discrete probability distribution $\Delta(s)$.

In the second stage, each consumer picks the product (or outside option) from her consideration set which gives her the highest utility. Precisely, for consideration set s let $u_{0|s}$ denote the utility associated with the outside option, and let $u_{j|s}$ be the utility from buying product $j \in s$. Also let $F_s(\mathbf{u})$ denote the joint cumulative distribution over the utilities of items contained in consideration set s. We can then write the probability that a consumer with consideration set s buys product $j \in s$ as

$$Pr_{j|s} = \int \mathbf{1}(u_{j|s} \geq u_{k|s} \forall k \in s) \mathrm{d}F_s(\mathbf{u}) \,,$$

where $\mathbf{1}(.)$ is an indicator function that takes the value 1 when element j offers the highest utility. Demand for product j is computed by summing the conditional discrete choice demands over all consideration sets which include product j:

$$D_j(\mathbf{p}) = \sum_{s \ni j} \Delta(s) \times Pr_{j|s} \,, \tag{1}$$

where $\mathbf{p}$ denotes the vector of prices.

Notice that because the joint distribution over utilities $F_s(\mathbf{u})$ can vary across consideration sets, this general model allows for an arbitrary link between consideration and choice. Hence the above formulation nests any model of consumer search that bases search and choice on a common utility function. Here, rather than modeling the search process directly, we simply model utility conditional on consideration in a flexible way. We also note that our formulation so far does not require us to take a stance on consumers' information sets before and after search, or other details of the search process such as search protocol (sequential or simultaneous). Furthermore, the model is consistent with other ways of forming consideration sets, for instance through choice heuristics.

Starting from this general formulation, we introduce one important assumption that is required to make the model tractable. Namely, we assume that price does not affect consideration set formation. Price not affecting consideration could be either because price is not known to consumers when they form consideration sets, or because consumers choose not to use price information. It is a common assumption in many models of consumer search and consideration set formation (e.g. Goeree (2008), Honka (2014), Pires (2016)) but not all (e.g. Kim, Albuquerque, and Bronnenberg (2010) assume price is known prior to search). We discuss this further in Section 5.

We can write the semi-elasticity of demand for product j with respect to the price of product k as

$$\eta_{j,k} = \frac{\sum_{s \ni j,k} \Delta(s) \times \frac{\partial Pr_{j|s}}{\partial p_k}}{\sum_{s \ni j} \Delta(s) \times Pr_{j|s}}. \tag{2}$$

Notice that $\Delta(s)$ does not change with price due to the assumption that price does not affect consideration.

In order to illustrate which factors determine semi-elasticities given by equation (2), consider Figure 1 from earlier. Suppose we are interested in establishing whether product A competes more fiercely with product B or with product F. A natural way to measure this is to compare the magnitudes of $\eta_{A,B}$ and $\eta_{A,F}$. Using equation (2) it is clear that other things equal $\eta_{A,B}/\eta_{A,F}$ is larger when $\sum_{s \ni A,B} \Delta(s)/\sum_{s \ni A,F} \Delta(s)$ is larger i.e. when A and B are considered together more often than are A and F. However it is also clear from (2) that in general this joint viewing probability ratio is not sufficient to determine which pair of products compete more closely. Intuitively the size and composition of the rest of the consideration set also matters. For example returning to Figure 1, it may be that A and F appear relatively infrequently in the consideration set (compared to A and B), however when they do appear they are the only products in it (whereas when A and B appear together, so do other products C, D and E which are often preferred by consumers over B). Consequently it is possible for F to exert more of a competitive influence on A than does B.

In summary cross-price elasticities in our model are determined by both the frequency of co-occurrence in consideration sets, as well as by substitution patterns within the considerations sets that contain the pair of products. Allowing for both forces to influence elasticities makes our model a hybrid between two existing approaches. On the one hand, perfect information demand models do not typically incorporate search information and hence substitution patterns are driven by preference parameters alone. At the other extreme a small literature (which we discussed in the introduction) uses summary statistics of the search process (or related similarity measures) such as the frequency of co-occurrence to describe the degree of substitutability between products.

Our proposed model allows for consideration sets and preferences (conditional on consideration) to jointly determine elasticities.

2.2 Utility Specification and Choice Probabilities

In order to take the model to the data we need to specify $F_s(\mathbf{u})$, the joint distribution function which arises under each consideration set s. We assume that after deciding which set of products to search s, consumer i obtains the following utility from purchasing a given product j:

$$\begin{aligned} u_{ij|s} &= \delta_{js} - \alpha_{is} p_{jt} + \beta_{is} X_j + \varepsilon_{ij} = \bar{u}_{ij|s} + \varepsilon_{ij} \\ u_{i0|s} &= \varepsilon_{i0}, \end{aligned}$$

where δ_{js} denotes the intercept for product j in consideration set s. p_{jt} denotes the price of product j on day t and α_{is} is a consumer and consideration set specific price coefficient. X_j is a vector of product attributes and β_{is} is a vector of (i,s)-specific coefficients. Finally, ε_{ij} denotes a taste shock that is iid extreme value. Utility (excluding the taste shock) is normalized to zero for the outside option.

This setting corresponds to a logit model with product fixed effects and consumer heterogeneity in preferences over price and other product characteristics. Heterogeneity can be modeled along both observed and unobserved dimensions. The one difference to demand models without a consideration stage is the fact that we allow preferences to be a function of the consideration set s. As we outline in more detail in Section 5, many models of consumer search entail such a dependence between consideration set choice and preferences. Here we can allow for such dependence albeit in a post-hoc and non-theoretical fashion. We note however that, with sufficient data, this dependence can be made arbitrarily flexible i.e. in principle one could estimate a model which allows each consideration set to have different preference parameters. Of course in a market with many products, some consideration sets may occur only very infrequently. Consequently in our application, we will eventually specify a more parsimonious parameterization regarding the heterogeneity of coefficients across consideration sets.

Given the above utility function, the probability of the consumer choosing product j from a consideration set s that contains the product is given by the standard expression

$$Pr_{ij|s}(\Omega_{is}) = \frac{\exp(\bar{u}_{ij|s})}{1 + \sum_{l \in s} \exp(\bar{u}_{il|s})},$$

where $\Omega_{is} = (\alpha_{is}, \beta_{is}, \delta_{ls} \forall l \in s)$ denotes the set of (i,s)-specific preference parameters. When including unobserved heterogeneity, the equivalent expression would be derived by numerically integrating out unobserved preferences. We estimate this conditional choice model by maximum likelihood.

Apart from estimating conditional preference parameters in this way, one also needs to view the observed consideration set frequencies in our sample as estimates of their population counterparts. Therefore we implicitly estimate a high-dimensional vector of consideration set probabilities. However the specific structure of our model allows us to handle this aspect in a computationally convenient way that imposes no additional computational burden. Specifically, the counterfactual price changes that allow us to infer elasticities, exhibit the same conditioning on consideration sets as the conditioning used during estimation. Consequently we can simply apply price changes to the estimation sample of consumers while holding consideration sets fixed. The sample of consumers used in estimation (trivially) exhibits the estimated consideration set frequencies. Hence there is no need to store the consideration set probability vector during estimation.

2.3 Elasticities

As described earlier, expected demand for product j is given by

$$D_j = \sum_{s \ni j} \Delta(s) Pr_{j|s}(\Omega_{is}).$$

We can derive an estimate of $Pr_{j|s}$ from the estimated parameters of the conditional utility function Ω_{is}. Note that $Pr_{j|s}$ is obtained by averaging over the consumer-specific terms $Pr_{ij|s}$ within a given consideration set. In particular $Pr_{j|s} = \frac{1}{N_s} \sum_i \mathbf{1}(s_i = s) Pr_{ij|s}$, where N_s denotes the number of consumers with consideration set s and $\mathbf{1}(s_i = s)$ is an indicator function equal to one if consumer i choose consideration set s. The estimate of $\Delta(s)$ is given by the frequency estimator $\frac{1}{N} \sum_i \mathbf{1}(s_i = s) = \frac{N_s}{N}$. This allows to re-write demand as follows

$$\begin{aligned} D_j &= \sum_{s \ni j} \frac{N_s}{N} \left[\frac{1}{N_s} \sum_i \mathbf{1}(s_i = s) Pr_{ij|s}(\Omega_{is}) \right] \\ &= \frac{1}{N} \sum_i \sum_{s \ni j} \mathbf{1}(s_i = s) Pr_{ij|s}(\Omega_{is}) \\ &= \frac{1}{N} \sum_i \mathbf{1}(j \in s_i) Pr_{ij|s}(\Omega_{is}), \end{aligned}$$

where $\mathbf{1}(j \in s_i)$ is an indicator function equal to one if consumer i's consideration set s_i contains product j.

The derivative of demand for product j with respect to the price of product k is therefore equal to

$$\frac{\partial D_j}{\partial p_k} = \frac{1}{N} \sum_i 1(j \in s_i) \frac{\partial Pr_{ij|s}}{\partial p_k},$$

which is easily computed because consideration set formation is unaffected by the price change.

2.4 Optimal Prices

With the demand estimates and derivative matrix in hand, we now turn to solving for the vector of optimal prices. The business model of the retailer is such that some of the products in the category we study are "retail" products and others are "marketplace" products. The former are purchased by the retailer at a specific wholesale price and then resold to consumers, while the latter are items that are sold through the platform by third-party sellers. The retailer earns a commission (i.e. a percentage of the purchase price) for each sale of a marketplace product. We focus on the problem of setting retail prices optimally while taking the commission rate as given. The firm's optimization problem is given by

$$\max_{p^R} \; Profit = \sum_{j \in R} (p_j - mc_j) D_j(p^R, p^{MP}) + \sum_{k \in MP} \tau_k D_k(p^R, p^{MP})$$

where R denotes the set of retail products and MP denotes the set of marketplace products. mc_j and τ_k denote marginal costs and commission payment per unit of demand respectively.[4] Note that demand for any specific product depends on the full vector of prices (p^R, p^{MP}). We take prices of marketplace products as given and assume they do not change when retail prices change.

The first order condition with respect to the price of retail product l is given by

$$D_l + \sum_{j \in R} (p_j - mc_j) \frac{\partial D_j}{\partial p_l} + \sum_{k \in MP} \tau_k \frac{\partial D_k}{\partial p_l} = 0$$

First order conditions for the other retail products look similar. We can write the system of equations in vector form as

$$\overrightarrow{D^R} + \Delta D^R \overrightarrow{p} - \Delta D^R \overrightarrow{mc} + \Delta D^{MP} \overrightarrow{\tau} = 0$$

where $\overrightarrow{D^R}$ is a J^R-dimensional vector of demands for retail products. Vectors of prices and marginal costs are defined similarly. Commissions ($\overrightarrow{\tau}$) are a J^{MP}-dimensional vector. ΔD^R denotes a $(J^R \times J^R)$ matrix of derivatives, where the (a, b) element is equal to $\partial D_a / \partial p_b$. ΔD^{MP} denotes a $(J^{MP} \times J^R)$ matrix of derivatives of marketplace demands with respect to retail prices. Rearranging, we can solve for optimal prices

$$\overrightarrow{p} = \overrightarrow{mc} - (\Delta D^R)^{-1} \left[\overrightarrow{D^R} + \Delta D^{MP} \overrightarrow{\tau} \right]. \tag{3}$$

3 Data and Descriptive Statistics

In this section, we briefly outline the data we use to estimate the model outlined above. We also provide evidence that joint-search of pairs of products is not well predicted by the similarity of the products in characteristic space. Hence, the additional information on substitutability that our

[4] Commissions are set as a percentage of price and hence total commission payments per unit of demand are given by $\tau_k = comm \times p_k$, where $comm$ denotes the commission rate which is the same across all products.

Average # of Searches		2.54
		Percentage
Number of Searches	1	55.62
	2	17.59
	3	8.76
	4	5.21
	5	3.44
	6	2.24
	7	1.61
	8	1.14
	9	0.85
	>=10	3.54
List of Product Characteristics		Price
		Deal Dummy (Price Decrease that is Highlighted on the Webpage)
		Ratings (Customer Reviews): 1-5 Scale
		Number of Customer Reviews
		Brand (27)
		3 Discrete Characteristics (Anonymized)
		3 Continuous Characteristics (Anonymized)

Table 1: **Descriptive Statistics.**

model leverages by incorporating the search data is particularly useful in this setting and constitutes a strength of our modeling approach.

We use data from an online retailer selling home improvement products. We focus on one product category, which is one of the largest sold by the retailer and that contains a total of 579 products during our sample period. We observe the entire history of consumers' browsing and purchase behavior during a 13 week period from 4/20/2016 to 7/16/2016. The final data set contains 470,000 searches and 13,000 purchases (basket additions) by 186,000 distinct users. We treat basket additions as the choice outcome in the demand model and assume that the conversion rate of basket additions to purchases does not vary across products and is unaffected by price. Under these assumptions the conversion term drops out of the first order condition in equation (3) and hence does not affect optimal prices. We focus on basket additions because actual purchases are even less frequent and the retailer did not store purchase information in a way that was easily accessible to us. Basket additions however were tracked as part of the browsing data. Roughly 6.8% of all search sessions end in a purchase and the average search session contains 2.5 products. The distribution of search set size is reported in the top panel of Table 1.

For each product, we observe price, whether the product was "on deal" in a given time period, the average customer review ratings, the number of reviews, brand as well as 6 further product characteristics which we are not able to disclose. The latter 6 constitute physical characteristics of products and are hence invariant over time. The number of reviews and the average rating can

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	# Joint Searches	Joint Search Ratio	Log # Joint Searches	# Joint Searches	Joint Search Ratio	Log # Joint Searches
Mean	9.21	0.43	1.31	9.21	0.43	1.31
S.D.	46.46	1.16	1.18	46.46	1.16	1.18
Same Brand	22.438***	0.806***	0.861***			
	(0.407)	(0.010)	(0.010)			
Price Difference	-2.292***	-0.050***	-0.236***			
	(0.117)	(0.003)	(0.003)			
Same Char. 1	15.026***	1.048***	0.746***			
(Discrete)	(0.945)	(0.023)	(0.022)			
Same Char. 2	3.812***	0.078***	0.164***			
(Discrete)	(0.261)	(0.006)	(0.006)			
Same Char. 3	2.666***	0.181***	-0.003			
(Discrete)	(0.306)	(0.007)	(0.007)			
Diff. Char. 4	-1.460***	-0.116***	-0.109***			
(Continuous)	(0.116)	(0.003)	(0.003)			
Diff. Char. 5	-1.913***	-0.066***	-0.117***			
(Continuous)	(0.114)	(0.003)	(0.003)			
Diff. Char. 6	-2.076***	-0.014***	-0.093***			
(Continuous)	(0.119)	(0.003)	(0.003)			
Similarity Score				7.148***	0.316***	0.231***
				(0.113)	(0.003)	(0.003)
Products	579	579	579	579	579	579
Observations	165,600	165,600	165,600	165,600	165,600	165,600
R-squared	0.038	0.083	0.162	0.024	0.074	0.038

Table 2: **Determinants of Co-search of Product Pairs.** Joint-search ratio is defined as [# Searches (i,j)]/[# Searches (i)×# Searches (j)]. Characteristics 1-3 are discrete variables. Regressors are defined as a dummy equal to one if the characteristic has the same value for both products. Characteristics 4-6 are continuous. Regressors are defined as the absolute difference between characteristics. All continuous variables (price difference and characteristics 4-6) are standardized to facilitate the interpretation of effect magnitudes.

in principle change over time. In reality, the variation in both variables over our sample period is very limited. For easier reference, we list the set of observed product characteristics in Table 1.

3.1 Browsing Data and Product Characteristics

One key advantage of our approach is the fact that we are able to let search behavior, and in particular the co-occurrence of products in a given search spell, inform substitution patterns. Our approach allows us to capture product similarity as revealed by co-occurrence patterns even if the products that are searched together do not share any observed characteristics. To get a sense of the degree to which similarity along unobserved dimensions matters, we explore if similarity in characteristic space can predict whether two products tend to be searched together. To this end, we run a regression at the product-pair level where we regress the number of co-occurrences in the same search session on various measures of similarity in characteristics space and investigate the r-squared.

Specifically, we use a set of 8 characteristics that the firm uses internally to define product similarity. These are price, brand and the 6 undisclosed product characteristics mentioned above. For all discrete variables we define a dummy that is equal to one if the variable takes the same value for both products. For example we define a dummy which is equal to one if the products belong to the same brand. Similar dummy variables are defined for the 3 other discrete characteristics. For all continuous characteristics, we compute the absolute value of the difference between the two products. To facilitate interpretation, we standardize all continuous difference variables so that the coefficient can be interpreted as capturing the impact of a one standard deviation change.

Results from this regression are reported in column (1) of Table 2. All characteristics have a significant impact and the coefficients have the expected sign. Sharing the same characteristic increases the occurrence of co-search and a larger difference in continuous characteristics lowers the numbers of co-searches. Some coefficients are relatively large in magnitude. For instance, belonging to the same brand shifts the occurrence of joint search by 0.5 standard deviations. However, the r-squared is equal to only 0.038 and hence a large part of the variation is not explained by similarity in the 8 characteristics included in the regression.

We probe robustness to different functional forms for the outcome variables in columns (2) and (3). In column (2) we use the ratio of joint searches divided by the product of the number of searches that include each product individually (# Searches (i,j))/[# Searches (i)$\times$# Searches (j)). This metric adjusts for the fact that co-occurrence with any other product is mechanically higher for products that are generally searched more frequently. In column (3) we use a logarithmic transformation of the number of joint searches as the dependent variable. The r-squared is higher in both specifications but still low in absolute value. We also run the same set of regressions using a similarity score (which the firm computed) as a regressor. The similarity score is itself based on the 8 characteristics used in the regressions in columns (1) to (3). Again, the predictive power of the regressions is relatively low. We further probe robustness to removing outliers and to including higher order terms of all covariates and find the results do not change qualitatively.

In summary, we take these regressions as evidence that a large part of variation in joint search patterns cannot be explained by observed characteristics. Hence it constitutes a distinct advantage of our approach to allow search co-occurrence to directly influence substitution patterns.

4 Estimation Routine

We outlined the general structure of the demand model in Section 2 earlier. Below we describe how we specify the conditional utility function. We explore different models, starting with a simple specification of utility and then proceeding to richer model specifications.

4.1 Basic model

The most basic model we estimate contains a full set of product fixed effects and a price coefficient that is common across consumers. Due to the specific context of our data, we also add a dummy for whether the product was "on deal" in a given time period. Deals typically entail other changes such as more prominent placement on the webpage. Our focus here is on estimating elasticities for non-promotional price changes to inform optimal regular price levels for each product. We therefore control for deals in estimation and hold deal status constant when computing price elasticities.

Hence, the utility function is given by (we omit the normalized outside option for simplicity):

$$u_{ij|s} = \delta_j - \alpha_1 p_{jt} + \alpha_2 deal_{jt} + \varepsilon_{ij},$$

where $deal_{jt}$ is equal to one if product j is "on deal" in time period t. Note that this simple version of the model assumes that utility does not depend on the consideration set. We will relax this feature of the model below.

We also add one further parameter that governs the degree of substitution between products versus substitution towards the outside option. In particular, we assume that with probability θ, which is to be estimated, a search spell contains a product that yields higher utility than the outside good but breaks down (and hence no product is purchased). We introduce this parameter due to the fact that many search spells (93.1%) do not end with a purchase. The breakdown parameter allows for the fact that not all outside good choices are due to the consumer not liking any of the sampled products. This parameter helps us match the degree of substitution between products and versus the outside good in a parsimonious way. To see this, consider the case where all outside good choices are driven by a breakdown of search. In this case, a price change for any product will not divert any consumers to the outside option. If instead some outside option choices are intentional, then a price change will lead to substitution to the outside good. Hence, for a given level of demand response to a price change, the breakdown parameter governs the degree of substitution to or from the outside option. Due to the fact that all products are payoff-relevant for the retailer, a flexible approach to capturing the degree of substitution between "inside goods" and towards the outside option is particularly relevant here.

This introduction of the breakdown parameter θ requires a simple adjustment to the way in which we compute choice probabilities:

$$
\begin{aligned}
Pr_{ik|s}(\Omega_{is}) &= (1-\theta)\frac{\exp(\bar{u}_{ik|s})}{1+\sum_{j\in s}\exp(\bar{u}_{ij|s})} \\
Pr_{i0|s}(\Omega_{is}) &= \theta + (1-\theta)\frac{1}{1+\sum_{j\in s}\exp(\bar{u}_{ij|s})}.
\end{aligned}
$$

With regards to estimating the vector of fixed effects, we rely on the fact that (similar to a demand model without consideration sets) the product intercepts play the role of matching the predicted market shares to their empirical counterparts. Therefore the vector δ solves the following system of equations (δ is contained in the parameter vector Ω_{is}):

$$
\sum_i \mathbf{1}(k \in s_i) Pr_{ik|s}(\Omega_{is}) = \frac{1}{N}\sum_i \mathbf{1}_{ik},\ \forall k
$$

where $\mathbf{1}_{ik}$ denotes an indicator function that is equal to one if consumer i purchased product k. We employ a contraction mapping in the spirit of Berry, Levinsohn, and Pakes (1995) to solve for the vector of product intercepts. This estimation routine allows us to recover the large vector of fixed effects (around 1,000 in our setting) at relatively low computational cost.

Although the specification of the conditional utility function in this basic model is terse, the model does not exhibit the IIA property, a key shortcoming of the homogeneous logit model. This is because consumers will substitute within their respective consideration sets following a price change (the model exhibits the IIA property only *conditional on consideration*). By allowing consideration sets to drive substitution patterns, even this simple specification can flexibly capture substitution patterns across products. We present results from this model under the heading "Model 1" in Table 3. Results are discussed together with richer specifications below.

4.2 Extending the model

The more general version of the model takes the form

$$
\begin{aligned}
u_{ij|s} &= \delta_j - \alpha_{1,is} p_{jt} + \alpha_{2,is} deal_{jt} + \beta_{is} X_j + \varepsilon_{ij} \\
\alpha_{1,is} &= \bar{\alpha}_1 + \tilde{\alpha}_1 Z_{is}^{Price} \\
\alpha_{2,is} &= \bar{\alpha}_2 + \tilde{\alpha}_2 Z_{is}^{Price} \\
\theta_{is} &= \bar{\theta} + \tilde{\theta} Z_{is}^{Price} \\
\beta_{is} &= \bar{\beta} + \tilde{\beta} Z_{is}^{ProdChar},
\end{aligned}
$$

where Z_{is}^{Price} and $Z_{is}^{ProdChar}$ denote vectors of consumer/consideration-set characteristics (we use "consumer characteristics" as a shorthand for "consumer/consideration-set characteristics" going forward). $\bar{\alpha}_1$ and $\tilde{\alpha}_1$ denote the average price coefficient and changes in the price coefficient as a function of consumer characteristics. The same notation applies to the deal coefficient $\alpha_{2,is}$, the breakdown parameter θ_{is} and the vector of preference parameters for non-price characteristics β_{is}. We note that un-interacted (time-invariant) product characteristics are co-linear with the vector of product fixed effects and hence $\bar{\beta}$ cannot be separately identified.

We first extend the model by incorporating consumer characteristics Z_{is}^{Price} that interact the price and deal coefficients as well as the breakdown parameter. We intentionally use the same set of characteristics in those three places because the price, deal and breakdown coefficients all relate to price effects. Specifically, we use the following set of consumer characteristics: whether the consumer's search set is characterized by (1) an above median price level and/or (2) an above median share of products on deal. Note that both consumer characteristics are based on observed search behavior, i.e. the consideration set s. In principle, any variable summarizing some aspect of the consideration set can be used as an interaction term. Here, we take the approach of including characteristics that seem a priori likely to affect price sensitivity. For instance, the presence of low price / "on deal" products is likely to correlate with price sensitivity. This specification is presented as "Model 2" in Table 3.

We then extend the model further by allowing interactions between consumer characteristics and other (non-price related) product characteristics. We allow consumers that predominantly searched products with specific characteristics to have different preferences over those same characteristics when choosing which product to purchase. This mirrors our approach to interacting the price coefficient with the price level of products included in the consideration set. For 7 different product characteristics, we compute whether a consumer's consideration set contains products with disproportionately high values along the specific dimension. For ease of interpretation we code these consumer characteristics as binary values. The specific consumer characteristic is equal to one if the consideration set contains products with on average high values of the relevant product characteristic. We then interact this indicator for prevalence of the characteristic in the consideration set with the product characteristic itself; this allows sensitivity of choice (conditional on consideration) with respect to the product characteristic to be different for consumers that predominantly searched products that score highly along that characteristic dimension. This specification is presented as "Model 3" in Table 3.

We note the type of heterogeneity we introduce here is unusual because it is entirely based on the consumer's search behavior rather than any demographic information. This choice is partly due to the fact that we do not have access to any demographic information, a situation that is likely to be common for many online retailers. At the same time, summary measures of search behavior are also likely to be particularly meaningful in terms of explaining preference heterogeneity regarding specific characteristics. Furthermore the nature of the interaction has a close relationship to models of consumer search which generally entail a link between consideration set formation and preferences. Specifically, if a product appears favorable along some characteristic dimension to a consumer and this characteristic is known prior to search, then the product will be more likely to be searched and purchased. Hence preferences for this product will be different depending on the specific observed consideration set. Here we introduce a relationship between consideration sets

Product Characteristic	Consumer/Search Characteristic	*Model 1* Coeff.	*Model 1* S.E.	*Model 2* Coeff.	*Model 2* S.E.	*Model 3* Coeff.	*Model 3* S.E.
Price (α_1)		-0.809	0.032	-1.022	0.042	-0.979	0.046
Deal Dummy (α_2)		0.080	0.020	0.239	0.056	0.255	0.056
Breakdown (θ)		0.592	0.006	0.518	0.015	0.469	0.019
Price	High Price Searches			0.396	0.050	0.362	0.051
	On-Deal Searches			-0.238	0.067	-0.252	0.069
Deal Dummy	High Price Searches			-0.281	0.051	-0.290	0.051
	On-Deal Searches			-0.128	0.063	-0.145	0.064
Breakdown	High Price Searches			0.194	0.020	0.168	0.023
	On-Deal Searches			-0.114	0.032	-0.181	0.038
Customer Rating	High Rating Searches					0.121	0.022
# Reviews	Large # Reviews Searches					0.003	0.022
Characteristic 2	High Char. 2 Searches					-0.037	0.037
Characteristic 3	High Char. 3 Searches					0.349	0.043
Characteristic 4	High Char. 4 Searches					0.062	0.031
Characteristic 5	High Char. 5 Searches					0.029	0.026
Characteristic 6	High Char. 6 Searches					0.609	0.047
Product FEs		Yes		Yes		Yes	
# Observations		185,963		185,963		185,963	
Product-Level Absolute Price Changes (in %) Relative to Previous Specification	Median	n/a		0.081		0.004	
	90th Percentile	n/a		0.150		0.013	
	Max	n/a		0.196		0.019	

Table 3: **Estimation Results.** Characteristic 2 - Characteristic 6 are the same anonymized characteristics used in Table 2. Characteristic 1 is not used in the demand model.

and conditional choices that mimics this aspect of search models.

Due to the nature of our approach, any dependencies between consideration and preferences we allow for will not be rooted in theory. However, we can in principle allow for very flexible dependencies that are less restrictive than the ones implied by specific models of consumer search. With infinite data one could allow preferences to be entirely different conditional preferences in each consideration set. While we are not able to allow for such a level of flexibility here, the implementation of a more parsimonious specification is due to data limitations rather than conceptual constraints.

4.3 Estimation Results

We present results from the simplest specification under the heading "Model 1" in Table 3. This basic model contains three parameters to be estimated: α_1, α_2 and θ as well as a set of product fixed effects. Results for the large set of fixed effects are suppressed. We find that the price and deal coefficients are precisely estimated and have the expected sign. The breakdown probability is equal to 59.2% and hence a large share of all search spells without a purchase (93.1%) are due to search breaking down.

In Model 2 we add interaction terms to the price, deal and breakdown coefficients. As expected, we find that consumers that search more expensive items are less price sensitive, and consumers that predominantly search products which are on deal are more price sensitive. Coefficients on both interaction terms are precisely estimated and large in magnitude. High price and on deal search respectively lead to a shift of 0.396 and -0.238 in the price coefficient (relative to a baseline coefficient of -1.022).[5] As noted above, this suggests that the search-behavior-based characteristics used as conditional preference shifters have substantial explanatory power. We also find that search spells of consumers that search expensive and on deal products are less likely to break down and hence such consumers are more likely to substitute to the outside option following a price increase.

Model 3 further adds heterogeneity in conditional preferences for other characteristics. In particular, we add heterogeneity in preference over the customer review rating and the number of reviews, as well as 5 anonymized horizontal product attributes. The latter set consists of the characteristics that were described in Section 3 and entered the regressions reported in Table 2. We omit one characteristic (Characteristic 1 in Table 2) because it takes on a large set of discrete values that have no ordinal interpretation. Contrary to price related parameters, these interaction terms relate to time-invariant characteristics and hence no baseline effect associated with preferences over those characteristics is included due to co-linearity with the product fixed effects. As expected we find that all coefficients are positive (except for one insignificant, negative coefficient), because consumers that predominantly search products with a particular characteristic are likely to value that product characteristic more. In particular 4 out of 7 coefficients are positive and significant, while the other 3 are statistically insignificant. The coefficients on the price related variables (including the interaction terms) only change marginally compared to Model 2.

We note that the characteristics we choose to include constitute an almost exhaustive set of the product information that was available to us and that is recorded by the retailers. We did not include brand identity and "Characteristic 1" which was used earlier because both variables take on many discrete values. Apart from these two variables, all available characteristics were included in the model. We therefore view Model 3 as using the available information in the data fairly exhaustively. It is also important to keep in mind that any flexibility in preferences will lead to flexibility with regards to substitution patterns *conditional on consideration set choice*. As we showed earlier, unconditional substitution patterns will then be driven by the frequency of co-occurrence of products in consideration sets and the cross-elasticity of demand conditional on consideration. It is the latter part that will be affected by richer preferences specifications. However due to the fact that consideration sets tend to be small (on average 2.54 products out

[5]We note that one of the deal interactions has a counter-intuitive sign. Namely having a lot of deal products in the consideration set makes the consumer less likely to choose a deal product from that set.

of 579), flexibility in conditional choice might not be as important with regards to impacting the inferred elasticity matrix and optimal prices.

With this in mind, we proceed to compare the implied optimal prices based on each of the model specifications. While we discuss optimal prices in more detail in the next section, we now report some simple summary measures of how much optimal prices change as we enrich the model. Specifically, we compute the absolute percentage change in the optimal price for each product and report percentiles of that distribution in the lower panel of Table 3. We find that moving from Model 1 to Model 2 does affect the implied optimal prices significantly. For the median product there is an 8% difference in price, and at the extreme one product's price changes by 20%. Hence allowing for flexibility in price-related terms seems important in terms of optimal prices. However when moving from Model 2 to Model 3, the price changes are much smaller and equal to only 0.4% and 1.9% at the median and maximum respectively. We therefore conclude that there are "decreasing returns" from adding more variables to the utility specification, and hence further flexibility (if data on more characteristics were available) might not alter price recommendations derived from the model very much.

4.4 Solving for Optimal Prices

In this section we give a brief overview of the nature of optimal prices derived from the model and how they relate to current prices of the retailer. In the interest of keeping the identity of the retailer anonymous, we do not report current and optimized prices (or cost-related variables) directly, but frame our discussion mostly in terms of mark-ups at current prices and at the optimal prices. We derive optimal prices as outlined earlier in Section 2.4, using Model 3 parameter estimates. The key ingredient that the demand model provides when solving for prices is the derivative matrix (ΔD^R in equation 3) across all products in the assortment. As discussed earlier, we only solve for prices of retail products and assume that the prices of marketplace products, over which the retailer has no control, remain constant.

Interestingly, we find that the average mark-up (over the wholesale price) under the optimal price vector is equal to 140% percent and much higher than the current average mark-up of only 26%. From preliminary discussions with the company we know that they focused on maximizing revenue rather than profit, and that they did not consider cross-product substitution patterns when analyzing demand for individual products. To get a sense of the relevance of these two aspects, we derive optimal prices under two alternative assumptions. First, we base prices on revenue rather than profit maximization and second, we set the off-diagonal elements of the derivative matrix ΔD^R equal to zero to focus on own-price effects. Switching to revenue maximization lowers the average mark-up from 140% to 51%. When also ignoring cross-terms in the derivative matrix, the average mark-up further decreases to 29% percent. The latter is close to the current mark-up of 26%. The difference in mark-ups from including cross-terms highlights how important cross-product substitution is for optimal price setting.

We are currently in the process of discussing next steps with the retailer. Very likely we will try to validate the model by running selective pricing experiments for individual products and confronting the observed changes in demand across the assortment with the changes predicted by the demand model.

5 Relationship to other Frameworks for Demand Estimation

In this section we relate our model to three existing approaches for modeling consumer demand under limited information. We first contrast our approach with consideration set and search models. Following the taxonomy introduced earlier, we define those two approaches respectively as models where consideration and choice are determined in two stages versus models where both are derived from one unified optimization framework. Furthermore, we compare our framework with models employing heuristics to determine consideration sets. Finally, we draw a comparison with perfect information demand models.

5.1 Consideration Set Models

Consideration set models have been applied in a wide variety of settings such as the demand for personal computers (Goeree (2008)), the automobile market (Barroso and Llobet (2012)), online purchases on eBay (Dinerstein, Einav, Levin, and Sundaresany (2017)), and consumer packaged goods (Bronnenberg and Vanhonacker (1996), Mehta, Rajiv, and Srinivasan (2003), Swait and Erdem (2007)). Our approach has a close relationship to such models because we also distinguish between consideration and then choice conditional on consideration. However a key difference from most papers in this literature is the fact that we do not unpack the process by which consideration sets are formed, whereas typically this process is explicitly modeled.

In principle, the researcher can be quite flexible in terms of modeling the process driving consideration set formation. In practice, a set of three assumptions has frequently been maintained when estimating such models: (1) consideration probabilities are independent across products conditional on characteristics, (2) a function of product characteristics and (3) variables driving consideration are excluded from preferences (which determine choice conditional on search). Many prominent papers in this literature make all three assumptions. Here we discuss in turn the consequences of these assumptions and contrast them with our approach.

The first and second assumptions restrict the way in which consideration sets are formed and hence the flexibility of substitution patterns that the model generates. With a characteristics-based consideration process, the only way in which the model can rationalize the co-occurrence of products in consideration sets is through observables. For example, if consideration depends positively on advertising and two products are advertised heavily, then the model will predict a high likelihood of joint-consideration for this product pair. Hence the elasticity matrix generated by the model will be driven by observable characteristics that enter consideration (as well as choice), albeit in a different non-linear way than in a perfect information demand model. However, if two products are frequently considered together but are not similar along any observed dimension, the model will not be able to rationalize the joint-consideration frequency for those products and will predict their cross-elasticity to be low.

Our approach instead conditions on observed consideration sets and hence product groupings in consideration sets are allowed to directly drive substitution patterns. Our approach does not require us to trace joint consideration back to the underlying drivers. In a context where product characteristics are hard to define by the researcher and joint consideration might be driven by unobserved factors, such flexibility is desirable. Additionally, in online markets where retailers stock a larger set of products, but have a relatively small set of observable characteristics at hand to describe individual products, a characteristics-driven search model will be unlikely to capture rich consideration (and hence substitution) patterns. This is nicely demonstrated by our earlier Table 2, where we showed that eight product characteristics observed by the retailer explain a relatively low amount of variation in joint consideration patterns.

The third assumption regarding the separation of consideration and choice can be captured in our model by removing the s subscript from all parameters in the utility function that allows preference parameters to vary as a function of the consideration set. The reason for this assumption is that in most studies no data on consideration is available and separate identification of the influence of one variable on both consideration and choice would be based on functional form. Therefore, many papers resort to separating the variables driving consideration and choice. For instance Goeree (2008) assumes advertising only affects consideration, whereas price and other product characteristics drive choice but do not affect consideration. When data on consideration is available there is no need to maintain this assumption.

In general, some link between consideration and choice seems likely. For instance, if advertising (which is likely to affect consideration) is targeted towards consumers with specific preferences, utility in the choice stage will differ across consumers with different consideration sets. Therefore relaxing the assumption of separation between consideration and choice, which our approach achieves, seems desirable in many settings.

5.2 Consumer Search Models

Relative to the consideration set models just discussed, models of consumer search derive consideration / search (we will use the two terms interchangeably in this subsection) and choice from one unified framework of utility maximization. Structural models of search differ in the assumptions made about whether search is simultaneous or sequential, and also the information structure, i.e. what information consumers have before and after search. In more detail, under simultaneous search the consumer pre-commits to search a particular set of products, whereas under sequential search she searches products one at a time. In both cases consumers pay a cost per product searched, and along with utility this search cost determines the set of products searched (and for sequential search, also the order). In terms of information structure, the typical approach assumes that consumers know the utility they derive from a product up to a match value error term that is independent across consumers and products (e.g. Kim, Albuquerque, and Bronnenberg (2010)). In other words consumers are assumed to know at least some characteristics of each product prior to search. Some papers estimate models with different informational assumptions and model search over price (Honka (2014), Honka and Chintagunta (2016)) or assume that no part of utility is known prior to search. In either case, it is assumed that consumers form correct expectations based on the empirical distribution of the unknown part of utility.

Simultaneous search models have a relatively close mapping to our setting. Such models feature two decisions, a first decision pertaining to which products (and how many) to search and a choice conditional on search. This process is not too dis-similar to the two stage models described above, but for the fact that the first stage takes a very specific form. More specifically, search models prescribe a specific way in which consideration and choice relate to each other. When uncertainty is only over the match value for instance, all observed characteristics influence both search and choice. The way that consideration and choice are linked is of course a specific one and it is a link that is driven by the underlying assumption of the search process and information structure. Models of sequential search similarly impose a specific (but different) link between preferences and consideration set formation, although they involve the consumer generally taking more than one decision (a stop-or-continue-searching decision is taken after each search).[6] Our model instead allows for flexibility in a post-hoc fashion that is not derived from theory.

There are several reasons why search models of the kind just described are less suitable for settings such as ours where choice sets contain a large number of products. First, such models tend to be computationally burdensome to estimate and structural search models have therefore primarily been applied to markets with a relatively small number of choices. Second, the typical models assume a high degree of sophistication with regards to consumers' information processing and cognitive abilities. In a model with uncertainty over match values only, the standard assumption is that consumers know the characteristics of all products in the market and form their optimal search strategies based on them. Our approach is agnostic about the nature of the informational environment and is consistent with cognitively "lighter" approaches such as the choice heuristics described in the next section.

Finally, we note that for search models that are based on a utility function in characteristics space, it is the case that substitution patterns are entirely driven by product characteristics. We described the same criticism in the previous section in the context of consideration set models and the same logic also applies here. When we trace consideration set formation back to a characteristics based utility function, it is characteristics that determine elasticities via their influence on search and choice.

[6]In terms of the functional form of preferences conditional on search, sequential models do not have a close mapping to our utility specification. The structure of utility conditional on the observed consideration set takes a more complicated form due to the fact that the consumer must have received relatively unfavorable realizations of unknown variable whenever he continued searching. For instance, if uncertainty is over a match value term that is iid normal, then the distribution of this error in any observed consideration set follows a truncated distribution (because a high realization early in the search would have led to the search process being terminated earlier).

5.3 Choice Heuristics

Another approach to modeling how consumers evaluate products is heuristic-based models. Such models describe rule-based strategies used by consumers to simplify decisions and reduce cognitive effort. They may be particularly relevant in online settings, where consumers must choose from a large number of products but each purchase is relatively low involvement.

One strand of this literature assumes that consumers make their choice entirely using rules. The most well-known is Tversky (1972)'s elimination-by aspects (EBA), in which products are viewed as a collection of attributes, and the consumer (randomly) eliminates products that do not contain a given attribute, until at most one product is left. Jedidi and Kohli (2005) show that when aspects have utility which follows an extreme value distribution, EBA can be represented as a generalized logit model. Consequently our modeling approach is consistent with EBA, and one could interpret the consideration set observed in our dataset as a snapshot of the consumer's elimination process.

Another strand of this literature assumes that consumers use a two-stage approach which is closer to our model. In the first stage consumers use simple non-compensatory rules to narrow down the set of options and thereby form their consideration set. Typical examples of rules studied in the literature are EBA, lexicographic (Fishburn (1974)), and disjunctive/conjunctive (Dawes (1964)).[7] In the second stage consumers then use a compensatory mechanism to make a choice from their consideration set. For example Fader and McAlister (1990) estimate demand for ground coffee, and assume that at the first stage consumers use EBA with the attributes being recency of purchase and whether a brand is on promotion; Gilbride and Allenby (2004) use experimental data on camera purchase intentions, and allow for conjunctive and disjunctive decision rules at the first stage; Bronnenberg and Vanhonacker (1996) consider demand for detergent, and allow factors such as shelf space and recency of purchase to affect first-stage consideration. Notice that our approach is fully consistent with these and other two-stage heuristic choice models. However similar to the other approaches mentioned earlier, rule-based approaches impose certain dependencies between consideration sets and the choice stage, and moreover require the researcher to take an a priori stance on what factors influence consideration. An advantage of our approach is that we can be agnostic about these.

5.4 Perfect Information Demand Models

Our approach shares many features with perfect information demand models. In particular, the only differences in our approach are that we model demand conditional on consideration set formation, and we allow preferences parameter to depend on the consideration set. Choice conditional on consideration has the same general structure as most discrete choice demand models in the tradition of Berry, Levinsohn, and Pakes (1995). In this section we draw a parallel between the way in

[7]All of these approaches view products as collections of attributes. A disjunctive rule includes a product in the consideration set if it is acceptable on at least one relevant attribute, whereas a conjunctive rule only includes a product in the consideration set if it is acceptable on all relevant attributes. See Aribarg, Otter, Zantedeschi, Allenby, Bentley, Curry, Dotson, Henderson, Honka, Kohli, Jedidi, Seiler, and Wang (2017) for a recent summary of the literature.

which random coefficients inform substitution patterns in perfect information demand models and how consideration set patterns (together with random coefficients in the utility function) inform substitution patterns in our setting.

To illustrate the point, consider again the fictional car market outlined in the introduction (and visualized in Figure 1), which consists of 10 car models that can be classified into 5 high horsepower and 5 low horsepower cars. It seems likely that consumers vary in the extent to which they value horsepower (relative to price) and therefore some set of consumers will tend to consider primarily the subset of cars with high horsepower and some consumers will tend to consider the low horsepower cars. In a random coefficient demand model one would capture such behavior by including a random coefficient on horsepower in the consumer's utility function. In our setting instead, we capture such behavior partly through the search process. In particular, we would expect some subset of consumers to only search cars from the high or the low horsepower subsets with fewer searches cutting across those two groups. In the same way that the random coefficient on horsepower will then lead to larger cross-elasticities within each of the two groups, our search model will also predict that substitution happens mostly to other cars within the same horsepower group.

There are two advantages to our approach. First, consumers' consideration sets are observed in the data and they directly inform which type of products are close substitutes to one another. There is hence no issue of how these sets are identified. Instead, random coefficients models typically rely on either choice set variation in cross-sectional data or consumer panel data. Neither of these data requirements needs to be fulfilled in order to estimate our model. Second, there is no need to define the set of characteristics for which heterogeneous tastes are modeled. It is conceivable that characteristics that matter strongly to some consumers (and hence drive substitution patterns) are hard to measure or simply not included in the data available to the researcher. Furthermore, including a large set of characteristics in the utility function and estimating heterogeneous preferences over these characteristics in a flexible way can be a computationally challenging. Our model instead does not require us to specify the set of characteristics that drive the observed consideration sets and we can hence remain agnostic about the relevant characteristics driving set formation and simply let the data "tell us" which products tend to be grouped together in consideration sets.

However, characteristics that drive consideration (which we do not model) are likely to also influence choice conditional on consideration (which we do model explicitly). For instance, using the same example above, heterogeneity in preferences over horse power are likely to influence both consideration and choice i.e. consumers that like high horsepower cars are more likely consider them and are more likely to choose them from a given consideration set. Hence choice and consideration set formation are unlikely to be independent, and therefore we do allow for such dependencies between the two decision stages in our demand model.

Finally, we note that our estimation routine is almost identical to the estimation of a perfect information demand model with an identical utility specification. We simply need to store a (consumer-$*$-product)-dimensional matrix that contains information on consideration sets. We then use this matrix to “remove” non-searched option from the choice set when computing purchase probabilities conditional on consideration. By narrowing down choice sets to typically a small set of options our model in fact becomes computationally easier to implement than the equivalent perfect information model. Furthermore, consideration set variation across consumers provides useful information to identify preference parameters whereas assortment changes (which a perfect information model would have to rely on) are minimal in our setting.

6 Conclusion

We proposed a demand model based on search and purchase data that is computationally light and allows for great flexibility in substitution patterns across products. Substitution patterns are directly informed by consumers’ consideration sets and do not have to be rooted in similarity in characteristics space. Furthermore, we are able to sidestep assumptions regarding the structure of the search process and consumers’ information sets which are required in many structural search models. Instead, we treat consideration set probabilities as objects to be estimated and allow conditional preferences to depend on characteristics of the consideration set.

Our model is particularly useful in settings with a large number of products, sparse purchases and limited information on product characteristics. This setting is typical in online retail and we apply it to such a setting. Our data contains search and purchase information for a large category of an online retailer which contains almost 600 products. We solve for profit-maximizing prices based on our demand estimates and find optimal prices to be significantly higher than current prices. We also find that taking cross-product substitution into account is crucial for optimal price setting and mark-ups are substantially lower when ignoring cross-product price effects.

References

Abaluck, J., and A. Adams (2016): "Discrete Choice Models with Consideration Sets: Identification from Asymmetric Cross-Derivatives," .

Andrews, R. L., and T. C. Srinivasan (1995): "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data," *Journal of Marketing Research*, 32(1), 30–41.

Aribarg, A., T. Otter, D. Zantedeschi, G. M. Allenby, T. Bentley, D. J. Curry, M. Dotson, T. Henderson, E. Honka, R. Kohli, K. Jedidi, S. Seiler, and X. S. Wang (2017): "Advancing Non-compensatory Choice Models in Marketing," *Customer Needs and Solutions*.

Barroso, A., and G. Llobet (2012): "Advertising and Consumer Awareness of New, Differentiated Products," *Journal of Marketing Research*, 49(6), 773–792.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Bronnenberg, B. J., and W. R. Vanhonacker (1996): "Limited Choice Sets, Local Price Response and Implied Measures of Price Competition," *Journal of Marketing Research*, 33(2), 163–173.

Chen, Y., and S. Yao (2016): "Sequential Search with Refinement: Model and Application with Click-stream Data," *Management Science*, p. forthcoming.

Chiong, K. X., and M. Shum (forthcoming): "Random Projection Estimation of Discrete-Choice Models with Large Choice Sets," *Management Science*.

Dawes, R. M. (1964): "Social selection based on multidimensional criteria," *The Journal of Abnormal and Social Psychology*, 68(1), 104–109.

De Los Santos, B. I., A. Hortacsu, and M. Wildenbeest (2012): "Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102(6), 2955–2980.

Dinerstein, M., L. Einav, J. Levin, and N. Sundaresany (2017): "Consumer Price Search and Platform Design in Internet Commerce," NBER Working Paper No. 20415.

Draganska, M., and D. 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.

Fader, P. S., and L. McAlister (1990): "An elimination by aspects model of consumer response to promotion calibrated on UPC scanner data," *Journal of Marketing Research*, 27(3), 322–332.

Fishburn, P. C. (1974): "Lexicographic Orders, Utilities and Decision Rules: A Survey," *Management Science*, 20(11), 1442–1471.

Gaynor, M., C. Propper, and S. Seiler (2016): "Free to Choose? Reform, Choice and Consideration Sets in the English National Health Service," *American Economic Review*, 106(11), 3521–3557.

Gilbride, T. J., and G. M. Allenby (2004): "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23(3), 391–406.

Goeree, M. S. (2008): "Limited Information and Advertising in the US Personal Computer Industry," *Econometrica*, 76(5), 1017–1074.

Haviv, A. (2016): "Does Purchase Without Search Explain Counter-Cyclic Pricing?," unpublished manuscript.

Honka, E. (2014): "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry," *RAND Journal of Economics*, 45(4), 847–884.

Honka, E., and P. Chintagunta (2016): "Simultaneous or Sequential? Search Strategies in the US Auto Insurance Industry," forthcoming.

Jedidi, K., and R. Kohli (2005): "Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers," *Journal of Marketing Research*, 42(4), 483–494.

Kawaguchi, K., K. Uetake, and Y. Watanabe (2016): "Identifying Consumer Attention: A Product-Availability Approach," .

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010): "Online Demand under Limited Consumer Search," *Marketing Science*, 29(6), 1001–1023.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2011): "Mapping Online Consumer Search," *Journal of Marketing Research*, 48(1), 13–27.

Koulayev, S. (2014): "Search for Differentiated Products: Identification and Estimation," *RAND Journal of Economics*, 45(3), 553–575.

Lee, T. Y., and E. T. Bradlow (2011): "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48(5), 881–894.

Mehta, N., S. Rajiv, and K. Srinivasan (2003): "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation," *Marketing Science*, 22(1), 58–84.

Netzer, O., R. Feldman, J. Goldenberg, and M. Fresko (2012): "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31(3), 521–543.

Pancras, J. (2010): "A Framework to Determine the Value of Consumer Consideration Set Information for Firm Pricing Strategies," *Computational Economics*, 35(3), 269–300.

Pires, T. (2016): "Costly Search and Consideration Sets in Storable Goods Markets," *Quantitative Marketing and Economics*, 14(3), 157–193.

Ringel, D. M., and B. Skiera (2016): "Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data," *Marketing Science*, 35(3), 511–534.

Roberts, J. H., and J. M. Lattin (1991): "Development and Testing of a Model of Consideration Set Composition," *Journal of Marketing Research*, 28(4), 429–440.

Seiler, S. (2013): "The Impact of Search Costs on Consumer Behavior: A Dynamic Approach," *Quantitative Marketing and Economics*, 11(2), 155–203.

Smith, A. N., P. E. Rossi, and G. M. Allenby (2016): "Inference for Product Competition and Separable Demand," *working paper.*

Swait, J., and T. Erdem (2007): "Brand Effects on Choice and Choice Set Formation Under Uncertainty," *Marketing Science*, 26(5), 679–697.

Tversky, A. (1972): "Choice by elimination," *Journal of Mathematical Psychology*, 9(4), 341 – 367.