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   **Andrew Rhodes**   **Stephan Seiler**

Harvard Business School   Toulouse School   Stanford University
of Economics

This draft: March 20, 2019

In many online markets, traditional methods of demand estimation are difficult to implement because assortments are very large and individual products are sold infrequently. At the same time, data on consumer search (i.e., browsing) behavior are often available and are much more abundant than purchase data. We propose a demand model that caters to this type of setting. Our approach is computationally light and allows for flexible cross-price elasticities that are informed by search patterns. We apply the model to a data set containing search and purchase information from a retailer stocking almost 600 products, recover the elasticity matrix, and solve for optimal prices for the entire assortment.

Keywords: Demand Estimation, Large-Scale Estimation, Consideration Sets, Consumer Search

1 Introduction

Demand estimation is a cornerstone of quantitative marketing and empirical IO, and the basis for optimal price setting and many other applications. However, online markets pose several challenges for demand estimation. One challenge is that the number of products is usually very large, especially relative to the number of observed product characteristics. This feature of online markets reduces the flexibility of substitution patterns that can be estimated from the data. Moreover, it raises the computational burden of common estimation methods, such as those with random coefficients. Another challenge is that purchases are typically sparse, making precise estimation of the parameters that determine substitution patterns difficult. A final challenge is that in online markets, the researcher may not observe many important drivers of demand, such as product rankings and recommendations.

This paper proposes a new demand model that is simple to estimate and allows for flexible substitution patterns even in markets with large numbers of products and sparse purchases. Our approach leverages data on consumer search (i.e., browsing) behavior. These data directly reveal which products a consumer had in her consideration set, and so provide an additional source of information on substitution patterns. Moreover, search data are typically much more abundant than purchase data. For example, in our empirical application, the average product is searched 36 times more frequently than it is purchased.

Our approach has three main advantages. The first advantage is computational. We apply our demand model to a setting with almost 600 products, and find it is roughly 75 times faster to estimate than a full-information model that allows for random coefficients on product characteristics.[1] A second advantage is fit. Specifically, we find our demand model fits the data better than the full-information model. In our approach, products that are frequently included together in consideration sets are allowed to have higher cross-elasticities (everything else equal). Intuitively, in a setting with many products, "revealed similarity" through co-occurrence in consideration sets may better capture substitution patterns than can a limited number of product characteristics. Another advantage of our approach is that it imposes less structure than most other models. For example, in contrast to structural models of consumer search, we do not need to make assumptions about information sets, search protocol, and so forth.

In more detail, we jointly model consideration-set formation and product choice in the following way. We begin by showing that any model of consideration and choice can be decomposed into (i) a probability distribution over consideration sets, and (ii) choice probabilities conditional on each consideration set. We then make two key modeling choices in order to implement this decomposition empirically. First, we do not "unpack" the consideration process and model how consideration sets are formed. Instead, we treat consideration-set probabilities as objects to be estimated, and we model the influence of price on consideration in a reduced-form way. Second, we allow conditional preferences to vary across consideration sets in a way that is informed by models of consumer

---

[1]We did not attempt to estimate a structural search model using consumer-level search and purchase data, because it would be too computationally burdensome given the number of products.
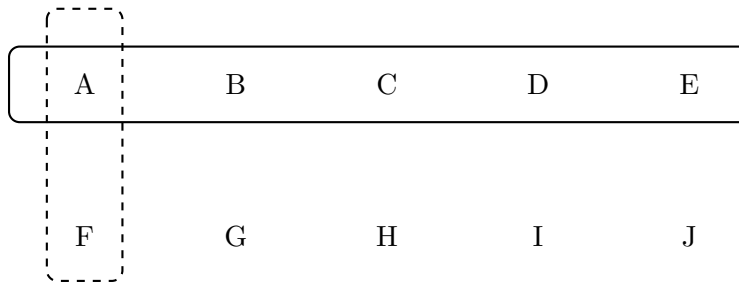
Figure 1: **Illustrative Search Patterns.**

search and consideration-set formation. We then estimate consideration-set probabilities and the parameters of the conditional distributions of preferences. Our modeling framework nests several common approaches in the literature on consumer search and consideration-set formation, and can be seen as an approximation to others. Our approach allows us to conduct price-related counterfactuals such as the estimation of elasticities and optimal prices.[2]

The main idea behind our approach is that search data are informative about substitution patterns. Figure 1 illustrates this via a simple example with 10 products, denoted by $A, \ldots, J$. Assume 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, in the special case in which price affects conditional choice but not consideration,[3] 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. Patterns such as the ones illustrated in Figure 1 could be driven by similarity in product characteristics. For instance, the top and bottom rows could denote respectively high and low quality, and each column could represent a different color. Alternatively, the rows and columns could represent similarity with regards to a characteristic such as product design that is typically unobserved to the researcher or proximity on the webpage. A major advantage of our approach is that the groupings of substitute products are directly observed in the data and do not need to be inferred indirectly from consumers' purchases. Defining the relevant set of characteristics that might drive the substitutability of products is unnecessary. This feature of our modeling approach is particularly valuable in online markets where the number of products is large relative to the number of observable characteristics, and where other drivers of substitution patterns, such as similarity in design or proximity on a webpage, are either hard to quantify or are simply not observed by the researcher.

We illustrate the power of our approach using search and purchase data from an online retailer

---

[2]As we discuss later, we cannot do certain other counterfactuals. For example, because we do not assume a particular model of search, we cannot analyze changes in the search process such as a decrease in search costs.

[3]In our full model, price affects both conditional choice and consideration, and hence substitution patterns are more complex than those outlined here.

selling home-improvement products. We focus on one product category that contains almost 600 items. Our data cover a three-month period and consist of roughly 455,000 product searches and 13,000 purchases by 186,000 users. Apart from search and purchase data, we also obtained data on prices and other (time-invariant) product characteristics. Using our estimates, we solve for the entire cross-price elasticity matrix and then compute optimal prices. We find that optimal prices are, on average, 23% lower than current prices and revenue would increase by 7.1% when switching to the optimal price vector.[4]

We also show our approach outperforms a full-information, characteristics-based random-coefficient demand model in terms of (out-of-sample) model fit. Specifically, the average likelihood for a leave-one-week-out holdout sample is higher for our proposed model than for a full-information model with random coefficients on all characteristics that the firm collects and uses to establish product similarity (nine characteristics). In fact, adding random coefficients does not lead to an improvement in fit relative to a simple logit model without random coefficients. However, our approach based on search data does yield an improvement in fit. We also show that similarity in observed characteristics does not predict very well which products are searched together. Search data therefore capture relevant aspects of substitutability that are missed by the product characteristics recorded by the company. In addition to giving superior fit, our preferred specification is also 75 times faster to estimate than the full-information random-coefficient demand model.

Related literature. Our paper relates to several distinct streams of literature. First, it relates to a literature that employs descriptive methods 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, whereas Lee and Bradlow (2011) use customer reviews and the products they mention. The two papers in this realm that are closest to our approach are Ringel and Skiera (2016) and Kim, Albuquerque, and Bronnenberg (2011), which use online search data to analyze competitive market structure and product substitution. However, neither of these papers estimates an elasticity matrix, due to the absence of information on prices and purchases. Instead, these papers provide a visualization of closeness in product space (a "perceptual map") with an implicit understanding (but no formal derivation) that this visualization informs substitution patterns and hence demand elasticities.[5] Our approach leverages co-occurrence data similarly to these descriptive papers, but it embeds this information into a model of demand and uses it to derive an elasticity matrix.[6]

Our paper is also related to models of consideration-set formation and consumer search. 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

---

[4]As we discuss later, the firm's objective function was to maximize revenue rather than profit.

[5]Kim, Albuquerque, and Bronnenberg (2011) clearly articulate the implied relationship of the perceptual map to substitution patterns (see p.14): "the map can be used to shed some light on substitution patterns. Local subsets of products on the map can be interpreted as stereotypical products or consideration sets that are searched together and, presumably, compete more intensely."

[6]A related application is by Li, Netessine, and Koulayev (2018), who use search data to compute instrumental variables that are used to estimate the relationship between different firms' pricing decisions.

3

underpins both search and purchase decisions. The first stage in a consideration model is typically either characterized as a more passive stage in which consumers become aware of products due to external factors such as advertising or product displays (Bronnenberg and Vanhonacker (1996), Mehta, Rajiv, and Srinivasan (2003), Pancras (2010)), or can be understood as a reduced-form approximation of a structural search model. To organize the literature and the relation of different papers to our approach, we categorize papers into those that conceptualize choice into two separate stages versus those that adopt a unified optimization framework to jointly describe search and choice.

The literature on consideration sets typically assumes some variables affect consideration-set formation but not choice.[7] For instance, Goeree (2008) assumes 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 product displays and feature advertising 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)). For example, high-quality products being considered together can be captured by allowing for quality (possibly interacted with demographics) to affect the consideration probability. Therefore, observed characteristics are what determine how often products are considered together and hence whether they are close substitutes. Despite a different functional form, substitution patterns ultimately trace back to observable characteristics as in the case of perfect-information demand models.

Models of consumer search (e.g., Kim, Albuquerque, and Bronnenberg (2010), De Los Santos, Hortacsu, and Wildenbeest (2012), Honka (2014), Chen and Yao (2016), and Ursu (2018))[8] instead present a unified framework of consumers' utility maximization that rationalizes observed search and purchase patterns. Typically, these models are computationally burdensome and are therefore estimated for markets with a relatively small number of products. Furthermore, in search models, both search and choice are driven by the specified utility function. Typically utility is defined in characteristic space and hence, similar to consideration set models, cross-elasticities are determined by the observable characteristics that enter utility.

We view our approach as combining the strengths of the descriptive approaches outlined above and structural models of search. Similar to the descriptive papers, we let information on co-occurrence in search inform substitution patterns directly without the need to rationalize co-occurrence through similarity in characteristics. By embedding the search information into a model of consideration and choice, we are able to combine it with information on purchases and price variation, which allows us to estimate the elasticity matrix. At the same time, relative to the above

---

[7]Several recent papers analyze identification in consideration-set models without exclusion restrictions. 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. Crawford, Griffith, and Iaria (2018) show how to estimate preferences when unobserved consideration sets need to be integrated out.

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

papers, our approach allows us to remain agnostic about the process that drives search or consideration. Furthermore, we are able to set up estimation such that the computational burden is much lower than in a typical structural search model, thus allowing us to estimate the model for a setting with several hundred products.

Finally, our paper is related to an emerging literature on flexible large-scale demand estimation that is based on purchase data and does not make use of additional data sources such as information on consumer search behavior. 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. Ruiz, Athey, and Blei (2017) estimate a sequential probabilistic model of basket demand.

The remainder of the paper is structured as follows. Section 2 outlines our approach to estimating demand and discusses how it relates to alternative approaches. Section 3 describes the data and provides descriptive statistics. Section 4 applies the model to data from an online retailer. Section 5 reports elasticity estimates and optimal prices. Section 6 concludes.

2   Model Framework

We first provide a general framework that nests any model of consumer search or consideration-set formation. This framework decomposes the choice process into two objects: (i) a probability distribution over consideration sets, and (ii) choice probabilities conditional on a consumer having selected into a particular consideration set. We then introduce our estimation strategy, which seeks to directly estimate these two objects from the data. Finally, we compare our approach with others from the literature.

2.1   A General Model of Consideration and Choice

Consider a setting with J differentiated products, whose prices are given by the vector $\mathbf{p}$ and whose characteristics are given by the matrix $\mathbf{X}$. An outside option exists that we label as good 0. A consumer's preferences and search-related parameters (e.g., search costs) are summarized by a vector θ whose distribution is $F_\theta(\theta)$.[9]

A consideration set s is a subset of the $J + 1$ products from which a consumer makes her final choice. We can write down the following discrete probability distribution (conditional on characteristics) over consideration sets:

$$\Pr(s|\mathbf{p}, \mathbf{X}) = \int \mathbf{1}(s|\mathbf{p}, \mathbf{X}, \theta) \mathrm{d}F_\theta(\theta), \tag{1}$$

where $\mathbf{1}(s|\mathbf{p}, \mathbf{X}, \theta)$ is an indicator function that equals 1 if the consumer picks consideration set s. This function will differ depending on the assumptions made regarding the process of consideration-

---

[9]The type vector θ fully describes the consumer; that is, it includes preferences over product characteristics, realizations of product-level taste shocks, and so forth. Hence, choices conditional on θ are deterministic.

set formation and/or search. For the moment, we simply assume $\mathbf{1}(s|\mathbf{p}, \mathbf{X}, \theta)$ represents *some function* that maps types θ and product characteristics $(\mathbf{p}, \mathbf{X})$ to consideration sets.

We can also derive choice probabilities conditional on selecting a particular consideration set s. To do so, let u_0 denote the utility associated with the outside option, let u_j be the utility (net of price) from buying product $j \in s$, and let $F_{\theta|s}(\theta|s, \mathbf{p}, \mathbf{X})$ be the distribution of θ conditional on choosing set s.[10] The conditional probability that a consumer with consideration set s buys product $j \in s$ can then be written as

$$\Pr(j|s, \mathbf{p}, \mathbf{X}) = \int \mathbf{1}(u_j(\mathbf{p}, \mathbf{X}, \theta) \geq u_k(\mathbf{p}, \mathbf{X}, \theta) \ \forall k \in s) \mathrm{d}F_{\theta|s}(\theta|s, \mathbf{p}, \mathbf{X}), \tag{2}$$

where $\mathbf{1}(.)$ is an indicator function that takes the value of 1 when good j offers the highest utility. We also write $\Pr(j|s, \mathbf{p}, \mathbf{X}) = 0$ for any product $j \notin s$.

Demand for product j is then obtained by summing the conditional demands in (2) over all consideration sets weighted by the probability distribution (1):

$$D_j(\mathbf{p}, \mathbf{X}) = \sum_{s \in \mathcal{S}} \Pr(s|\mathbf{p}, \mathbf{X}) \times \Pr(j|s, \mathbf{p}, \mathbf{X}), \tag{3}$$

where $\mathcal{S}$ denotes the set of all consideration sets. Using this expression, we can write the elasticity of demand for product j with respect to the price p_k of product k as

$$\eta_{j,k} = \underbrace{\frac{p_k \times \sum_{s \in \mathcal{S}} \left(\Pr(s|\mathbf{p}, \mathbf{X}) \times \frac{\partial \Pr(j|s, \mathbf{p}, \mathbf{X})}{\partial p_k} \right)}{D_j(\mathbf{p}, \mathbf{X})}}_{\text{Within-set substitution}} + \underbrace{\frac{p_k \times \sum_{s \in \mathcal{S}} \left(\frac{\partial \Pr(s|\mathbf{p}, \mathbf{X})}{\partial p_k} \times \Pr(j|s, \mathbf{p}, \mathbf{X}) \right)}{D_j(\mathbf{p}, \mathbf{X})}}_{\text{Between-set substitution}}. \tag{4}$$

The first term describes substitution patterns among consumers who do not change their consideration set in response to a change in p_k. Substitution through this channel can only arise for consumers who jointly consider products j and k. The second term describes substitution patterns caused by consumers responding to the change in p_k by changing the set of products they consider.

The general demand expression in (3) nests several common approaches to demand estimation. First, perfect-information demand models correspond to the special case in which each consumer selects into the consideration set that contains every product. Second, as we discuss below, consideration-set models often assume price does not affect consideration, and that preferences are independent of which consideration set was chosen. This set of assumptions corresponds to the special case in which $\Pr(s|\mathbf{p}, \mathbf{X}) = \Pr(s|\mathbf{X})$ and $F_{\theta|s}(\theta|s, \mathbf{p}, \mathbf{X}) = F_\theta(\theta)$, and hence no between-set substitution occurs. Third, because our general framework permits $F_{\theta|s}$ to vary arbitrarily across consideration sets, it allows for any link between consideration and choice. As such, it is consistent with any model of consumer search as well as other ways of forming consideration sets, such as choice heuristics. We return to the comparison of our model to alternative approaches in section

---

[10]The distribution is conditioned on $(\mathbf{p}, \mathbf{X})$ because $(\mathbf{p}, \mathbf{X})$ may affect which consumers select into a given set. For example, if the price of a product in a given set decreases, more price-sensitive consumers might select that set.

2.3, and argue our empirical implementation of (3) either nests or approximates these different approaches.

2.2 Empirical implementation

We now describe in detail how we propose to estimate this general model. Instead of trying to recover the underlying distribution of consumer types θ, our approach directly estimates consideration-set probabilities $\Pr(s|\mathbf{p}, \mathbf{X})$ and the conditional distribution of preferences $F_{\theta|s}(\theta|s, \mathbf{p}, \mathbf{X})$. We parameterize both objects in a way that is informed by models of consideration-set formation and search. This approach allows us to estimate elasticities and to conduct counterfactuals that alter prices (e.g., evaluating profits at a counterfactual price vector) without making assumptions on the process by which consumers search or form consideration sets. The main disadvantage of our approach is that we cannot implement counterfactuals that require knowledge of the search or consideration process and the unconditional distribution of preferences, such as changes in search technology or the evaluation of consumer welfare.

2.2.1 Consideration-Set Formation

We model the probability of consideration set s occurring in period t at a given price vector as

$$\Pr(s|\mathbf{p}, \mathbf{X}) = \frac{\exp(\chi_s + \gamma \overline{price}_{st})}{\sum_{s' \in \mathcal{S}} \exp(\chi_{s'} + \gamma \overline{price}_{s't})}, \tag{5}$$

where χ_s denotes a set-specific fixed effect, $\overline{price}_{st}$ is the average price of the consideration set in that time period, and γ is a parameter to be estimated. One can loosely interpret $\chi_s + \gamma \overline{price}_{st}$ (plus an iid extreme-value shock) as the "utility" of choosing consideration set s. This formulation allows us to sidestep the task of modeling the process by which specific products are combined in consumers' consideration sets. Hence, one should think of this set-level choice model as a reduced-form model for an underlying product-level search process that (i) captures joint-search patterns in the data via consideration-set-specific fixed effects χ_s, and (ii) allows prices to influence the likelihood of different sets occurring. In the special case in which $\gamma = 0$ such that price does not affect consideration, the predicted probability of a given consideration set occurring is just its frequency in the data.

We do not explicitly model how consideration-set probabilities depend on characteristics $\mathbf{X}$, but take the estimated probabilities to be valid at the current realization of characteristics in the data. We thus have in mind a setting where non-price factors that drive consideration, such as product characteristics or webpage layout, do not vary during the sample period. (Therefore, characteristics that drive joint-search patterns are co-linear with the set fixed effects.) We assume the set of possible consideration sets S is given by those observed in the data, and hence the summation in the denominator of equation (5) is over those sets.[11]

---

[11]Thus, a potential limitation of our approach is that it places zero weight on consideration sets that do not appear in the data. However, we show later that in our application, most pairs of products are considered together at least

7

2.2.2 Conditional Choice

We assume a consumer i with a given consideration set obtains the following utilities:

$$
\begin{aligned}
u_{ij} &= \bar{u}_{ij} + \varepsilon_{ij} = \delta_j + \alpha_i price_{jt} + X_j' \beta_i + \varepsilon_{ij} \qquad (6) \\
u_{i0} &= \varepsilon_{i0},
\end{aligned}
$$

which comprise of a deterministic component $\bar{u}_{ij}$ (normalized to zero for the outside option) and an iid extreme-value taste shock ε_{ij}. The deterministic component associated with buying product j depends on an intercept term δ_j, the product's price $price_{jt}$ in period t, and a vector of product characteristics X_j.[12]

In principle, one could estimate the conditional distribution of consumer types flexibly for each consideration set. However, in most applications (including ours), not all consideration sets are observed frequently enough to do this. We therefore take a more parsimonious approach, but still allow (α_i, β_i) to vary across consumers via their choice of consideration set. In particular, we assume the coefficient β_{ki} on the kth characteristic X_{kj} is given by

$$
\beta_{ki} = \bar{\beta}_k + \tilde{\beta}_k \tilde{X}_{ki}, \qquad (7)
$$

where $\tilde{X}_{ki}$ is the average value of that characteristic across all products in consumer i's consideration set, normalized such that it lies between 0 and 1.[13] Hence, $\bar{\beta}_k$ denotes the preference for characteristic k among consumers with the lowest average value of that characteristic in their consideration set, and $\tilde{\beta}_k$ captures differences in conditional preferences for characteristic k as a function of $\tilde{X}_{ki}$. Heterogeneity in the price coefficient α_i is modeled similarly.

This approach captures the idea that consumers with a strong preference for a specific characteristic are more likely to include products with that characteristic in their consideration set, and so we expect $\tilde{\beta}_k > 0$. It allows us to approximate a type of dependence that is natural in a search model (see discussion in the next section), but has the advantage that we can estimate the nature of the dependence from the data rather than imposing it through assumptions on the search process. Notice that for simplicity, we are modeling coefficients as a deterministic function of the average value of the relevant characteristic in a given consideration set, rather than allowing for a distribution of preferences that varies across consideration sets.

---

once during the sample period, ensuring that substitution effects for most pairs of products are not set to zero.

[12] We omit the time-subscript t from utility, because each consumer is observed in only one time period.

[13] We first compute a simple average of the characteristic within the set $\bar{X}_{ki}$. We then normalize the variable as follows: $\tilde{X}_{ki} = \left[\bar{X}_{ki} - \min\left(\bar{X}_{ki} \right) \right] / \left[\max\left(\bar{X}_{ki} \right) - \min\left(\bar{X}_{ki} \right) \right]$, where the min and max operators are taken over all consideration sets in the data.

8

2.3 Comparison to Other Approaches

We now briefly compare our approach with search and consideration models that are commonly used in the literature, and also draw a parallel to demand models based on choice heuristics.

2.3.1 Consideration-Set Models

In a typical consideration-set model, the consideration set probabilities are

$$\Pr(s|\mathbf{p}, \mathbf{X}) = \int \prod_{j \in s} \Pr(search\ j|X_{1j}, Z_i, \gamma) \mathrm{d}F_Z(Z_i), \tag{8}$$

where Z_i are consumer characteristics with distribution function F_Z, and γ is a vector of parameters to be estimated. Furthermore, X_{1j} are product characteristics that influence consumers' awareness of the product such as advertising (Goeree (2008), Barroso and Llobet (2012)) or whether it is prominently displayed within a store (Bronnenberg and Vanhonacker (1996), Swait and Erdem (2007)). Price is assumed not to influence consideration, and the probability of a given set occurring is simply obtained by multiplying the relevant product-level consideration probabilities. Meanwhile, conditional utilities are typically defined as in our earlier equation (6), that is,

$$u_{ij} = \delta_j + \alpha_i price_{jt} + X'_{2j}\beta_i + \varepsilon_{ij}, \tag{9}$$

where X_{2j} is another set of product characteristics that do not overlap with those in X_{1j}. Importantly, the characteristics X_{1j} are assumed to be uncorrelated with consumers' preferences over price and X_{2j}. This assumption implies the distribution of the preference parameters (α_i, β_i) is the same across all consideration sets.

Our approach nests the above model but is more flexible because we allow price to affect consideration, and because we allow conditional utilities to vary with consideration sets. Moreover, because we estimate $\Pr(s|\mathbf{p}, \mathbf{X})$ using a (modified) frequency estimator, we avoid making assumptions on the nature of the process by which consideration sets are formed. Crucially, we do not impose that product-level inclusion probabilities are independent of each other. We believe relaxing this assumption is important because conditional viewing probabilities are unlikely to be independent in online settings due to recommendations of similar products ("consumers who viewed this product also viewed ..."). Other elements of webpage layout, such as product rankings, are also likely to influence consideration sets and so should be part of X_{1j} in equation (8), and yet are hard to measure and are rarely recorded. Our approach sidesteps the need to estimate the influence of such variables by directly estimating consideration-set probabilities.

2.3.2 Consumer Search Models

In a typical search model, both search and choice are driven by an underlying (characteristics-based) utility function similar to the one in equation (6). Consumers search either sequentially or

9

simultaneously, and may know the realizations of some subset of characteristics prior to search.

In most search models, selection into different consideration sets will depend on consumer preferences. For example, if consumers tend to search products with similar price levels, a search model will rationalize this behavior through heterogeneity in preferences over price. Such heterogeneity would induce high (low) price sensitivity consumers to search sets of products with low (high) prices. However, if two products do not share any observable characteristics, a search model will not be able to rationalize that the pair of products might be frequently searched together. Our approach, by contrast, directly estimates consideration-set probabilities and can therefore capture co-search patterns that are not driven by similarity in observed characteristics.

Furthermore, selection based on characteristics will lead consumers who have a strong preference for a specific characteristic to select a consideration set containing products with high values of this characteristic. However, the exact way in which $\Pr(s|\mathbf{p}, \mathbf{X})$ varies with preference parameters will depend, for example, on what is assumed about the search protocol and what consumers know before engaging in search.[14] Our approach captures these kinds of dependencies by allowing preference parameters of consumers who consider a particular consideration set to depend on the characteristics of the selected consideration set rather than imposing the nature of selection via assumptions about the search process.

2.3.3 Choice-Heuristics Models

Choice-heuristics models are rule-based strategies that help consumers simplify decisions and reduce cognitive effort. They may be particularly relevant in online settings where assortments are large. In many of these models, consumers use a two-stage decision process, where in the first stage they use simple non-compensatory rules to narrow down the set of options, and then in the second stage they use a compensatory mechanism to make a final choice.[15]

Our approach nests the special case in which price is not used for screening products in the first stage and the two stages are independent (which is then similar to consideration-set models). Our approach also approximates the general case, which is similar to a search model in the sense that at the first stage, consumers will tend to select into consideration sets containing products with characteristics that they value highly. As already discussed, our approach captures this dependence without needing to take a stand on the particular heuristic being used.

2.3.4 Benefits of Our Approach

As discussed above, consideration-set models typically assume (i) price does not affect consideration, and (ii) the distribution of preferences does not differ across consideration sets. On the other hand, in search and choice-heuristics models, those assumptions do not always hold. Our approach nests

---

[14]Honka and Chintagunta (2016) show that in the context of search over price, the distribution of prices in a given consideration set differs under sequential and simultaneous search.

[15]See Aribarg, Otter, Zantedeschi, Allenby, Bentley, Curry, Dotson, Henderson, Honka, Kohli, Jedidi, Seiler, and Wang (2017) for a recent summary of the literature.

consideration-set models and provides an approximation to structural models of search and choice heuristics. In both cases, we believe our approach has some important advantages.

One benefit of our approach is that it does not "unpack" the consideration process, and instead estimates it directly from the data. Therefore, relative to the models outlined above, our approach avoids imposing strong (and often untested) assumptions on what consumers know or on their search/consideration process.[16] Avoiding such assumptions is particularly advantageous in online markets, where consumers face a large number of choices and may differ substantially in what information they have and how they use it to evaluate products. An approach like ours, which approximates the underlying search or consideration process, could dominate an incorrectly specified model.

Another benefit of our approach is that it flexibly allows factors that are not observed by the researcher to influence substitution patterns. One example would be the layout of a webpage, which is likely to greatly influence which products are searched together (either by displaying certain products together, or by recommending products similar to those already viewed by the consumer). To our knowledge, search and consideration models would typically not capture such drivers of pair-specific search.[17] Furthermore, in online markets, the number of products is large relative to the number of product characteristics that are observed, and the researcher might not observe relevant characteristics that drive substitution between products. These features of the data make using random-coefficient models to flexibly capture substitution patterns difficult. Because our approach directly estimates the distribution of consideration sets from the data, it allows these unobserved factors to influence estimated elasticities.

A final benefit of our approach is computational. Structural search models are typically very computationally burdensome, and thus are usually estimated for small numbers of products. Given the current methods of estimation, we think scaling a structural model of search to a large assortment size of several hundreds of products is computationally infeasible. As we demonstrate later, our demand model is considerably simpler to estimate.

3   Data and Descriptive Statistics

We estimate the model using data from an online retailer that sells home-improvement products (hardwood flooring, tiles, etc.). We focus on a single product category, which is one of the largest sold by the retailer and contains 576 products during our sample period. We observe the entire history of consumers' search and purchase behavior during a 13-week period from April 20, 2016, to July 16, 2016. The final data set contains 454,977 searches and 12,626 purchases (basket additions) by 185,963 distinct users. A consumer is considered to have searched a product if she accessed the product description page. We treat basket additions as the choice outcome in the demand model,

---

[16]With regards to search protocol, De Los Santos, Hortaçsu, and Wildenbeest (2012) and Honka and Chintagunta (2016) test whether search is sequential or simultaneous. To the best of our knowledge, other assumptions (e.g., whether search is over price or a "match value") have not been tested in the literature.

[17]Because utility is often defined at the product level, factors that guide a consumer from one product to another (e.g., recommendations) are hard to include, because they are specific to pairs of products.

		Percentage
Purchase probability		6.8%
Average # of Searches		2.45
		Percentage
Number of	1	56.41
Searches	2	17.66
	3	8.74
	4	5.17
	5	3.34
	6	2.13
	7	1.52
	8	1.11
	9	0.80
	≥10	3.12
List of Product		Price
Characteristics		Deal Dummy (Price Decrease that is Highlighted on the Webpage)
		Ratings (Customer Reviews): 1-5 Scale
		Number of Customer Reviews
		Brand
		3 Discrete Characteristics (Anonymized)
		3 Continuous Characteristics (Anonymized)

Table 1: **Descriptive Statistics.**

because the retailer did not store purchase information in a way that was easily accessible to us. (Basket additions were, however, tracked as part of the browsing data.) We assume the rate of conversion from basket additions to purchases does not vary across products and is unaffected by price. Under this assumption the conversion rate simply scales up demand, and hence has no effect on estimated elasticities or optimal prices. Table 1 provides some descriptive statistics. Roughly 6.8% of all search sessions in this category end in a "purchase." On average, a search session contains 2.45 products.

We also have some information about each of the products. Each product belongs to one of 27 different brands. In addition to observing weekly prices, we also observe whether a product is "on deal," that is, is on sale in a given week and is highlighted as such on the website. We also have data on the number of reviews posted about a product, as well as their score on a scale from 1 to 5. Although these ratings could vary over time, given our short sample period, there is minimal variation. Finally, the retailer also provided us with six additional product characteristics that we are not allowed to disclose. Three of them are discrete and three are continuous. They measure physical characteristics of the products and are therefore time invariant.

In Table 2, we document search and purchase patterns at the product level and describe the distribution of consideration sets that occur in the data. The first two columns report the average number of purchases and searches separately for the top 100, 200, and so on products in the

	Average Number of Purchases	Average Number of Searches	Product-pair Co-search Patterns			
			Number of Product Pairs	Average Number of Searches	Median Number of Searches	Share of Zeros
Top 100 products	102	3142	4,950	93	39	0.26%
Top 200 products	58	1902	19,900	38	13	3.17%
Top 300 products	40	1378	44,850	22	6	8.69%
Top 400 products	31	1083	79,800	14	4	16.10%
Top 500 products	25	894	124,750	10	2	23.66%
All products (576)	22	789	165,600	8	2	28.44%

Table 2: **Purchase, Search and Joint-Consideration.**

assortment by market share. As mentioned earlier, we find that, consistently across the assortment, products are searched significantly more frequently than they are purchased. The ratio of searches to purchases is roughly 36 to 1.

We also document that for large parts of the assortment, most pairs of products are searched together (possibly with other products) relatively frequently. For example, for the top 100 products, we observe almost all pairs of products (out of $(100 * 99)/2 = 4950$ possible pairs) being searched together by at least some consumers with an average (median) number of joint searches of 93 (39) across all product pairs. Even when including 300 products (more than half of the assortment), the share of pairs not being searched together is below 10%.

3.1 Price Variation

Prices in our data vary both across products (at a given point in time) and over time (for a given product). Time-series variation in prices is driven either by temporary deals or by changes in the regular price implemented by the retailer's data-analytics team. Deals tend to be accompanied by other changes such as more salient display (a colored price tag alerting the consumer to the deal) and more prominent placement on the webpage. As a result, consumers' reaction to deals is likely to be attributable not just to the price change, but also to other elements that change alongside it. We therefore control for deal status and estimate the impact of price on consideration and conditional choice entirely from within-product changes in regular price. We later hold deal status constant when computing price elasticities, and use those elasticities to solve for optimal *regular* prices.

We have ample within-product price variation. During our 13-week sample, the average product changes price 1.2 times. Many products (300) change price at least once, and among these products, the magnitude of the average price change is equal to 10.8% of the average product price over the 13 weeks. Most products (453) are never "on deal." Conditional on having deal status at least once, the median product has it for three out of the 13 weeks.

We treat regular price changes as exogenous. Based on conversations with the company, this

Dependent Variable	(1) # Joint Searches	(2) Joint Search Ratio	(3) Log # Joint Searches	(4) # Joint Searches	(5) Joint Search Ratio	(6) Log # Joint Searches
Mean	7.96	0.42	1.22	7.96	0.42	1.22
S.D.	41.93	1.17	1.14	41.93	1.17	1.14
Same Brand	19.534***	0.809***	0.834***			
	(0.364)	(0.010)	(0.009)			
Price Difference	-1.922***	-0.046***	-0.213***			
	(0.106)	(0.003)	(0.003)			
Same Char. 1	13.667***	1.047***	0.749***			
(Discrete)	(0.854)	(0.023)	(0.022)			
Same Char. 2	2.881***	0.070***	0.097***			
(Discrete)	(0.239)	(0.007)	(0.006)			
Same Char. 3	2.675***	0.190***	0.045***			
(Discrete)	(0.275)	(0.007)	(0.007)			
Diff. Char. 4	-1.336***	-0.115***	-0.110***			
(Continuous)	(0.105)	(0.003)	(0.003)			
Diff. Char. 5	--1.678***	-0.064***	-0.110***			
(Continuous)	(0.103)	(0.003)	(0.003)			
Diff. Char. 6	-1.890***	-0.020***	-0.108***			
(Continuous)	(0.107)	(0.003)	(0.003)			
Similarity Score				6.587***	0.321***	0.249***
				(0.102)	(0.003)	(0.003)
Products	576	576	576	576	576	576
Observations	165,600	165,600	165,600	165,600	165,600	165,600
R-squared	0.036	0.083	0.161	0.025	0.075	0.048

Table 3: **Determinants of Co-search of Product Pairs.** The joint-search ratio is defined as [# Searches (j,j')]/[# Searches (j)×# Searches (j')]. Characteristics 1-3 are discrete variables. Regressors are defined as a dummy equal to 1 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.

assumption is reasonable because most price changes were part of an attempt to induce price variation in order to understand how responsive demand is to such changes. Of course, some price changes could be triggered by changes in demand that the firm is trying to adapt to. However, within the short time frame of our data (13 weeks), we think that large changes in product-level demand are unlikely.

3.2 Browsing Data and Product Characteristics

Next, we examine the relationship between the search data and product characteristics. First, we show that similarity in characteristics helps predict the likelihood of two products appearing in the same consideration set. Hence, the search data allow us to capture product similarity that is driven by observed characteristics. Second, however, we find that a large part of the variation in search patterns remains unexplained by flexible measures of similarity in observed characteristics.

In more detail, we measure the closeness of products in characteristic space by using price, brand identity, and the six physical characteristics that we mentioned earlier. For all discrete variables, we define a dummy that is equal to 1 if the variable takes the same value for both products j and j'. (For example, one of the regressors is a dummy that takes the value of 1 if two products belong to the same brand.) For all continuous variables, we compute the absolute value of the difference between the two products, and then to facilitate comparisons, we normalize it by the variable's standard deviation.

In column (1) of Table 3, we regress the number of times a pair of products (j, j') was searched together during the entire sample period on the measures of product closeness that we just described. All characteristics have a significant impact and the coefficients have the expected sign. Sharing the same discrete characteristic increases joint search, whereas a larger difference in any continuous characteristic lowers joint search. Moreover, some coefficients are relatively large in magnitude, for example, belonging to the same brand increases the number of joint searches by roughly half a standard deviation. However, importantly, the r-squared is only equal to 0.036, and therefore most of the variation in search patterns is not explained by closeness in characteristic space.

The remaining columns of Table 3 probe the robustness of this result. In column (2), the dependent variable is

$$\frac{\sum_i \mathbf{1}((j, j') \in s_i)}{(\sum_i \mathbf{1}(j \in s_i)) \times (\sum_i \mathbf{1}(j' \in s_i))} , \tag{10}$$

where, for example, $\mathbf{1}(j \in s_i)$ takes the value of 1 if product j is contained in consumer i's consideration set. In other words, the dependent variable is the number of consumers who searched (j, j') together, divided by the product of the number of consumers who searched j and j', respectively. This metric adjusts for the fact that products that are searched more often will automatically have higher joint search with any other product. In column (3), we use a logarithmic transformation of the number of joint searches as the dependent variable.[18] The r-squared is higher in both specifications but still low in absolute value. In columns (4) to (6), we run the same set of regressions but use as the regressor a similarity score that is computed by the firm and is based on the characteristics we have already used. Again, the predictive power of these regressions is relatively low. In unreported regressions, we also probe robustness to removing outliers and including higher-order terms of all covariates, and find the results do not change qualitatively.

In summary, these regressions suggest much of the variation in search patterns cannot be explained by observed characteristics. Therefore, an important advantage of our approach is that we

---

[18]Before taking the logarithm, we add 1 to the number of joint searches.

15

directly use information on joint-search patterns and leverage them to estimate cross-elasticities. We note the patterns described above do not establish that the search data (and our way of using this data) allow us to obtain better estimates of substitution patterns. We return to this point in section 4.2.1 when we assess the fit of our model against a characteristics-based (full-information) random-coefficient model.

4 Estimation and Results

We outlined the general structure of the demand model in section 2. We briefly re-cap the relevant equations here and describe a few modifications that are required to adapt the model to our empirical setting.

With regards to consideration-set formation, we slightly modify equation (5) from earlier to include deal status. We assume the probability of consumer i choosing set s is given by

$$\Pr(s|\mathbf{p}, \mathbf{X}) = \frac{\exp(\chi_s + \gamma_1 \overline{price}_{st} + \gamma_2 \overline{deal}_{st})}{\sum_{s' \in \mathcal{S}} \exp(\chi_{s'} + \gamma_1 \overline{price}_{s't} + \gamma_2 \overline{deal}_{s't})}, \tag{11}$$

where $\overline{price}_{st}$ and $\overline{deal}_{st}$ denote, respectively, the average price and the fraction of "on deal" products in consideration set s. The parameters to be estimated are γ_1 and γ_2 as well as a vector of set fixed effects.

With regards to conditional choice, we also slightly modify equation (6) from earlier to include deal status. We assume a consumer's utility function takes the form

$$
\begin{aligned}
u_{ij} &= \bar{u}_{ij} + \varepsilon_{ij} = \delta_j + \alpha_{1i} price_{jt} + \alpha_{2i} deal_{jt} + X'_j \beta_i + \varepsilon_{ij} \tag{12} \\
u_{i0} &= \varepsilon_{i0},
\end{aligned}
$$

where $price_{jt}$ is product j's price, and $deal_{jt}$ is a dummy variable that equals 1 if product j is "on deal." Following our earlier discussion, we assume the coefficients on the kth product characteristic, price, and deal are given by, respectively,

$$
\begin{aligned}
\beta_{ki} &= \bar{\beta}_k + \tilde{\beta}_k \tilde{X}_{ki}, \tag{13} \\
\alpha_{1i} &= \bar{\alpha}_1 + \tilde{\alpha}_1^{price} \widetilde{price}_i + \tilde{\alpha}_1^{deal} \widetilde{deal}_i, \\
\alpha_{2i} &= \bar{\alpha}_2 + \tilde{\alpha}_2^{price} \widetilde{price}_i + \tilde{\alpha}_2^{deal} \widetilde{deal}_i,
\end{aligned}
$$

where $\tilde{X}_{ki}$, $\widetilde{price}_i$, and $\widetilde{deal}_i$ are the (normalized) average values of characteristic k, price, and deal status across all products in consumer i's consideration set. Because the price level and deal status are both price-related variables, we allow the coefficients on those two variables to depend on both the average price and average deal status in consumer i's consideration set.

We also assume that with probability ϑ, which is independent of the utilities that a consumer draws, a search spell breaks down and therefore leads to no purchase. An example for the type of behavior this parameter captures is that of consumers who want to purchase in a physical

16

store but use the webpage to find their preferred product. Such consumers are better classified through search breakdown rather than the choice of the outside option, because they did find their utility-maximizing product, and their decision not to purchase on the webpage is unrelated to price. Empirically, the breakdown parameter helps us match the degree of substitution between products and the outside good in a parsimonious way. To understand this, notice that when the breakdown parameter is larger, fewer consumers respond to a price decrease on a particular product by substituting away from the outside option. Equivalently, substitution between products is more important than substitution away from the outside option. Intuitively, ϑ is identified by the responsiveness of outside-option choices to movement in prices.

We allow the breakdown parameter to vary as a function of the (normalized) average price and average deal dummy in the same way as the price and deal coefficients; hence,

$$\vartheta_i = \bar{\vartheta} + \tilde{\vartheta}^{price}\widetilde{price}_i + \tilde{\vartheta}^{deal}\widetilde{deal}_i. \tag{14}$$

The probability that consumer i with consideration set s chooses product $j \in s$ is then given by

$$\Pr(j|s, \mathbf{p}, \mathbf{X}) = [1 - \vartheta_i(\tilde{\mathrm{X}}_i)] \frac{\exp[\bar{u}_{ij}(\tilde{\mathrm{X}}_i)]}{1 + \sum_{l \in s} \exp[\bar{u}_{il}(\tilde{\mathrm{X}}_i)]}, \tag{15}$$

where $\tilde{\mathrm{X}}_i = \left\{\widetilde{price}_i, \widetilde{deal}_i, \tilde{X}_{1i}, ..., \tilde{X}_{Ki}\right\}$ denotes the average price, deal dummy, and product characteristics in the consumer's consideration set (and K denotes the number of characteristics that drive selection into consideration sets). The probability of choosing the outside option is given by

$$\Pr(0|s, \mathbf{p}, \mathbf{X}) = \vartheta_i(\tilde{\mathrm{X}}_i) + [1 - \vartheta_i(\tilde{\mathrm{X}}_i)] \frac{1}{1 + \sum_{l \in s} \exp[\bar{u}_{il}(\tilde{\mathrm{X}}_i)]}. \tag{16}$$

The set of parameters to be estimated includes product fixed effects, the parameters governing the conditional distribution of preferences over price, deal, and other characteristics, as well as the breakdown probability.

We estimate both the conditional-choice model and the consideration-set model by maximum likelihood. Because the two models have no common parameters, they can be estimated separately. In both models, we employ a contraction mapping in the spirit of Berry, Levinsohn, and Pakes (1995) to solve for the vector of product and set fixed effects, respectively.

4.1 Estimation Results

The top panel of Table 4 reports results for the consideration-set formation process. We do not report the large number of consideration-set fixed effects. The impact of average price and average deal dummy both have the expected sign; that is, a consideration set becomes less likely when its average price is higher and more likely to be chosen if more products in the set are on deal. Both coefficients are precisely estimated.

The bottom panel of Table 4 reports results for three different specifications of the conditional-

17

Panel A: Set Formation

Consideration-Set Characteristic	Coeff.	S.E.
Average Price	-1.329	0.039
Average Deal	0.371	0.010
Set Fixed Effects	Yes	

Panel B: Conditional Choice

Product Characteristic	Heterogeneity	*Model 1* Coeff.	S.E.	*Model 2* Coeff.	S.E.	*Model 3* Coeff.	S.E.
Price		-0.844	0.163	-1.187	0.192	-1.068	0.192
Deal Dummy		0.132	0.035	0.811	0.097	0.899	0.097
Breakdown		0.634	0.012	0.488	0.026	0.452	0.031
Price	High Price Consideration Sets			2.074	0.324	1.687	0.338
	High Deal Consideration Sets			-0.107	0.122	-0.089	0.127
Deal Dummy	High Price Consideration Sets			-3.891	0.474	-4.189	0.473
	High Deal Consideration Sets			-0.355	0.114	-0.447	0.114
Breakdown	High Price Consideration Sets			0.500	0.029	0.534	0.034
	High Deal Consideration Sets			-0.048	0.045	-0.051	0.050
Customer Rating	High Rating Consideration Sets					0.126	0.032
# Reviews	Large # Reviews Consideration Sets					-0.071	0.035
Characteristic 2	High Char. 2 Consideration Sets					-0.027	0.044
Characteristic 3	High Char. 3 Consideration Sets					0.313	0.094
Characteristic 4	High Char. 4 Consideration Sets					0.110	0.040
Characteristic 5	High Char. 5 Consideration Sets					0.061	0.059
Characteristic 6	High Char. 6 Consideration Sets					0.551	0.231
Product Fixed Effects		Yes		Yes		Yes	
Out-of-sample Likelihood		-0.57535		-0.57498		-0.57533	
# Observations		185,963		185,963		185,963	

Table 4: **Estimation Results.** Characteristics 2-6 are the same anonymized characteristics used in Table 3. Characteristic 1 is not used in the demand model. The out-of-sample likelihood is the average (per observation) likelihood of 13 "leave-one-week-out" estimates, where 12 weeks are used for estimation and fit is evaluated for the one week left out of estimation.

choice model. As a benchmark, we present results from a simple specification under the heading "Model 1." This basic model contains three parameters to be estimated: α_1, α_2, and ϑ as well as a

set of product fixed effects, but does not allow for heterogeneity in the conditional distribution of preferences as a function of the chosen consideration set. Results for the large set of fixed effects are suppressed. We find the price and deal coefficients are precisely estimated and have the expected sign. The breakdown probability is equal to 63.4%, and hence a large share of the 93.2% of search spells without a purchase is due to search breaking down.

In Model 2, we add interaction terms to the price, deal, and breakdown coefficients. As expected, we find consumers who search more expensive items are less price sensitive, and consumers who search more "on deal" products are more price sensitive. The price-interaction term is statistically significant, but the deal interaction is not. In terms of magnitude, a one-standard-deviation shift in the average price in the consideration set leads to a 0.188 shift in the price coefficient.[19] We also find search spells of consumers who search cheap and on-deal products are less likely to break down.

Model 3 also adds heterogeneity in conditional preferences for other characteristics. The characteristics that we use are the average customer rating of a product and the number of reviews, as well as five of the six anonymized product attributes that we described earlier in section 3 and that were used in the regressions reported in Table 3. (We do not use the sixth product attribute or brand data, because both variables take 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 most of the coefficients are positive, and four are statistically significant. This finding is intuitive because, as we discussed earlier, consumers who predominantly search products with a particular characteristic are likely to value that product characteristic more. The coefficients on the price-related variables (including the interaction terms) only change marginally compared to Model 2.

We assess out-of-sample fit by estimating each model on 12 out of 13 weeks and evaluating fit based on the unused week. We estimate each model 13 times for each possible permutation of training and test sample and compute the average (per observation) likelihood across the 13 samples.[20] The fit statistic for each model is reported at the bottom of Table 4. When comparing Models 1 and 2, we find that allowing price-related variables to differ across consideration sets improves the model fit. Interestingly, adding further heterogeneity in preferences over other characteristics across consideration sets (in Model 3) actually reduces fit. We hence consider Model 2 to be our preferred specification, and elasticities and optimal prices (which we report below) are based on this specification. We also note that despite the change in fit, elasticities and optimal prices based on Models 2 and 3 are very similar to each other.[21]

---

[19]We note the price coefficient is positive for large values of the average set price. However, the share of observations where this happens is very small (0.5%) and is due to a few outlier values of the average set price.

[20]We focus on the purchase likelihood. Predicted purchase probabilities are based on equation (3) and depend on predicted consideration-set probabilities $Pr(s|\mathbf{p}, \mathbf{X})$ and predicted conditional-choice probabilities $Pr(j|s, \mathbf{p}, \mathbf{X})$ given the relevant week-specific price vector.

[21]Across all products, the median (maximum) change in the optimal price (in absolute value) between Models 2 and 3 is 1.1% (3.7%). Moving from Model 1 to 2 leads to a median (maximum) change of 21.6% (30.5%).

4.2 Comparison to Full-Information Demand Model

Next, we assess the performance of our approach relative to a more standard characteristics-based demand model that ignores the search data.

To provide a fair comparison, we use a relatively flexible utility function for the perfect-information model that contains a full set of product fixed effects and allows for random coefficients on nine product characteristics (including two price-related variables). Specifically, utility is

$$
\begin{aligned}
u_{ij} &= \psi_j + \zeta_{1i} price_{jt} + \zeta_{2i} deal_{jt} + X_j' \lambda_i + e_{ij} && (17) \\
\zeta_{1i} &= \bar{\zeta}_1 + \sigma_1 \nu_{1i} \\
\zeta_{2i} &= \bar{\zeta}_2 + \sigma_2 \nu_{2i} \\
\lambda_i &= \bar{\lambda} + \sigma_\lambda \nu_{\lambda i},
\end{aligned}
$$

where ψ_j is a product fixed effect and $(price_{jt}, deal_{jt}, X_j)$ denote price, a deal dummy, and other product characteristics. The various coefficients on price and other characteristics allow for a mean effect and a set of variance terms $(\sigma_1, \sigma_2, \sigma_\lambda)$ that capture unobserved heterogeneity. $(\nu_1, \nu_2, \nu_\lambda)$ are standard normally distributed. $\bar{\lambda}$ is not separately identified, because X_j does not vary over time and the model includes product fixed effects. The taste shock e_{ij} is extreme-value iid.

Note we attempt to keep the general structure of utility similar to the one used in our consideration-and-choice model.[22] In both our main model and the full-information model, we have a full set of product fixed effects, and price and deal status enter utility. Our main model leverages the search data by estimating consideration-set probabilities and the distribution of preferences *conditional on consideration*. Naturally, those aspects are not part of the full-information model. Instead, the full-information model is enriched by allowing for random coefficients on price, deal, and a set of product characteristics X_j. The latter are the same characteristics used in Model 3 for the purpose of linking consideration and choice. Here, we instead allow for random coefficients on those characteristics in order for similarity in those characteristics to drive substitution patterns. We also note that we include all characteristics that are routinely coded by the retailer and that are used internally to compute similarity scores between product pairs.[23] Detailed estimation results for the full-information model are presented in Table A1 in the appendix.

We do not attempt to estimate a structural search model of the kind estimated elsewhere in the literature (e.g., Honka (2014), Chen and Yao (2016)). As we show below, the full-information model is already computationally demanding due to the size of our data. A structural model of search would be substantially more computationally involved and does not appear feasible in a context with as many products (and consumers) as our setting.[24]

---

[22]We also attempted to include an equivalent term to the "breakdown parameter" in the full-information model, but found its estimated value was close to 1. Hence, we dropped the parameter from the model specification.

[23]Similarity scores are used to recommend products ("consumers who viewed this product also viewed...").

[24]Estimating structural search models is challenging because the joint likelihood of search and purchase decisions has no closed form and must be simulated for every evaluation of the likelihood function. Therefore, most papers in

	Preferred Specification (Model 2)	Full Information Model A	Full Information Model B	Full Information Model C
		Product FEs, Price & Deal Coeff.	Full Info. Model A + RC on Price & Deal	Full Info. Model B + RC on 7 Char.
Model Fit (Out-of-sample Likelihood)	-0.57498	-0.57530	-0.57532	-0.57801
Computational Burden (Estimation Time in Minutes)	38	23	217	2,812

Table 5: **Model Comparison: Out-of-sample Fit and Computational Burden.**

4.2.1 Model Fit

To provide a benchmark against which to evaluate our main model and the full-information random-coefficient demand model outlined above, we first estimate a simple model that does not allow for flexibility in substitution patterns. In particular, our baseline specification is a full-information logit model that follows the specification outlined in equation (17), but contains no random coefficients, that is, the utility function includes product fixed effects, price, and a deal dummy. This model suffers from the well-known restrictive nature of substitution patterns of logit models and contains no parameters that cater specifically to flexibility in cross-price elasticities. We then add random coefficients, first only on price and deal, and then on an additional set of seven characteristics.

To assess fit, we compute the average out-of-sample likelihood for 13 leave-one-week-out samples, following the procedure described at the end of section 4.1. We also report the same fit statistic for the preferred model specification (Model 2). To make predictions comparable, we focus on predictions in terms of purchase probabilities. In the case of our consideration-and-choice model, we derive predicted purchase probabilities based on equation (3) from predicted consideration-set probabilities $Pr(s|\mathbf{p}, \mathbf{X})$ and predicted conditional-choice probabilities $Pr(j|s, \mathbf{p}, \mathbf{X})$ given the relevant week-specific price vector. In other words, we do not use information on the realizations of consideration sets in the test sample.

The main results in terms of model fit are reported in the first line of Table 5. We find our preferred model leads to a higher out-of-sample likelihood than any of the full-information demand models. Interestingly, adding random coefficients to the full-information model worsens out-of-sample fit. Fit decreases slightly when adding random coefficients on price and the deal dummy, and decreases even further when adding random coefficients on the additional seven characteristics. As Table A1 in the appendix shows, the majority of the variance terms of the random coefficients are not statistically significant, which is in line with the finding that adding such terms does not improve the overall model fit.

---

the literature estimate them for choice sets containing a relatively small number of products.

	Mean	25th Perc.	Median	75th Perc.
All Cross-Price Elasticities	0.00142	0.00018	0.00061	0.00160
Max Cross-Price Elasticities	0.080	0.070	0.085	0.092
Top 10 Cross-Price Elasticities	0.039	0.029	0.034	0.041
Top 11-20 Cross-Price Elasticities	0.018	0.015	0.018	0.020
Top 21-30 Cross-Price Elasticities	0.011	0.009	0.011	0.012
Own-Price Elasticities	-2.055	-1.744	-1.988	-2.280

Table 6: **Cross- and Own-Price Elasticities.**

In conjunction with the descriptive patterns documented in section 3.2, the results regarding model fit are intriguing. We documented earlier that search patterns are not well predicted by similarity in characteristics between pairs of products. Above, we have shown that enriching a full-information model with heterogeneous tastes for characteristics does not lead to an improvement in model fit. However, the search data used in our main specification do provide a higher out-of-sample fit. This finding suggests the search data capture something other than similarity in observed characteristics, and that they are more predictive of substitution patterns than similarity in characteristics. Elasticities and optimal prices derived from the full-information model also differ substantially, which we expand on below.

4.2.2 Computational Considerations

Apart from providing an improvement in fit, our approach is also substantially computationally lighter than the full-information model, due to two features of our model. First, the conditional-choice model is computationally light because, conditional on consideration, consumers choose from sets that contain 2.45 products on average. In the full-information model, we need to model choice from the full set of almost 600 products. Second, our model does not involve any numerical integration, which is required when modeling random coefficients in the full-information approach.[25]

Some key statistics on computational time are presented in the lower panel of Table 5. Our preferred specification, Model 2, is about 75 times faster to estimate relative to the full-information model with nine random coefficients.

5 Elasticities and Optimal Prices

5.1 Elasticities

Our model yields estimates of own- and cross-price elasticities for the full assortment of 576 products. We now present various aggregate statistics that summarize the distribution of these elasticities.

---

[25]We use 5 draws per consumer when integrating out the random-coefficient terms in the full-information model.
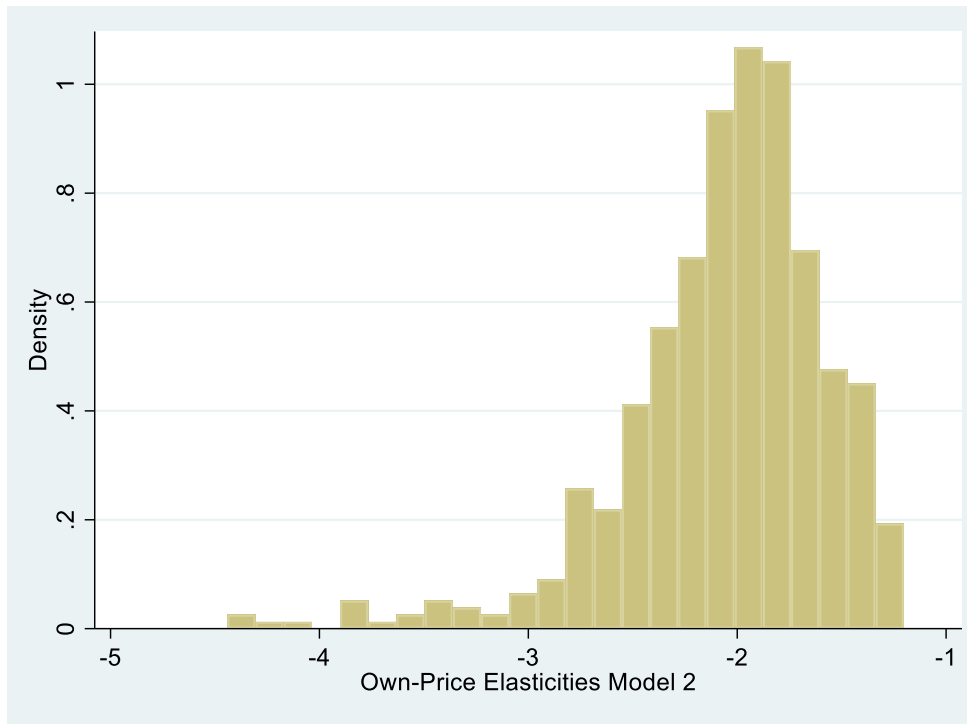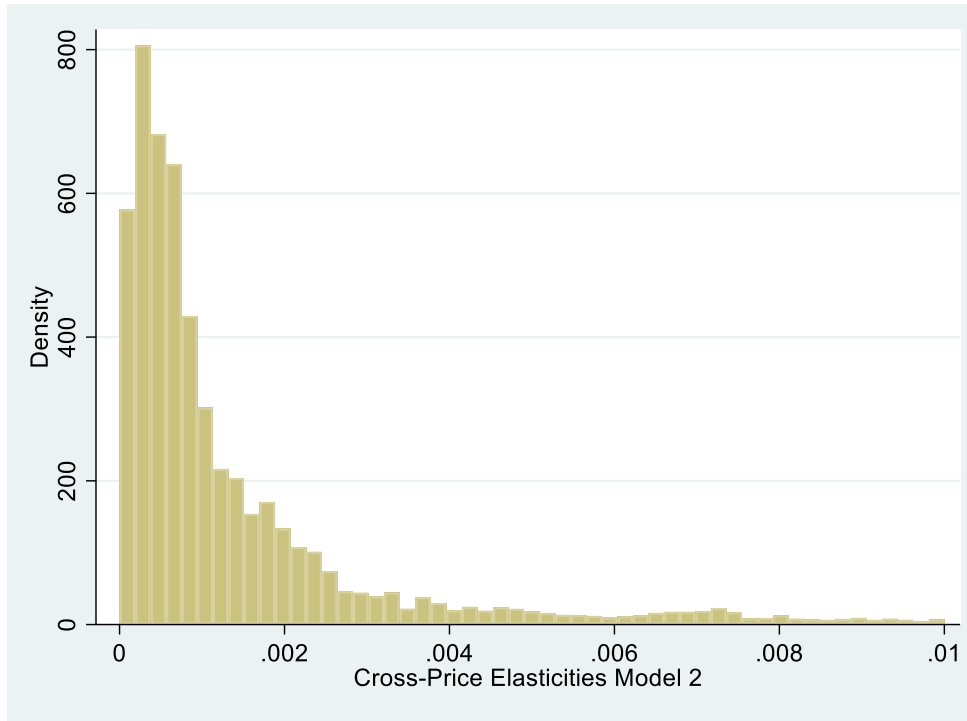
Figure 2: **Distribution of Cross- and Own-Price Elasticities.**

In Table 6, we report summary statistics of the distribution of cross-elasticities, whereas the top graph in Figure 2 presents a histogram of all cross-price elasticities.[26] We find the average cross-price elasticity across all product pairs is equal to 0.00142, whereas the median is equal to 0.00061. These numbers are small because any given product is likely to have few close substitutes, and will therefore have relatively small cross-price elasticities with most products in the assortment. To illustrate this pattern, we compute for each product the highest cross-price elasticity with any other product in the assortment. We find the average (across products) maximum cross-price elasticity is equal to 0.080, and the dispersion across products is relatively low. We also report the distribution of the top 10, top 11-20, and top 21-30 elasticities for each product. As one might expect, these elasticities decline relatively rapidly, again suggesting most products have a small set of important substitutes.

To highlight the strength of using search data to estimate cross-price elasticities, we report heat maps of cross-price elasticities for the top 40 products by market share. We report such a heat map for our preferred specification (Model 2) and for the random-coefficient full-information model (Model C) in Figure 3. The color coding is identical in both figures and spans a range of elasticities from 0 (white) to ≥ 0.03 (dark red). One notable feature is that the random-coefficient elasticity matrix displays little variation across rows for a given column, meaning a change in one product's price shifts demand relatively equally across all products, regardless of how (dis)similar the products are, because the random-coefficient variance terms are small in magnitude and imprecisely estimated. Consequently, elasticities from the random-coefficient model are almost identical to elasticities from a simple logit model.[27] By contrast, the elasticity matrix of the consideration-and-choice model displays much more variation at the product-pair level.

We also report the distribution of own-price elasticities in Table 6 and Figure 2. The average elasticity is equal to -2.055 and own elasticities are roughly normally distributed with a left tail of products with higher (in absolute terms) own elasticities.

5.2 Optimal Prices

In this section, we describe how optimal prices derived from our demand estimates relate to the retailer's current prices. The retailer in our application sells two different types of products—"retail" products, where the retailer sets prices, and "marketplace" products, where third-party sellers set prices and the retailer receives a percentage of the transaction price in case of a sale.[28] We solve for optimal prices for the retail products, assuming the prices of marketplace products remain constant. Based on conversations with the company, during our sample period, it was maximizing revenue in order to grow market share. Therefore, to facilitate a comparison with the

---

[26]We restrict the histogram to cross-price elasticities below 0.01, which corresponds roughly to the 95th percentile of the distribution.

[27]Based on the standard formula, logit elasticities of demand for product j with respect to the price of product k depend only on the price level and market share of product k. Specifically, the elasticity is given by $\zeta_1 \times p_k \times Pr(k)$, where ζ_1 is the price coefficient (see equation (17)).

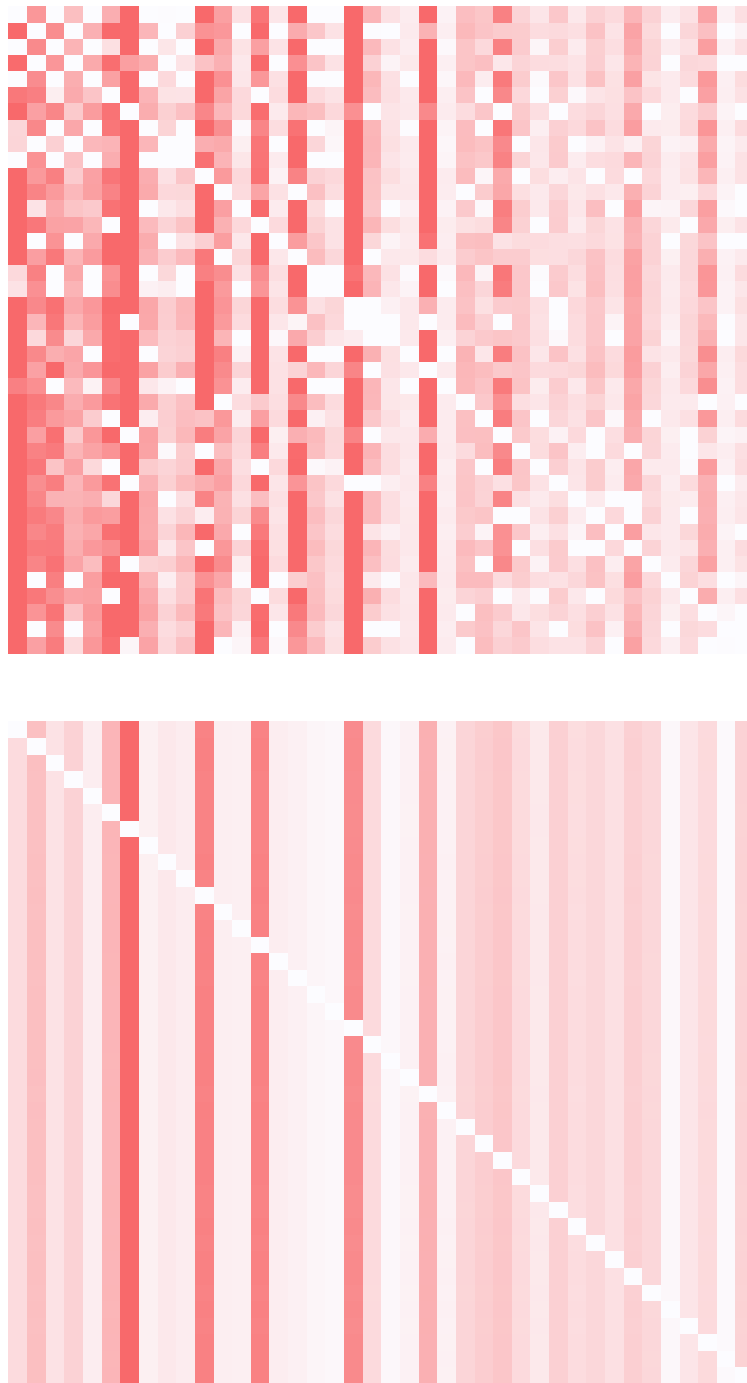[28]Of the 576 products, 200 are retail products and they make up 79% of all purchases.

Figure 3: **Heat Map of Cross-Price Elasticities. Top: Model 2 (Consideration-and-Choice Model), Bottom: Model C (Full-Information Random-Coefficient Model).** A cell represents the elasticity of row demand with respect to column price. The color range indicates elasticities between 0 (white) to ≥0.03 (dark red).
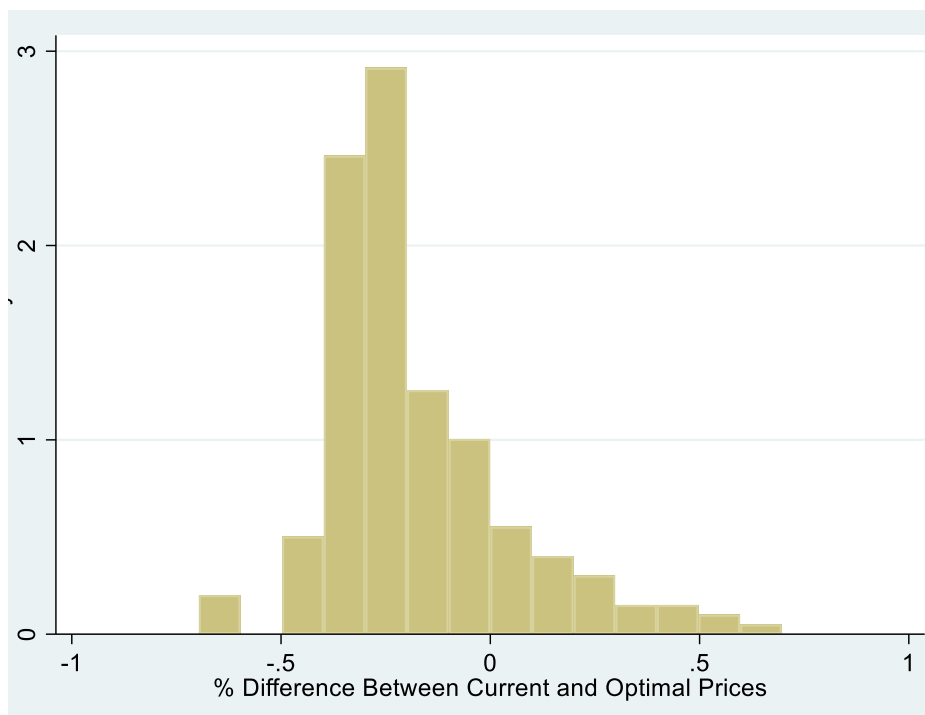
Figure 4: **Histogram of Percentage Price Difference between Current and Optimal Prices.**

company's prices, we solve for revenue-maximizing prices (details are available in section A of the appendix).

To keep the identity of the retailer anonymous, we do not report current and optimized prices directly, and instead frame our discussion in terms of percentage differences between optimal and current prices. Figure 4 shows a histogram of percentage deviations of optimal relative to current prices. On average, optimal prices derived from our preferred specification are lower by 23.1%. At the 25th percentile of the distribution of price differences, the optimal price is 31.2% lower than the current price of the specific product. At the 75th percentile, the optimal price is 6.5% lower. For 17% of retail products, the optimal price is higher than the one that is currently charged. We find that using the optimal price vector increases revenue by 7.1% relative to current prices.

We also compare the implied revenue from using our estimates with the full-information model. Using the optimal prices from the full-information random-coefficients model (Model C) increases revenues by 4.1% compared to the current price vector. Therefore, although revenue does improve relative to current prices, the full-information prices only achieve 58% of the gain in revenue relative to the optimal price vector implied by our model.

26

6 Conclusion

We propose a new demand model that is computationally light and allows for flexible substitution patterns. We leverage search data and use joint-search patterns as an additional source of information about the substitutability between products. The key modeling choice is that we do not "unpack" the consideration process, but instead treat consideration-set probabilities as objects to be estimated. We also allow the conditional distribution of preferences to depend on characteristics of the consideration set in a way that approximates a dependence that is common in search models. We believe our approach is particularly useful in online markets, where the number of products is large relative to the number of observed product characteristics, purchases are sparse, and demand may be driven by many factors, such as page layout, that are hard to record. We apply our approach to such a setting with almost 600 products, and show it has superior fit and significantly lower computational burden than a full-information random-coefficients model. We then use our demand estimates to solve for optimal prices, and find them to increase revenue by 7.1% relative to current prices.

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*.

CRAWFORD, G., R. GRIFFITH, AND A. IARIA (2018): "Preference Estimation with Unobserved Choice Set Heterogeneity using Sufficient Sets," .

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.

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.

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.

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.

LI, J., S. NETESSINE, AND S. KOULAYEV (2018): "Price to Compete ... with Many: How to Identify Price Competition in High-Dimensional Space," *Management Science*, 64(9), 4118–4136.

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.

RUIZ, F. J. R., S. ATHEY, AND D. M. BLEI (2017): "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements," .

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.

URSU, R. M. (2018): "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 37(4), 530–552.

A First-Order Conditions for Optimal Price Setting

As explained earlier, the retailer sells both retail and marketplace products, and in the case of the latter, it receives a percentage of the transaction price following a sale. Here, we show how to derive first-order conditions and solve for optimal prices for the retail products while taking into account that the firm also derives profits from marketplace products.

Assuming marketplace sellers do not respond to retail price changes, the website's problem is to

$$\max_{p^R} \ Profit = \sum_{j \in R}(p_j - mc_j)D_j(p^R, p^{MP}) + \sum_{k \in MP} \tau p_k D_k(p^R, p^{MP}), \tag{18}$$

where R denotes the set of retail products and MP denotes the set of marketplace products, whereas p^R and p^{MP} are the associated price vectors. Meanwhile, mc_j and τ denote, respectively, marginal costs and the commission rate received by the website. The first-order condition with respect to the price of retail product l is given by

$$D_l + \sum_{j \in R}(p_j - mc_j)\frac{\partial D_j}{\partial p_l} + \sum_{k \in MP} \tau p_k \frac{\partial D_k}{\partial p_l} = 0. \tag{19}$$

The system of first-order conditions can then be written in vector form as

$$\overrightarrow{D^R} + \Delta D^R \overrightarrow{p^R} - \Delta D^R \overrightarrow{mc^R} + \Delta D^{MP} \overrightarrow{p^{MP}} \tau = 0, \tag{20}$$

where $\overrightarrow{D^R}$ is a J^R-dimensional vector of retail product demands, and where price and marginal cost vectors $\overrightarrow{p^R}$, $\overrightarrow{p^{MP}}$, and $\overrightarrow{mc^R}$ are defined similarly. Meanwhile, ΔD^R denotes a $(J^R \times J^R)$ matrix of retail product derivatives, where the (a, b) element is equal to $\partial D_a / \partial p_b$. Finally, ΔD^{MP} denotes a $(J^{MP} \times J^R)$ matrix of derivatives of marketplace demands with respect to retail prices. Rearranging the system of equations (20), we can solve for optimal prices

$$\overrightarrow{p} = \overrightarrow{mc} - (\Delta D^R)^{-1}\left[\overrightarrow{D^R} + \Delta D^{MP}\overrightarrow{p^{MP}}\tau\right]. \tag{21}$$

Because the firm's objective is to maximize revenue, we set marginal costs equal to zero when solving for optimal prices.

B   Additional Tables

| | Full Info. Model A | | Full Info. Model B | | Full Info. Model C | |
| | Product FEs, Price & Deal Coeff. | | Full Info. Model A + RC on Price & Deal | | Full Info. Model B + RC on 7 Char. | |
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Price	-2.413	0.140	-2.413	0.140	-2.414	0.141
Deal Dummy	0.181	0.029	0.181	0.030	0.181	0.030
SD Price			0.005	0.033	0.005	0.038
SD Deal			0.019	0.084	0.111	0.082
SD Customer rating					0.130	0.055
SD Number of Reviews					0.037	0.057
SD Char 2					0.038	0.053
SD Char 3					0.064	0.139
SD Char 4					0.030	0.059
SD Char 5					0.003	0.077
SD Char 6					0.260	0.148
Product FEs	Yes		Yes		Yes	
# Observations	185,963		185,963		185,963	

Table A1: **Estimation Results for Full-Information Demand Models.**