An Empirical Model of Consideration through Search*

José Luis Moraga-González†

Zsolt Sándor‡

Matthijs R. Wildenbeest§

July 2024

Abstract

We propose a tractable method for the estimation of a consideration set model in which consideration sets are endogenously determined through search. We show that the widely-used alternative-specific consideration model is a special case in which consumers put zero weight on expected utility when making their search decisions. To deal with the dimensionality problem that may arise from a large number of consideration sets, we propose a novel, accurate, and computationally fast Monte Carlo estimator for the choice probabilities. We use several existing datasets to identify the extent to which search played a role when buyers formed their consideration sets.

Keywords: demand estimation, consumer search, consideration sets, differentiated products
JEL Classification: C51, D83, L13

*We thank Jason Abaluck, Abi Adams-Prassl, and Raluca Ursu for letting us use their data. An earlier version of this paper was circulated under the title "A Framework for the Estimation of Demand for Differentiated Products with Simultaneous Consumer Search." A simpler variant of the non-sequential search model proposed in this paper was first used in Moraga-González, Sándor, and Wildenbeest (2015). That paper underwent a major overhaul and the published version (Moraga-González, Sándor, and Wildenbeest, 2023) now features a sequential search model instead. For helpful comments on this paper we thank Javier Donna, Carl-Christian Groh, Elisabeth Honka, Tobias Klein, Stefan Seiler, Michael Sullivan, and Xuan Teng. The paper has also benefited from presentations at CREST (Paris), Tilburg University, Tinbergen Institute, University of Mannheim, EARIE 2022 (Vienna), MaCCI 2023 (Mannheim), 2023 Business Data Science-Tinbergen Institute Conference on Consumer Search and Markets (Rotterdam), and IIOC 2024 (Boston). Financial support from Marie Curie Excellence Grant MEXT-CT-2006-042471 and grant PN-II-ID-PCE-2012-4-0066 of the Romanian Ministry of National Education, CNCS-UEFISCDI is gratefully acknowledged.

†Vrije Universiteit Amsterdam and Télécom Paris, E-mail: j.l.moragagonzalez@vu.nl. Moraga-González is also affiliated with the Tinbergen Institute, CEPR, and the Public-Private Sector Research Center at IESE (Barcelona).

‡Sapientia Hungarian University of Transylvania, E-mail: sandorzsolt@uni.sapientia.ro. Sándor is also affiliated with the Babeş-Bolyai University.

§University of Arizona and CEPR, E-mail: wildenbeest@arizona.edu.

1 Introduction

Consumers often consider only a limited number of products before making a purchase decision. Understanding what determines which products are considered and how this affects purchase decisions has been the subject of a sizable literature in economics and marketing. Consideration set models generalize the standard (full consideration) discrete choice framework to allow individuals to consider subsets of products. These models typically use the probabilistic choice framework of Manski (1977), in which the probability of choosing an alternative can be separated into the probability of choosing an alternative conditional on a consideration set and the probability of considering the alternatives in that consideration set. Even though the full consideration model is often the default choice for modeling purchase decisions, several papers have pointed out that assuming full consideration in settings in which consumers have limited information could lead to biased demand estimates. For instance, Sovinsky Goeree (2008) studies a setting in which consideration sets are endogenously determined through advertising and finds that ignoring the effects of advertising and therefore incorrectly assuming consumers consider all products, results in demand estimates that are too elastic.

A widely used consideration set model is the *alternative specific consideration* (ASC) model.[1] This model, which dates back to Swait and Ben-Akiva (1987), features products that each have an independent probability of being considered. This probability may be related to a product's attributes (Abaluck and Adams-Prassl, 2021), product availability (Ben-Akiva and Boccara, 1995), or marketing mix variables such as product advertising (Sovinsky Goeree, 2008; Van Nierop, Bronnenberg, Paap, Wedel, and Franses, 2010; Dressler and Weiergraeber, 2023; Berger, 2024) and promotional activities (Ching, Erdem, and Keane, 2009; Ching, Erdem, and Keane, 2014). ASC models have gained popularity because the process of consideration set formation is modeled in a parsimonious way that allows the researcher to easily compute individual product consideration probabilities. Moreover, ASC models enable the researcher to be agnostic about which product characteristics affect product consideration and which affect utility (Abaluck and Adams-Prassl, 2021). However, the parsimonious structure of the consideration set probabilities in the ASC model also make it less suitable for handling consideration set probabilities that are determined through search. In search models consumers make a tradeoff between the benefits and costs of searching a set of

[1] We will use the name ASC model throughout this paper, which was introduced by Abaluck and Adams-Prassl (2021). Alternative names for this type of model are the *random constraint choice set* model (Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995), the *random consideration set* model (Lee, 2019), the *random choice set* model (Leslie, 2004; Sovinsky Goeree, 2008), and the *independent consideration* model (Chib and Shimizu, 2023).

products or firms. What makes the structure of these models more difficult to capture in an ASC-type framework is that the benefits of search depend not only on the features of the product in question but also on the characteristics of the products that are searched together with this product. This means a product's probability of being considered cannot be determined in isolation, as in the ASC model.

In Section 2 we introduce a novel consideration set model in which consideration sets are endogenously determined through optimal search and that gives the ASC model with logit consideration (a commonly-used parameterization of the ASC model used in Leslie, 2004; Sovinsky Goeree, 2008; Abaluck and Adams-Prassl, 2021; Berger, 2024, among others) as a limiting case.[2] We model search as non-sequential, which in our setting means consumers determine the optimal consideration set by making a tradeoff between the utility they expect to get from inspecting the products they include in their consideration sets and the costs of considering those products.[3] We assume search costs follow a Type I Extreme Value (TIEV) distribution and introduce a parameter that determines how much weight consumers put on expected utility versus consideration costs when determining their optimal consideration sets. We show that this parameter can be interpreted as capturing the variance of the unobserved part of search costs. More importantly, the ASC model with logit consideration can be obtained as a special case of our model in which consumers put zero weight on the expected benefits of search and hence allocate all the weight on the consideration costs when forming their consideration sets. Hence, estimating the weight parameter is important for pinning down the extent to which consideration sets are determined by the expected utility the products within these sets generate and allows us to distinguish between consideration sets that are determined through search and through other means.

We show that an individual's probability of purchasing a specific product can be computed in closed-form. This probability is equal to the sum of the products of two logit expressions: the multinomial logit probability that a group of products is considered times the multinomial logit probability that a product is chosen from the set of considered products. We show that our model belongs to the family of Generalized Extreme Value (GEV) models, which implies that it is consistent with utility maximization (McFadden, 1978).[4] More importantly, this implies that the

[2] For instance, in Sovinsky Goeree (2008) the probability that product k is considered is given by $\phi_k = \exp[\alpha_k]/(1+\exp[\alpha_k])$, where α_k is a consideration shifter for product k.

[3] For theoretical studies of models of non-sequential search for differentiated products see Section 7.6 of Anderson, de Palma, and Thisse (1992), Chade and Smith (2006), and Moraga-González, Sándor, and Wildenbeest (2021).

[4] For the case of no search cost shifters our model is similar to the two-stage choice set formation model in Swait (2001) (named the GenL model) in which preference parameters endogenously determine choice sets.

model has a full information discrete choice equivalent in which consumers pick the best product out of the set of all products assuming the utility shocks are no longer TIEV distributed but distributed according to the corresponding GEV joint distribution function. This is the analog of the *eventual purchase theorem* of Armstrong (2017) and Choi, Dai, and Kim (2018) for our setting.

Because the number of consideration sets increases exponentially in the number of products, calculation of the search and purchase probabilities may involve very large sums. For instance, with 25 products one has to sum over 33 million terms, which can be very time consuming to compute. We show in Section 3 that the search and purchase probabilities of our model have a similar structure to the ASC model, which allows us to treat the sums that create the dimensionality problem as expected values of a discrete random variable with probabilities as in the ASC model with logit consideration. We show how to exploit this to conveniently estimate the large sums by Monte Carlo, using an approach for randomly drawing consideration sets similar to Sovinsky Goeree (2008).[5]

The relatively straightforward computation of the search and choice probabilities allows us to derive the individual likelihood of a search and purchase decision and to estimate our model following the methods proposed by Goolsbee and Petrin (2004) and Train and Winston (2007). An advantage of using both search and purchase data for estimation is that it enables us to separately identify utility and search cost shifters, which allows us to be completely agnostic about which covariates enter utility, which enter consideration costs, and which enter both as well as which are endogenous. To estimate the model, we first construct the likelihood of our data on individuals' search and purchase behavior. Then, following Berry, Levinsohn, and Pakes (2004), we use the aggregate data and the contraction property of our system of market share equations to solve for the mean utilities of the products. These mean utilities substitute for the linear part of utility in the likelihood function, which implies that we maximize the likelihood function for the "non-linear" parameters only. This procedure is the maximum likelihood analog of the micro-moments method of Berry, Levinsohn, and Pakes (2004), which yields consistent estimates of the non-linear parameters of the model even in the presence of price endogeneity and endogeneity problems caused by other covariates. These parameters include the variance parameters of consumers' preferences and the mean and variance parameters of search costs. Moreover, the maximum likelihood approach yields efficient estimates of such parameters (see Grieco, Murry, Pinkse, and Sagl, 2023). In a final step

[5]Lee (2019) and Chib and Shimizu (2023) provide alternative approaches for calculating choice probabilities in ASC-type models.

we regress the mean utilities on prices and other exogenous characteristics using an instrument for prices to get estimates of the marginal effects of each of these characteristics, which makes controlling for price endogeneity relatively straightforward.

In Section 4 we investigate the performance of our estimator using Monte Carlo simulations. In particular, we assess the performance of the Monte Carlo estimator by comparing the parameter estimates obtained using the simulated market shares and individual search and purchase probabilities to those obtained using the actual expressions when the number of firms is not too large. We find that even though our Monte Carlo estimator is slower when the number of firms is five or lower, it is about twenty times faster when there are ten firms. Moreover, in all our experiments the parameter estimates based on the Monte Carlo estimator are almost identical to those based on the actual probabilities, so the computational gains do not come at the cost of accuracy.

In Section 5 we validate our approach using several existing data sets. In our first application, we estimate demand in the Dutch market for automobiles. In addition to the usual aggregate data, survey data allows us to see which brands consumers considered before making a purchase decision. Even though we find that distance from a consumer to a car dealer is important as a search cost shifter, consumers put disproportional weight on expected utility when making search decisions. In our second application we use the experimental data from Abaluck and Adams-Prassl (2021). Participants to the experiment faced consideration sets that were generated using a known (probability function, which means we can use this data to see if our approach can correctly identify this as an ASC model. Estimates of the weight parameter are not significantly different from zero, suggesting that our approach can also be used to estimate models in which consideration sets are not endogenously determined through search. Finally, in our third application we use Expedia search and click data that was originally used in Ursu (2018). Here we make a direct comparison between estimates of our search and consideration model and estimates according to an ASC model. As in the car application we find that consumers put most of the weight on expected utility when making search decisions. Moreover, we find that the search and consideration model outperforms an ASC model with logit consideration (which has the weight parameter normalized to zero) in terms of fit, even when all utility variables are part of the consideration probabilities as well.

Related literature

Our paper fits into the large theoretical and empirical literature on consumer search behavior. Our paper relates to a strand in the theoretical consumer search literature that focuses on search for differentiated products (Wolinsky, 1986; Anderson and Renault, 1999). Our theoretical search model is most closely related to the logit search model discussed in Anderson, de Palma, and Thisse (1992), but we allow for asymmetric multi-product firms and for consumer heterogeneity in both preferences and search costs, and we allow consumers to put different weight on expected utility versus search costs.[6] The search model in our paper is also related to the non-sequential search model in Moraga-González, Sándor, and Wildenbeest (2021), who, as in our model, allow for search cost heterogeneity and differentiated products, and provide conditions under which a price equilibrium in pure strategies exists.

The empirical literature has seen a number of recent contributions that estimate models of search for differentiated products (Kim, Albuquerque, and Bronnenberg, 2010; De los Santos, Hortaçsu, and Wildenbeest, 2012; Seiler, 2013; Dinerstein, Einav, Levin, and Sundaresan, 2018; Honka, 2014; Koulayev, 2014; Pires, 2016). Our paper is most closely related to Moraga-González, Sándor, and Wildenbeest (2023). An important difference with that paper is in the way search behavior is modeled: while in this paper we assume consumers search non-sequentially, Moraga-González, Sándor, and Wildenbeest (2023) assume consumers search sequentially. With sequential search, search decisions are based on realized search outcomes, which makes it less straightforward to obtain closed-form expressions for the buying probabilities. Using recent insights from Armstrong (2017) and Choi, Dai, and Kim (2018), Moraga-González, Sándor, and Wildenbeest (2023) show how to obtain closed-form expressions for the buying probabilities. In this paper we provide an alternative approach which relies on non-sequential search combined with the search cost shocks being choice-set specific. Both approaches lead to closed-form expressions for the buying probabilities when making specific assumptions about the search cost distribution. An advantage of the model we lay out in this paper is that it is easier to obtain expressions for the search probabilities, which makes it especially suitable for use with individual-specific search and choice data, as we illustrate in Section 5 for various empirical settings. Another advantage of non-sequential search is that it can easily handle settings in which firms (or retailers) sell overlapping sets of products, since search decisions are determined before any realizations of the match values are observed. In contrast, with sequential search consumers determine after each search whether to continue or not, so search decisions are conditional on observed match values. This means that if firms sell overlapping sets of products, searching one firm may give valuable information about searching another firm, which makes the optimal search rule more challenging than in the standard setting in which firms sell

[6] See Section 7.6 (pp. 246–248) of Anderson, de Palma, and Thisse (1992).

non-overlapping sets of products (see Weitzman, 1979).

A number of recent papers have built on earlier versions of the search model presented here. For instance, Lin and Wildenbeest (2020) develop a method to non-parametrically estimate search costs using a conditional logit version of our model, in which search costs are assumed to be consumer-specific but identical across firms. Murry and Zhou (2020) use individual-level transaction data for new products to quantify how geographical concentration among product sellers affects competition and search behavior. Donna, Pereira, Pires, and Trindade (2022) estimate the welfare effects of intermediation in the Portuguese outdoor advertising industry using a demand model that extends our search model to allow for nested logit preferences. Ershov (2018) develops a structural model of supply and demand to estimate the effects of search frictions in the mobile app market and uses our search model on the demand side. Pires (2018) studies the effect of search frictions on prices and profits in the laundry detergent market and follows our approach of using choice-set specific logit errors to smooth the choice set probabilities. Finally, De los Santos, Hortaçsu, and Wildenbeest (2012) estimate a related non-sequential search model using individual-specific data in which consumer search behavior and individual choice sets are observed—our approach uses aggregate data in addition, which allows the researcher to deal with price endogeneity. Moreover, we specifically deal with the dimensionality problem that arises when the number of possible choice sets is large, which is not necessary in their application because of the low number of choice sets.

Our paper is also closely related to the consideration set literature in economics and marketing.[7] Early contributions focused on search as a main determinant of consideration sets (Hauser and Wernerfelt, 1990; Roberts and Lattin, 1991), while later work focused on marketing mix variables such as promotion, display, advertising, and price as a main source of consideration set heterogeneity (Bronnenberg and Vanhonacker, 1996; Ching, Erdem, and Keane, 2009; Sovinsky Goeree, 2008; Van Nierop, Bronnenberg, Paap, Wedel, and Franses, 2010; Ching, Erdem, and Keane, 2014). More recent work has focused on the identification of unobserved consideration sets (Abaluck and Adams-Prassl, 2021; Barseghyan, Coughlin, Molinari, and Teitelbaum, 2021; Crawford, Griffith, and Iaria, 2021). Abaluck and Adams-Prassl (2021) study models in which the consideration set probabilities have similar functional forms as those in Sovinsky Goeree (2008) and show that cross-derivatives of buying probabilities with respect to product characteristics are asymmetric, which can be employed to separately identify how a product feature (such as price) affects preferences and consideration

[7] See Honka, Hortaçsu, and Wildenbeest (2019) for a recent survey of the search and consideration set literature.

without using data on consideration.[8]

2 The Model

In this section we introduce a consideration set model in which consideration sets are endogenously determined through search. We model search as non-sequential, which in our setting means consumers determine the optimal set of firms to consider by making a tradeoff between the expected utility from visiting and buying from a subset of firms and the costs of considering these firms. We introduce a parameter that determines how much weight consumers put on expected utility versus search costs when determining optimal consideration sets, which we show can be interpreted as capturing the variance of the unobserved part of search costs. Furthermore, we show that the ASC model with logit consideration is a special case of our model in which consumers put zero weight on the expected benefits of search and consideration costs include both benefits and costs.

2.1 Utility, demand, and search costs

Consider a market that consists of F different firms (indexed by $f = 1, 2, \ldots, F$), selling J different products (indexed by $j = 1, 2, \ldots, J$). We assume that product j is sold by a single firm f, but allow firms to sell multiple products. Specifically, let firm $f \in \mathcal{F}$ sell a subset of products $\mathcal{G}_f \subset \mathcal{J}$, where $\mathcal{J}$ denotes the set of products and $\mathcal{F}$ represents the set of firms.

We posit that, conditional on having inspected product j, consumer i's indirect utility of consuming product j is given by

$$u_{ij} = \alpha_i p_j + x_j' \beta_i + \xi_j + \varepsilon_{ij}, \tag{1}$$

where α_i is consumer i's price coefficient, p_j is the price of product j, x_j captures product j's attributes, β_i is a consumer-specific parameter that captures the marginal utility of each of these attributes, ξ_j captures characteristics not observed by the econometrician, and ε_{ij} is a consumer-product specific utility shock that is not observed by the econometrician either. We assume ε_{ij} is Type I Extreme Value (TIEV) distributed across consumers and products and captures whether product j is a good match for consumer i. Consumers have the option of not buying any product and opt for the outside option, which gives utility $u_{i0} = \varepsilon_{i0}$. Before inspecting a product j,

[8] Abaluck and Adams-Prassl (2021) show that this source of identification is also present in so-called default specific consideration (DSC) models (Ho, Hogan, and Scott Morton, 2017; Hortaçsu, Madanizadeh, and Puller, 2017), in which consumers either choose a default option without considering other options or make an informed choice among all available options.

consumers know the characteristics x_j and ξ_j but are not aware of the exact match value ε_{ij} and the exact price of the product p_j. The purpose of search is thus to figure out the realized values of the stochastic utility shocks ε_{ij} and the actual prices at which products sell.

We adopt a consideration set formulation for the probability of choosing product j, where consideration sets are determined through optimal non-sequential search. Let $\boldsymbol{S}$ be the set of all subsets of firms in $\mathcal{F}$, including the empty set, and let S be an element of $\boldsymbol{S}$. Letting $\boldsymbol{S}_f \subset \boldsymbol{S}$ denote the set of all consideration sets containing firm f, the choice probabilities are then given by

$$s_{ij} = \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S}, \tag{2}$$

where P_{iS} is the probability that the set of firms S is considered and $P_{ij|S}$ is the probability that product j is chosen conditional on considering the products in consideration set S.

We assume the consideration set probabilities P_{iS} are a function of both the expected utility of visiting a subset of firms and the cost of considering this subset. We specify consumer i's cost for visiting all the firms in the subset S, denoted c_{iS}, as:

$$c_{iS} = \sum_{f \in S} \kappa\left(t'_{if}\gamma_i\right) + \lambda_{iS}. \tag{3}$$

Here κ is a known function, t_{if} is a vector of cost shifters that are consumer and/or firm specific (such as employment status and distance to the firm), γ_i a vector of random coefficients, and λ_{iS} is a consumer-specific search cost shock for visiting a set of firms S that is not observed by the econometrician. It is the addition of this consumer consideration-set specific error term to the costs of searching subsets of firms that allows us to solve the search problem while maintaining tractability. We interpret this search cost shock as consideration-set specific variation in search costs that the observed search cost shifters are unable to pick up.

The consideration-set specific error term allows us to compute the probability that any given search-set is chosen. This idea is analogous to adding an error term to utility in discrete-choice models. If we further assume that $-\lambda_{iS}$ follows a TIEV distribution, then we can compute the probability with which any subset of sellers is chosen in closed form, no matter how large the number of available options is. We derive this probability in the next section.

2.2 Consideration and choice probabilities

Given consideration costs c_{iS}, the expected gain to consumer i from inspecting all the products sold by the sellers in a subset S is equal to:

$$w \cdot \mathbb{E}\left[\max_{j \in \mathcal{G}_f \cup \{0\},\ f \in S} \{u_{ij}\}\right] - (1-w)c_{iS},$$

where $\mathbb{E}$ denotes the expectation operator, taken in this case over the search characteristics ε_{ij}'s, and $w \in [0,1)$ is a parameter that reflects how much weight the consumer puts on expected utility versus consideration costs when deciding on which (subset of) sellers to consider. We now define m_{iS} as the non-stochastic part of the expected gain to consumer i from inspecting subset S, i.e.,

$$m_{iS} \equiv w \cdot \mathbb{E}\left[\max_{j \in \mathcal{G}_f \cup \{0\},\ f \in S} \{u_{ij}\}\right] - (1-w)\sum_{f \in S} \kappa\left(t'_{if}\gamma_i\right).$$

The expected maximum utility for set S is given by the logit inclusive value for all products in that set, which means we can write m_{iS} as

$$m_{iS} = w\left(\log\left(1 + \sum_{j \in \mathcal{G}_f,\ f \in S} \exp[\delta_{ij}]\right)\right) - (1-w)\sum_{f \in S} \kappa\left(t'_{if}\gamma_i\right), \tag{4}$$

where $\delta_{ij} \equiv \alpha_i p_j + x'_j \beta_i + \xi_j$.[9]

Consumer i will pick the subset of sellers to visit that maximizes the overall expected gain $m_{iS} - (1-w)\lambda_{iS}$, i.e., consumer i's optimal consideration set by S_i^* is:

$$S_i^* = \arg\max_{S \in \boldsymbol{S}} \left[m_{iS} - (1-w)\lambda_{iS}\right].$$

As discussed above, conditional on the λ_{iS}'s, the explicit characterization of the set S_i^* is extremely difficult. However, because $-\lambda_{iS}$ is assumed i.i.d. TIEV distributed, we obtain a closed-form expression for the consideration set probability P_{iS}:

$$P_{iS} = \frac{\exp\left[m_{iS}/(1-w)\right]}{\sum_{S' \in \boldsymbol{S}} \exp\left[m_{iS'}/(1-w)\right]}, \tag{5}$$

where the sum in the denominator is for all the possible consideration sets. Note that this sum may be over a large number of consideration sets; in Section 3.1 we show how to deal with this

[9] We have left out the Euler constant from the expected maximum utility since it does not affect choices.

dimensionality problem.

The expression in equation (5) is the multinomial logit probability that consumer i considers the products of the sellers contained in the set S. Once these products are inspected, the probability that she buys alternative j (sold by one of the visited sellers) is equal to the probability that product j provides the highest utility out of the products of the firms in S. Denoting this probability by $P_{ij|S}$, we have:

$$P_{ij|S} = \frac{\exp[\delta_{ij}]}{1 + \sum_{r \in S} \exp[\delta_{ir}]}, \tag{6}$$

where product r is a product sold by one of the firms in S.

Using equations (4) and (5), we can write P_{iS} (for $S \neq \varnothing$) as follows:

$$\begin{aligned} P_{iS} &= \frac{\exp\left[\frac{w}{1-w} \log\left(1 + \sum_{j \in \mathcal{G}_f,\ f \in S} \exp[\delta_{ij}]\right) - \sum_{f \in S} \kappa\left(t'_{if}\gamma_i\right)\right]}{1 + \sum_{S' \in \boldsymbol{S} \setminus \varnothing} \exp\left[\frac{w}{1-w} \log\left(1 + \sum_{j \in S'} \exp\left[\delta_{ij}\right]\right) - \sum_{f \in S'} \kappa\left(t'_{if}\gamma_i\right)\right]} \\ &= \frac{\left(1 + \sum_{j \in S} \exp\left[\delta_{ij}\right]\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS}\right]}{1 + \sum_{S' \in \boldsymbol{S} \setminus \varnothing} \left(1 + \sum_{j \in S'} \exp\left[\delta_{ij}\right]\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]}, \end{aligned} \tag{7}$$

where we use the notation $\bar{c}_{iS} \equiv \sum_{f \in S} \kappa\left(t'_{if}\gamma_i\right)$. Using equations (6) and (7), we can write individual choice probabilities $s_{ij} = \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S}$ as

$$\begin{aligned} s_{ij} &= \sum_{S \in \boldsymbol{S}_f} \frac{\left(1 + \sum_{j \in S} \exp\left[\delta_{ij}\right]\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS}\right]}{1 + \sum_{S' \in \boldsymbol{S} \setminus \varnothing} \left(1 + \sum_{j \in S'} \exp\left[\delta_{ij}\right]\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]} \times \frac{\exp[\delta_{ij}]}{1 + \sum_{r \in S} \exp[\delta_{ir}]} \\ &= \exp[\delta_{ij}] \frac{\sum_{S \in \boldsymbol{S}_f} \left(1 + \sum_{j \in S} \exp\left[\delta_{ij}\right]\right)^{\frac{w}{1-w} - 1} \exp\left[-\bar{c}_{iS}\right]}{1 + \sum_{S' \in \boldsymbol{S} \setminus \varnothing} \left(1 + \sum_{j \in S'} \exp\left[\delta_{ij}\right]\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]}. \end{aligned} \tag{8}$$

Given these individual choice probabilities, the aggregate probability that product j is chosen is equal to the integral:

$$s_j = \int s_{ij}\ f_\tau(\tau_i) d\tau_i, \tag{9}$$

where $f_\tau(\tau_i)$ is the joint density function of the random coefficients and the demographic characteristics of consumer i that enter the utility and search cost specifications.

The weight parameter w captures the relative importance of the benefits of search versus the cost of searching when making search decisions, where the distribution of the search benefits depends on the distribution of the utility shock ε. Since the the scale parameter of this distribution is normalized to one, the weight parameter w can also be interpreted as capturing the variance of the consideration-set specific search cost shock λ. In fact, denoting the scale parameter of the distribution of λ by σ_λ, it can be shown that $\sigma_\lambda = (1-w)/w$, which indicates that less weight being put on expected utility (w closer to zero) corresponds to larger variance of search costs.[10]

Finally, we note that the computation of the individual purchase probabilities using equation (8) as is becomes extremely tedious in situations where there is a large number of firms. The reason is twofold: the denominator of equation (8) requires us to sum over all 2^F choice sets and the numerator over all 2^{F-1} choice sets that contain firm f. Depending on the number of firms, these sums may have hundreds of thousands of summands. The existing literature that built on an earlier version of this paper (Moraga-González, Sándor, and Wildenbeest, 2015) avoided this dimensionality problem in various ways. We ourselves normalized the weight parameter to $w = \frac{1}{2}$, in which case the individual purchase probabilities can be simplified and the dimensionality problem disappears.[11] Pires (2018) followed the same approach. Murry and Zhou (2020) did not adopt this normalization but assumed that consumers search across clusters of sellers, rather than across sellers, and restricted the number of miles a buyer is willing to travel to visit a cluster to 40 miles.[12] Finally, Donna, Pereira, Pires, and Trindade (2022) did not adopt this normalization either but in their empirical application consumers search among only nine retailers. An important contribution of the current paper is to provide a novel Monte Carlo estimator of equation (8) that is accurate and fast even if the number of alternatives in the market is large (see Section 3.1). For example, in Section 4 we show that with ten firms the Monte Carlo estimation takes twenty times less than

[10] Allowing for a scale parameter σ_λ in equation (3) gives $c_{iS} = \sum_{f\in S} \kappa\left(t'_{if}\gamma_i\right) + \sigma_\lambda \lambda_{iS}$. For the case in which there is no weight parameter (i.e., $w = \frac{1}{2}$), this changes the individual choice probabilities in equation (8) to

$$s_{ij} = \exp[\delta_{ij}] \frac{\sum_{S\in \boldsymbol{S}_f} \left(1+\sum_{j\in S}\exp\left[\delta_{ij}\right]\right)^{\frac{1}{\sigma_\lambda}-1} \exp\left[-\overline{c}_{iS}\right]}{1+\sum_{S'\in \boldsymbol{S}\setminus\varnothing}\left(1+\sum_{j\in S'}\exp\left[\delta_{ij}\right]\right)^{\frac{1}{\sigma_\lambda}} \exp\left[-\overline{c}_{iS'}\right]}.$$

Note that even though γ in $\overline{c}_{iS}$ is now normalized by σ_λ, i.e., $\overline{c}_{iS} \equiv \sum_{f\in S}\kappa\left(t'_{if}\gamma_i/\sigma_\lambda\right)$, for $\gamma_i = \tilde{\gamma}_i\sigma_\lambda$ this version of the model is identical to our main model in case $\sigma_\lambda = (1-w)/w$.

[11] In such a special case, we can integrate out the choice-set probabilities and write out the buying probabilities in equation (8) as (see Appendix B):

$$s_{ij} = \frac{\exp\left[\delta_{ij} - \ln\left(1+\exp\left[\kappa\left(t'_{i\{f\}}\gamma_i\right)\right]\right)\right]}{1+\sum_{k=1}^{J}\exp\left[\delta_{ik} - \ln\left(1+\exp\left[\kappa\left(t'_{i\{g\}}\gamma_i\right)\right]\right)\right]}, \tag{10}$$

where $\kappa\left(t'_{i\{f\}}\gamma_i\right)$ contains the search cost of firm f only.

[12] Murry and Zhou (2020) put it this way: "*This restriction dramatically reduces the computational burden of computing consumers' optimal search sets. Otherwise, it is computationally infeasible to compute them by allowing consumers to optimally choose their search sets among 248 clusters.*"

computing the actual sum.

2.3 Alternative specific consideration model

Because we let consumers put arbitrary weight on the relative importance of expected utility and consideration costs, our setup includes the ASC model with logit consideration as a special case in which zero weight is put on expected utility when deciding which firms to visit.[13] As such our model can be used to provide a micro foundation for the ASC consideration set probabilities used by Sovinsky Goeree (2008), which have been further studied in the consideration set literature (see e.g. Abaluck and Adams-Prassl, 2021).

For $w = 0$, the expression m_{iS} in equation (4) becomes

$$m_{iS} = -\sum_{f \in S} \kappa\left(t'_{if}\gamma_i\right)$$

and therefore the choice set probability P_{iS} becomes

$$P_{iS} = \frac{\exp\left[-\sum_{f \in S} \kappa\left(t'_{if}\gamma_i\right)\right]}{\sum_{S' \in \boldsymbol{S}} \exp\left[-\sum_{f \in S'} \kappa\left(t'_{if}\gamma_i\right)\right]} \equiv \frac{\exp[-\bar{c}_{iS}]}{\sum_{S' \in \boldsymbol{S}} \exp[-\bar{c}_{iS'}]}.$$

Denoting $\bar{c}_{i\{f\}} \equiv \kappa\left(t'_{if}\gamma_i\right)$, since

$$\exp\left[-\bar{c}_{iS}\right] = \prod_{f \in S} \exp\left[-\bar{c}_{i\{f\}}\right] \tag{11}$$

and

$$\sum_{S \in \boldsymbol{S}} \exp\left[-\bar{c}_{iS}\right] = \sum_{S \in \boldsymbol{S}} \prod_{f \in S} \exp\left[-\bar{c}_{i\{f\}}\right] = \prod_{f \in F} \left(1 + \exp\left[-\bar{c}_{i\{f\}}\right]\right),$$

we obtain that

$$P_{iS} = \frac{\prod\limits_{f \in S} \exp\left[-\bar{c}_{i\{f\}}\right]}{\prod\limits_{f \in F} \left(1 + \exp\left[-\bar{c}_{i\{f\}}\right]\right)}.$$

Note that this can be written as

$$P_{iS} = \prod_{f \in S} \phi_{if} \prod_{f \notin S} \left(1 - \phi_{if}\right), \quad \text{where} \quad \phi_{if} = \frac{\exp\left[-\bar{c}_{i\{f\}}\right]}{1 + \exp\left[-\bar{c}_{i\{f\}}\right]} \tag{12}$$

[13] Note that utility components might still enter the model through consideration costs.

which has the same structure as the ASC consideration set probabilities in equation (3) of Sovinsky Goeree (2008).

2.4 Utility maximization in a GEV model of discrete choice

In this section we show that our non-sequential search model is equivalent to a standard discrete-choice model in which consumers choose from the J products to maximize their utility. This is the analog of the *eventual purchase theorem* of Armstrong (2017) and Choi, Dai, and Kim (2018) for our setting. We do so by appealing to the so-called generalized extreme value (GEV henceforth) family of discrete-choice models developed by McFadden (1978). The GEV generating function that leads to our non-sequential search model is

$$\begin{aligned} G\left(y_0, y_1, \ldots, y_J\right) &= \sum_{S \in \boldsymbol{S}} \rho_S \left(\sum_{f \in S} \sum_{r \in \mathcal{G}_f} y_r \right)^{w/(1-w)} \\ &= \sum_{S \in \boldsymbol{S}} \rho_S \left(y_0 + \sum_{f \in S \setminus \{0\}} \sum_{r \in \mathcal{G}_f} y_r \right)^{w/(1-w)} \qquad \text{for} \quad y_0, y_1, \ldots, y_J \geq 0, \end{aligned} \tag{13}$$

where $\rho_S = \exp\left(-\overline{c}_S\right)$ and $\overline{c}_S$ denotes a generic search cost corresponding to choice set S. Using the GEV Theorem from Swait (2001), which is slightly more general than that from McFadden (1978), we can verify that G defined in equation (13) satisfies the required conditions, provided that $0 < w \leq \frac{1}{2}$. Indeed, it is straightforward to verify that G is homogeneous of degree $w/(1-w)$ and that $\lim_{y_j \to \infty} G\left(y_0, y_1, \ldots, y_J\right) = \infty$ for any j. Further, using that for any $j_1, \ldots, j_k \in \{0, 1, \ldots, J\}$ that can be the outside alternative or belong to firms $f_1, \ldots, f_\ell$ we have

$$\frac{\partial^k G\left(y_0, y_1, \ldots, y_J\right)}{\partial y_{j_1} \ldots \partial y_{j_k}} = \frac{w}{1-w} \left(\frac{w}{1-w} - 1 \right) \ldots \left(\frac{w}{1-w} - k + 1 \right) \sum_{S \in \boldsymbol{S}_{f_1} \cap \ldots \cap \boldsymbol{S}_{f_\ell}} \rho_S \left(\sum_{f \in S} \sum_{r \in \mathcal{G}_f} y_r \right)^{\frac{w}{1-w} - k},$$

where recall that $\boldsymbol{S}_f$ denotes the set of all choice sets that contain firm f. One can easily see that these derivatives are nonnegative if k is odd and nonpositive if k is even, as required. The GEV generating function implies the following joint CDF of the $J+1$ error terms in the utility

$$
\begin{aligned}
F^{GEV}\left(\epsilon_0,\epsilon_1,\ldots,\epsilon_J\right) &= \exp\left[-G\left(\exp\left[-\epsilon_0\right],\exp\left[-\epsilon_1\right],\ldots,\exp\left[-\epsilon_J\right]\right)\right], \quad -\infty<\epsilon_j<\infty; \\
&= \exp\left[-\sum_{S\in\boldsymbol{S}}\left(\sum_{f\in S}\sum_{r\in\mathcal{G}_f}\exp\left[-\left(\epsilon_r+(1-w)\,\overline{c}_S/w\right)\right]\right)^{w/(1-w)}\right],
\end{aligned}
$$

where the choice-set specific search costs $\overline{c}_S$ act as shifters of the "mean" of the distribution.

Based on McFadden (1978), we can now claim that in our non-sequential search model, consumer i chooses product j that maximizes the utility $\overline{u}_{ij}=\delta_{ij}+\epsilon_{ij}$, where the vector of error terms $(\epsilon_{i0},\epsilon_{i1},\ldots,\epsilon_{iJ})$ has the CDF F^{GEV} with $\overline{c}_S$ replaced by $\overline{c}_{iS}$ (defined below equation (7)). The components of the error vector $(\epsilon_{i0},\epsilon_{i1},\ldots,\epsilon_{iJ})$, unlike the ε_{ij}'s from equation (1), can be regarded as usual error terms in the sense that they are observed by consumer i and they are not observed by the econometrician. The probability of choosing j can be derived from F^{GEV} using the following formula from McFadden (1978):

$$
P(j)=\frac{\exp\left[\delta_{ij}\right]}{\mu G\left(\exp\left[\delta_{i0}\right],\exp\left[\delta_{i1}\right],\ldots,\exp\left[\delta_{iJ}\right]\right)}\frac{\partial G\left(\exp\left[\delta_{i0}\right],\exp\left[\delta_{i1}\right],\ldots,\exp\left[\delta_{iJ}\right]\right)}{\partial y_j}, \tag{14}
$$

where in the definition of G from equation (13) we replace ρ_S by $\rho_{iS}=\exp\left(-\overline{c}_{iS}\right)$. It is straightforward to verify that this is the same as the choice probability derived above.

As in Armstrong (2017) and Choi, Dai, and Kim (2018), the *eventual purchase theorem* for our non-sequential search model implies that it can be reformulated as a discrete-choice problem without search frictions. This means that in order to distinguish between a search model and a full information model, we need search data.

3 Estimation

In this section we show how to estimate the model using individual-level data on consideration sets and purchases, as well as product characteristics and aggregate sales. We first discuss how to estimate the purchase probabilities in our model by Monte Carlo methods, which allows us to deal with the dimensionality problem that arises due to the number of consideration sets increasing exponentially in the number of firms. Next we discuss more generally how to estimate our model using maximum likelihood. We end this section with an informal discussion on identification of the main parameters in our model.

3.1 Monte Carlo estimator of the market shares

Market share expressions are needed for the BLP contraction mapping. In Appendix A we show that the expression for the individual purchase probabilities given in equation (8) can be written as follows:

$$s_{ij} = \frac{\exp\left[\delta_{ij}\right]\exp\left[-\bar{c}_{i\{f\}}\right]}{\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)} \frac{\sum_{S\in\boldsymbol{S}_{-f}}\left(1+\bar{\delta}_{i\{f\}}+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1}\frac{\exp[-\bar{c}_{iS}]}{\prod_{g\in\mathcal{F}\setminus\{f\}}\left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)}}{\sum_{S'\in\boldsymbol{S}}\left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}}\frac{\exp[-\bar{c}_{iS'}]}{\prod_{g\in\mathcal{F}}\left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)}}, \tag{15}$$

where, for notation simplicity, we write

$$\bar{\delta}_{i\{f\}} \equiv \sum_{j\in\mathcal{G}_f}\exp\left[\delta_{ij}\right], \text{ and } \bar{\delta}_{iS} \equiv \sum_{f\in S}\bar{\delta}_{i\{f\}}.$$

We now show how to estimate equation (15) by Monte Carlo. Both the numerator and the denominator of the second fraction in equation (15) contain potentially very large sums. Consider the denominator of such a fraction, which contains the largest sum:

$$D_i \equiv \sum_{S'\in\boldsymbol{S}}\left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}}\frac{\exp\left[-\bar{c}_{iS'}\right]}{\prod_{g\in\mathcal{F}}\left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)}$$

Notice now that, because $\exp\left[-\bar{c}_{iS}\right] = \prod_{f\in S}\exp\left[-\bar{c}_{i\{f\}}\right]$,

$$Q_{iS} \equiv \frac{\exp\left[-\bar{c}_{iS}\right]}{\prod_{g\in\mathcal{F}}\left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)} = \prod_{g\in S}\phi_{ig}\prod_{g\notin S}(1-\phi_{ig}), \tag{16}$$

where

$$\phi_{ig} = \frac{\exp\left[-\bar{c}_{i\{g\}}\right]}{1+\exp\left[-\bar{c}_{i\{g\}}\right]},$$

which is a factorization that allows sampling for Monte Carlo estimation of the sum (similar to Sovinsky Goeree, 2008). Further, notice that $\sum_{S'\in\boldsymbol{S}} Q_{iS'} = 1$, which implies that we can regard $(Q_{iS'})_{S'\in\boldsymbol{S}}$ as a probability mass function. As a result, we can interpret the denominator D_i as the expectation of a discrete random variable taking on values $\left((1+\bar{\delta}_{iS'})^{\frac{w}{1-w}}\right)_{S'\in\boldsymbol{S}}$ with probabilities $(Q_{iS'})_{S'\in\boldsymbol{S}}$.

This interpretation gives rise to our Monte Carlo estimator of D_i. Let $\{\mathbf{u}^r\}_{r=1}^R = \{(u_1^r, u_2^r, \ldots, u_F^r)\}_{r=1}^R$ be a sample of size R of uniform $[0,1]$ vectors of dimension F (the number of firms), where each u_i^r is a random draw from $[0,1]$. For a given vector $\mathbf{u}^r$, let $(\mathbf{1}(u_1^r \leq \phi_{i1}), \mathbf{1}(u_2^r \leq \phi_{i2}), \ldots, \mathbf{1}(u_F^r \leq \phi_{iF}))$ be the resulting vector of ones and zeros indicating whether a firm is included in the search set S (one) or not (zero). Letting

$$\bar{\delta}_i(\mathbf{u}^r) \equiv \sum_{g=1}^{F} \mathbf{1}\left(u_g^r \leq \phi_{ig}\right) \exp\left[\delta_{ig}\right],$$

we can then rewrite D_i as:

$$D_i = \sum_{S' \in \boldsymbol{S}} \left(1 + \bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} Q_{iS'} = \int_{[0,1]^F} \left(1 + \bar{\delta}_i(\mathbf{u})\right)^{\frac{w}{1-w}} d\mathbf{u}. \tag{17}$$

Therefore, the Monte Carlo estimator of D_i is

$$\widehat{D}_i = \frac{1}{R} \sum_{r=1}^{R} \left(1 + \bar{\delta}_i(\mathbf{u}^r)\right)^{\frac{w}{1-w}}.$$

It is well-known that the estimator $\widehat{D}_i$ is not continuous in the search costs parameters γ_i that enter the ϕ_{ig}'s because the indicators $\mathbf{1}\left(u_g^r \leq \phi_{ig}\right)$ may jump from 0 to 1 or from 1 to 0 as γ_i changes slightly. This is problematic during the estimation of the model. Hence, instead of using the indicators $\mathbf{1}(u_g \leq \phi_{ig})$, we use a smoothed version of them given by $\Phi\left(\frac{\phi_{ig} - u_g}{h}\right)$, where Φ is the standard normal CDF and h is small (e.g., $h = 0.001$ or smaller).[14] Letting

$$\widetilde{\delta}_i(\mathbf{u}^r) = \sum_{g \in \mathcal{F}} \Phi\left(\frac{\phi_{ig} - u_g^r}{h}\right) \exp\left[\delta_{ig}\right],$$

the smooth Monte Carlo estimator of D_i that we use in estimation is

$$\widetilde{D}_i = \frac{1}{R} \sum_{r=1}^{R} \left(1 + \widetilde{\delta}_i(\mathbf{u}^r)\right)^{\frac{w}{1-w}}. \tag{18}$$

The numerator of the second fraction in equation (15) is

$$N_{if} \equiv \sum_{S \in \boldsymbol{S}_{-f}} \left(1 + \bar{\delta}_{i\{f\}} + \bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1} \frac{\exp\left[-\bar{c}_{iS}\right]}{\prod_{g \in \mathcal{F} \setminus \{f\}} \left(1 + \exp\left[-\bar{c}_{i\{g\}}\right]\right)}.$$

[14] Here Φ plays the role of a kernel function and h the role of a smoothing parameter similar to nonparametric kernel estimation. In Section S2 of the Supplementary Material we provide Monte Carlo evidence that justifies our choice of the smoothing parameter value.

In Appendix C we show that this can be approximated by the smooth-in-parameters integral

$$\widetilde{N}_{if}=\int_{[0,1]^F}\left(1+\bar{\delta}_{i\{f\}}+\widetilde{\delta}_{if}\left(\mathbf{u}_{-f}\right)\right)^{\frac{w}{1-w}-1}d\mathbf{u}. \tag{19}$$

By collecting the approximations provided in equations (18) and (19) we obtain the following estimator of equation (15):

$$\widetilde{s}_{ij}=\frac{\exp\left[\delta_{ij}\right]\exp\left[-\bar{c}_{i\{f\}}\right]}{\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)}\frac{\widetilde{N}_{if}}{\widetilde{D}_i}. \tag{20}$$

To derive an estimator of the aggregate market share of product j in equation (9) we can now use $\widetilde{s}_{ij}$ instead of s_{ij}. We need to integrate out the individuals' unobserved heterogeneity including the random coefficients and demographic characteristics. Hence we obtain the estimator of s_j

$$\begin{aligned}\widehat{s}_j &= \int\widetilde{s}_{ij}f_\tau\left(\tau_i\right)d\tau_i=\int\frac{\exp\left[\delta_{ij}\right]\exp\left[-\bar{c}_{i\{f\}}\right]}{\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)}\frac{\widetilde{N}_{if}}{\widetilde{D}_i}f_\tau\left(\tau_i\right)d\tau_i\\ &= \int\int_{[0,1]^F}\frac{\exp\left[\delta_{ij}\right]\exp\left[-\bar{c}_{i\{f\}}\right]}{\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)}\frac{\left(1+\bar{\delta}_{i\{f\}}+\widetilde{\delta}_{if}\left(\mathbf{u}_{-f}\right)\right)^{\frac{w}{1-w}-1}}{\widetilde{D}_i}f_\tau\left(\tau_i\right)d\mathbf{u}d\tau_i.\end{aligned} \tag{21}$$

This integral can be estimated by Monte Carlo by drawing simulatenously from τ_i and $\mathbf{u}$. Denote the sample by $\left(\tau_i,\mathbf{u}_i\right)_i$, $i=1,\ldots,ns$; we estimate s_j by

$$\widetilde{s}_j=\frac{1}{ns}\sum_{i=1}^{ns}\frac{\exp\left[\delta_{ij}\right]\exp\left[-\bar{c}_{i\{f\}}\right]}{\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)}\frac{\left(1+\bar{\delta}_{i\{f\}}+\widetilde{\delta}_{if}\left(\mathbf{u}_{-fi}\right)\right)^{\frac{w}{1-w}-1}}{\widetilde{D}_i}.$$

We note that by treating the market share integral as a joint integral in equation (21) we can reduce computing time considerably because in this way we only need to deal with two Monte Carlo sums instead of three (i.e. $\widetilde{D}_i$ and $\widetilde{s}_j$ instead of $\widetilde{D}_i$, $\widetilde{N}_{if}$ and $\widetilde{s}_j$).

3.2 Estimation approach

Following Goolsbee and Petrin (2004) and Train and Winston (2007), we first efficiently estimate the non-linear demand parameters (variance of preferences and mean and variance of search costs) by maximum likelihood using the individual-level data on consideration (firms an individual chooses to visit) and purchase (final choice an individual makes). In this step, we do not estimate the product fixed effects δ_j directly; instead, we follow Berry, Levinsohn, and Pakes (2004) and exploit the data on aggregate market shares to compute them by using the contraction mapping. In Section S1 of the Supplementary Material we show that our system of market share equations is a contraction. In a second step, we estimate the effects of price and product characteristics on (mean) utility using instruments for price as in Berry (1994). This two-step procedure is the maximum likelihood analog to the two-step GMM procedure of Berry, Levinsohn, and Pakes (2004).

In order to estimate the nonlinear demand parameters, we compute the likelihood of the consumers' observed search and purchase behavior. This likelihood is based on the joint probability that consumer i ($i = 1, \ldots, N$) searches the set of sellers S and chooses product $j \in \mathcal{G}_f$ with $f \in S$:

$$\Pr\left(i \text{ searches } S \text{ and } i \text{ chooses } j\right) = P_{iS} P_{ij|S}.$$

The probabilities P_{iS} and $P_{ij|S}$ depend on the demand and search cost parameters as well as observed individual characteristics and random coefficients. The probability that consumer i, whose individual characteristics are observed, searches S and buys j is

$$s_{iSj}\left(\theta\right) = \int P_{iS} P_{ij|S} f_v\left(v_i\right) dv_i, \tag{22}$$

where f_v is the density function of the random coefficients only. By equations (A26) and (A27) we have

$$\begin{aligned} s_{iSj}\left(\theta\right) &= \int \frac{\left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]} \frac{\exp[\delta_{ij}]}{1+\bar{\delta}_{iS}} f_v\left(v_i\right) dv_i \\ &= \int \frac{\exp[\delta_{ij}] \left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1} \exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]} f_v\left(v_i\right) dv_i \\ &= \int \frac{\exp\left[-\bar{c}_{iS}\right]}{\prod\limits_{f \in \mathcal{F}} \left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)} \frac{\exp[\delta_{ij}] \left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1}}{\frac{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]}{\prod_{f \in \mathcal{F}} \left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)}} f_v\left(v_i\right) dv_i, \end{aligned}$$

which, by equation (16), can be written as:

$$s_{iSj}\left(\theta\right) = \int Q_{iS} \frac{\exp[\delta_{ij}] \left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1}}{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} Q_{iS'}} f_v\left(v_i\right) dv_i.$$

This search and purchase probability can be estimated by Monte Carlo in the same way we estimated the individual purchase probabilities and market shares in Section 3.1. The Monte Carlo estimator is:

$$\widetilde{s}_{iSj}(\theta) = \sum_{q=1}^{ns} Q_{iS} \frac{\exp[\delta_{ij}(v_q)] \left(1 + \bar{\delta}_{iS}(v_q)\right)^{\frac{w}{1-w}-1}}{\widetilde{D}_i(v_q)},$$

where $\widetilde{D}_i$ is the smooth estimator of D_i given in equation (18) and the argument v_q reflects explicit dependence of the underlying expressions on the draws from the random coefficients.

Let y_{iSj} be 1 if consumer i searches S and buys j and 0 otherwise. Then the log-likelihood is

$$L_N(\theta) = \sum_{i=1}^{N} \sum_{j,S} y_{iSj} \log \widetilde{s}_{iSj}(\theta),$$

where the summation $\sum_{j,S}$ is over all possible j, S. Because in practice this sum only contains the terms corresponding to $y_{iSj} = 1$, the log-likelihood can also be written as

$$L_N(\theta) = \sum_{i=1}^{N} \log \widetilde{s}_{iSj}(\theta), \tag{23}$$

with S and j being i's search set and chosen product, respectively. Obviously, S is allowed to be the empty set and j is allowed to be the outside alternative.

As mentioned above, we do not estimate the product fixed effects δ_j that enter the likelihood directly. Instead, in every iteration of the maximum likelihood procedure we compute the ξ_j that enter δ_{ij} from the aggregate data as a function of the parameters. Following BLP, the predicted market share $s_j(\theta)$ of product j should match observed market shares s_j^o, or

$$s_j(\delta(\theta), \theta) - s_j^o = 0. \tag{24}$$

In Section S1 of the Supplementary Material we show that when $w \leq \frac{1}{2}$, this system of equations corresponds to a contraction mapping in $\delta(\theta)$, so it has exactly one solution. Hence, we can compute the components of the vector $\xi = (\xi_1, \ldots, \xi_J)$ of unobserved characteristics as

$$\xi_j = \delta_j(\theta) - \alpha p_j - x_j \beta, \tag{25}$$

where $\delta_j(\theta)$ is component j of the unique fixed point $\delta(\theta)$ of the contraction mapping. It is important to note that when we substitute ξ_j into δ_{ij} then αp_j and $x_j' \beta$ cancel. Consequently, the log-likelihood in equation (23) will depend on all the demand parameters except for α and β, and therefore, by maximizing $L_N(\theta) \equiv L_N(\theta_2)$ we can estimate all the demand parameters except for α and β by maximum likelihood. The linear utility parameters can be estimated in a second step using either OLS or, when dealing with price endogeneity, two-stage least squares.

3.3 Identification

We provide an informal discussion of the identification of the model parameters. In doing so we assume that, similar to estimation, individual-level search and purchase data are available. It is important to note that in the recent literature several papers provide frameworks for identification of demand without observing consideration sets (e.g., Abaluck and Adams-Prassl, 2021; Barseghyan, Coughlin, Molinari, and Teitelbaum, 2021; Lu, 2022). As we have pointed out in several places, in the most general version of our model, point-identification of the parameters is not possible in the absence of search data.

There are two main issues that need to be discussed regarding identification. The first is the potential endogeneity of price and search cost variables (e.g., distance from a consumer's home to the seller) while the second is the presence of common covariates in utility and search cost (e.g., the constants, which are important for both searching and buying decisions). Both of these issues can be tackled in the case when search and purchase data are available.

Consider first endogeneity. Recall from BLP that price endogeneity arises due to the fact that usually there are some unobserved characteristics that affect both utility and marginal cost and, therefore, affect also price, causing omitted variable bias. Endogeneity of a search cost variable like distance can occur in a similar way: products can have some characteristics (unobserved to the researcher) that affect both the utility that consumers derive from them and the location where they are sold. A crucial point of identification is to interpret the mean utility δ_j of product j as a product-specific fixed effect parameter. These fixed effect parameters incorporate the unobserved product characteristics responsible for endogeneity, so endogeneity is eliminated (Berry, Levinsohn, and Pakes, 2004). These fixed effect parameters can be identified together with the variance parameters from the utility and search cost based on individual-level variation in choices and search. Finally, the mean utility parameters included in $\delta_j = \alpha p_j + x_j' \beta + \xi_j$ can be identified as in Berry (1994) or BLP based on the usual conditional moment restriction that the unobserved product characteristics are mean independent of the observed characteristics and excluded (from utility) marginal cost shifters.

Consider the issue of common covariates next. The most unfavorable situation for identification is when a variable that affects both the utility and search cost enters the purchase probabilities as a sum of identical functional forms. To illustrate, one such example arises when $\kappa\left(t_{if}' \gamma_i\right) =$

$\ln\left(\exp\left(t'_{if}\gamma_i\right)-1\right)$ and $w=\frac{1}{2}$, in which case common covariates appear as a sum of linear terms. Using equation (A30), this specification leads to the consumer-specific buying probability

$$s_{ij}=\frac{\exp\left[\delta_{ij}-t'_{if}\gamma_i\right]}{1+\sum_{k=1}^{J}\exp\left[\delta_{ik}-t'_{ig}\gamma_i\right]},$$

which clearly shows that if both δ_{ij} and $t'_{if}\gamma_i$ contain a constant, these cannot be identified separately based on purchase or market share data, and likewise, if there are common covariates in the utility and the search cost, the coefficients of these cannot be identified separately either.

Now suppose we also observe which firms have been searched. Using the same functional form for κ as above as well as the normalization $w=\frac{1}{2}$, the probability that consumer i searches S and chooses j simplifies to

$$P_{iS}P_{ij|S}=s_{ij}\cdot\frac{\prod_{g\notin S}\left(\exp[t'_{ig}\gamma_i]-1\right)}{\prod_{g\in\mathcal{F}\setminus f}\exp[t'_{ig}\gamma_i]}.$$

Since the fraction in the second half of this equation only contains the search cost shifters, it is straightforward to see that common covariates should now be separately identified.

4 Monte Carlo Experiments

In this section we use Monte Carlo experiments to study the performance of the Monte Carlo estimator we use to deal with the dimensionality problem. This estimator allows us to deal with the dimensionality problem that arises because one has to sum over all possible choice sets to get these probabilities. However, with a small number of firms it is feasible to proceed without using the estimator and use the actual expressions for the market shares and individual search and purchase probabilities, allowing us to compare its performance for different number of firms.

We use the following setup for the Monte Carlo experiments. We simulate data for 25 different markets, where each market has up to 10 different firms, each selling one product. We allow for a utility constant as well as a product attribute that is randomly drawn from a normal distribution with mean 2 and standard deviation 0.5, with parameter values as given by the first column in Table 1. The unobserved characteristic ξ is drawn from a normal distribution with mean zero and standard deviation 0.1. We specify the search cost to be linear (that is, we take κ to be the identity function). We allow for a consumer-firm specific search cost shifter that is randomly drawn from a

Table 1: Results Monte Carlo Experiments

		(A)		(B)		(C)	
	true	3 firms		5 firms		10 firms	
Variable	coeff.	MC est	actual	MC est	actual	MC est	actual
Preference parameters							
Constant	-1.000	-0.974	-0.974	-1.015	-1.016	-0.987	-0.988
		(0.070)	(0.070)	(0.050)	(0.050)	(0.031)	(0.031)
Attribute 1	2.000	1.995	1.996	2.000	2.002	1.999	2.001
		(0.052)	(0.051)	(0.038)	(0.038)	(0.025)	(0.025)
Price	-2.000	-1.997	-1.998	-1.999	-1.999	-1.999	-2.000
		(0.027)	(0.026)	(0.020)	(0.020)	(0.013)	(0.013)
Search cost parameters							
Constant	1.500	1.512	1.510	1.479	1.477	1.524	1.522
		(0.075)	(0.075)	(0.059)	(0.059)	(0.043)	(0.042)
Shifter	1.000	1.025	1.025	1.014	1.015	0.993	0.993
		(0.080)	(0.080)	(0.058)	(0.058)	(0.038)	(0.038)
Weight parameter	0.500	0.498	0.500	0.498	0.500	0.499	0.501
		(0.058)	(0.057)	(0.043)	(0.042)	(0.030)	(0.030)
Estimation time (minutes)		1:50	0:26	2:41	2:23	4:47	94:55

Reported are the means and standard deviations (in parentheses) across 100 replications. All specifications use a combination of aggregate data and individual search and purchase data. Data are generated for 25 markets. The total number of individual-specific observations used for estimation is 2,500. The number of quasi-random draws for the Monte Carlo estimator is 529 and smoothing parameter value 0.001. Estimation time is the average time (in minutes) it takes for one replication to converge, and is obtained using MATLAB 2023b on a Apple Mac Studio M1 Max 64 GB.

lognormal distribution with variance of the variables' logarithm set to one and firm-specific means between 2 and 3.[15] The corresponding parameter values are shown in the first column of Table 1. For each replication we simulate prices, attributes, purchases, searches, and market shares. Prices are obtained by simulating equilibrium prices using the supply side model discussed in Appendix D, assuming consumers have to search for both price realizations and the matching term. As described in Section 3.1, we use a Monte Carlo estimator to estimate the buying probability according to equation (20), using 529 quasi-random draws and the smoothing parameter set to 0.001.

Table 1 reports for each specification the mean and standard deviation of the parameters of the model across 100 replications. As shown in column (A) of Table 1, when there are only 3 firms in the market it takes on average 26 seconds for a single replication to converge, which is more than three times faster than the Monte Carlo estimator. Estimation time is about the same when there are 5 firms, but when there are 10 firms the Monte Carlo estimator is about 20 times faster than when using the actual probabilities. This is in line with our expectations since for the Monte Carlo estimator the computing time is expected to increase linearly in the number of firms while for the estimator that uses the actual probabilities the computing time increases exponentially. Moreover,

[15]Specifically, the firm-specific means of the logarithm depend on the number of firms and are equally spread between 2 and 3, i.e., with three firms the means are 2, 2.5, and 3.

in all three cases the estimates when using the Monte Carlo estimator are almost identical to when using the actual probabilities, so the computational gains do not come at the cost of accuracy.

5 Validation

In this section we study several applications to illustrate the performance of our estimator with actual data.

5.1 Search for new cars

To illustrate the performance of our estimator in an empirical setting, in our first application we use our approach to estimate search costs in the Dutch market for new cars. The data we use for this is similar to the data used in Moraga-González, Sándor, and Wildenbeest (2023) and consists of new car sales, prices, and car characteristics for the period 2003-2008, supplemented with individual-specific choice data from a survey. The survey contains information for 2,530 respondents on the make and model of the most recently purchased car. The survey also contains information on the brands of the dealerships visited in the period before purchasing the car, which we use as a proxy for search visits. We only include respondents whose most recent car purchase happened during the period for which we have aggregate data (2003-2008), leaving us with 1,250 respondents for which we observe search and purchase choices.[16]

Estimation results are shown in Table 2. As a benchmark case, specification (A) gives estimates for the preference parameters when assuming full information and using only the aggregate data. We estimate this specification using two-stage least squares, using the same instruments as in Moraga-González, Sándor, and Wildenbeest (2023) to deal with price endogeneity (labor costs, steel prices interacted with weight, and annual production). The estimates are identical to those in column (A) of Table 3 of Moraga-González, Sándor, and Wildenbeest (2023), and show that all parameters have the expected sign and are precisely estimated. Specification (B) gives parameter estimates for the non-sequential search model when using only aggregate data for estimation. We include distance from consumers to the dealerships as a search cost shifter, but because a search cost constant is not separately identified from the utility constant when search data is not used for estimation, we do not include a search cost constant. The estimated distance parameter is precisely estimated and shows that search costs are higher for dealerships located further away.

[16] For a more detailed description of the data we refer to Section 4.1 of Moraga-González, Sándor, and Wildenbeest (2023).

In comparison to the full information results, all preference parameters decrease in magnitude, including the price coefficient. The latter implies that consumer demand is estimated to be less price sensitive in the search model than in the full information model.

Table 2: Estimation results cars data

	Full information		Non-sequential search					
	Aggregate data (A)		Aggregate data (B)		Aggregate and micro data (C)		Aggregate and micro data (D)	
PREFERENCE PARAMETERS								
constant	-20.728	(1.978)	-17.450	(1.787)	-16.250	(1.735)	-13.947	(1.528)
HP/weight	4.706	(1.294)	3.896	(1.143)	4.297	(0.994)	3.321	(0.874)
non-European	-1.237	(0.187)	-1.069	(0.165)	-1.139	(0.158)	-0.671	(0.139)
cruise control	0.508	(0.142)	0.263	(0.130)	0.348	(0.129)	0.146	(0.114)
fuel efficiency	2.627	(0.354)	2.511	(0.304)	2.538	(0.359)	2.130	(0.316)
size	13.453	(2.362)	10.608	(2.104)	11.770	(2.101)	8.835	(1.849)
luxury brand	1.923	(0.403)	1.758	(0.346)	1.863	(0.350)	1.802	(0.308)
price	-0.168	(0.035)	-0.131	(0.031)	-0.146	(0.029)	-0.112	(0.025)
SEARCH COST PARAMETERS								
constant	—		—		2.841	(0.027)	4.003	(0.070)
distance	—		0.192	(0.042)	0.085	(0.004)	0.054	(0.004)
weight parameter	—		0.500		0.500		0.962	(0.002)
Euro per mile	—		1048.92		556.70		204.72	
Log-likelihood					-10,465.86		-10,178.79	

Notes: The number of aggregate observations is 1,382. The number of individual-specific observations is 1,250. Standard errors are in parenthesis. The number of simulated consumers used for the estimation of specifications (B)-(D) is 300. The number of (Owen-randomized) quasi-random draws for the Monte Carlo estimator used in specifications (C) and (D) is 1,681, with bandwidth $1 \cdot 10^{-4}$. The linear part of utility includes car segment fixed effects. Instruments for price include unit labor costs (normalized by PPI), steel prices (normalized by PPI) interacted with weight, and annual production. We include average distance to dealerships as an instrument for specification (B). The Cragg-Donald Wald F test of excluded instruments for specification (A) is 15.68.

The remainder of the columns in Table 2 give estimates for various specifications of the non-sequential search model that are estimated using both aggregate and micro data. For these specifications we use the maximum likelihood estimation procedure outlined in Section 3, using the aggregate data to obtain estimates of mean utility δ_j and the individual search and purchase data to obtain estimates of the search cost parameters. Specification (C) normalizes the weight parameter w to 0.5, whereas in specification (D) we estimate the weight. The estimated ratio is precisely estimated and with 0.962 relatively large, which means consumers put relatively little weight on search costs when deciding which subset of dealers to visit and care more about differences in expected utility. Moreover, the estimated effect of distance on search costs decreases when estimating the weight parameter. This also affects the implied search costs of distance in dollars. Following Murry and Zhou (2020) we calculate the implied distance-related search costs by first calculating the sales change when locating all dealerships of a brand one kilometer further away from consumers, and then calculating how much the prices of all car models need to be lowered to compensate this sales loss. These estimates correspond to an implied search cost of €556.70 per kilometer for specification (C) and €204.72 for specification (D).

5.2 Experimental data

In this application we use the experimental data that Abaluck and Adams-Prassl (2021) utilize to validate their approach to estimating the ASC model.[17]. Since the consideration sets faced by participants in the experiment are generated using a known formula that is similar to the logit consideration set probabilities in Sovinsky Goeree (2008), the true consideration set probabilities are observed, so it can be directly tested whether their identification approach based on asymmetric demand responses works in a practical setting. Given that the ASC model with logit consideration is a special case of our model, this setting can also be used for validation of our estimation approach.

The experiment was carried out using 149 students from Yale University, who each received a budget of \$25. In each round of the experiment, participants had to make a choice from a randomly generated choice set that contained up to 10 items that were sold at the Yale Bookstore. After 50 rounds, one randomly chosen product from the set of chosen items was given to each of the participants as well as \$25 minus the price of the item in cash. The probability of an item j being part of a participant i's choice set in round r is given by

$$\phi_{ijr} = \frac{\exp\left[\gamma_j + p_{ijr}\gamma_p\right]}{1 + \exp\left[\gamma_j + p_{ijr}\gamma_p\right]},$$

where p_{ijr} is a product's randomly generated price, γ_p is a price coefficient, and γ_j is a product fixed effect. The true values of the consideration set parameters are given in the consideration column of specification (1) in Table 3. For further details regarding the experimental data we refer to Section IV of Abaluck and Adams-Prassl (2021).

Column (1) of Table 3 gives estimation results when we estimate the model conditioning on the actual consideration sets the respondents observed. Because we do not know the true preference parameters of the respondents, estimation of this model offers a benchmark we can to compare the preference parameter estimates of the full model. Estimates for the full model are shown in Column (2) of the table. We estimate the model using a maximum likelihood approach, where

[17]See https://doi.org/10.7910/DVN/ZVYOCQ for the replication data.

Table 3: Estimation Results Experimental Data

	(1) Conditional on consideration		(2) Search and consideration model	
	utility	consideration	utility	consideration
Price (dollars)	-0.173	0.150	-0.173	0.148
	(0.004)		(0.003)	(0.003)
Product 1	0.368	-2.500	0.375	-2.496
	(0.069)		(0.067)	(0.041)
Product 2	-0.497	-2.500	-0.499	-2.482
	(0.080)		(0.079)	(0.042)
Product 3	0.093	-2.500	0.092	-2.509
	(0.073)		(0.072)	(0.041)
Product 4	0.088	-2.500	0.085	-2.527
	(0.073)		(0.073)	(0.042)
Product 5	0.306	-2.500	0.309	-2.512
	(0.070)		(0.069)	(0.041)
Product 6	-0.581	0.000	-0.578	-0.009
	(0.045)		(0.045)	(0.039)
Product 7	-1.075	0.000	-1.075	-0.007
	(0.051)		(0.051)	(0.041)
Product 8	-0.909	0.000	-0.908	-0.002
	(0.048)		(0.048)	(0.040)
Product 9	-0.405	0.000	-0.405	-0.034
	(0.044)		(0.044)	(0.038)
Weight parameter			0.024	
			(0.046)	

Notes: This table reports estimation results using the experimental data in Abaluck and Adams-Prassl (2021). The estimates in column (1) correspond to those in column (4) of Table I in Abaluck and Adams-Prassl (2021) and are estimated using a conditional logit model that conditions on the actual consideration sets faced by consumers (with probabilities according to the ASC model). The estimates in column (2) are obtained using the search and consideration model. The consideration set probabilities include a constant. Standard errors are in parentheses.

the log-likelihood function is given by equation (23). Notice that in this application that we do not have market share data, which means we estimate the product fixed effects δ_j directly as part of the maximum likelihood procedure. The estimates in column (2) show that our estimation approach is able to recover the parameters of the consideration set probabilities—all consideration set parameters are close to the true ones. Moreover, the estimated preference parameters are similar to the ones in column (1) of the Table. The estimated weight parameter is not significantly different from zero, which indeed confirms that consideration probabilities do not depend on expected utility.

5.3 Expedia application

In our final application we use click and purchase decisions on Expedia.com for the period between November 1, 2012 and June 20, 2013, which was originally used in Ursu (2018).[18] The data was

[18] See https://services.informs.org/dataset/mksc/download.php?doi=mksc.2017.1072 for the replication data.

provided as part of a contest on Kaggle.com and contains search queries when Expedia's ranking of hotels is used as well as a random ranking of hotels. For estimation of our model, we use the random ranking only. In particular, we use the data for Destination 1 in Ursu (2018). For this destination the data contains 1,055 consumer-specific search queries. For each of these search queries, the data contains all hotels listed, their search ranking (position), as well as price, whether there was a promotion at the time of the search query, hotel stars, review score, whether the hotel is part of a chain, and a location score. We also know whether a hotel was clicked on (which is interpreted as a search), and whether the consumer booked the hotel. The total number of different hotels listed for Destination 1 is 47, bringing the total number of observations to 34,311.

We do not have data on market shares, so we cannot estimate hotel fixed effects using the Berry inversion. Instead, we estimate the model using search and purchase decisions only, using hotel characteristics and position as explanatory variables. We also only observe search queries when one of the hotels is clicked on, which means that all consideration sets contain at least one of the hotels.[19] However, conditional on search a consumer can opt for the outside good by not purchasing.

Table 4 gives estimates when using the data for Destination 1 from Ursu (2018). Specification (1) normalizes the weight parameter to $w = 0.5$.[20] Search costs include a constant and position–both are highly significant. Specification (2) also estimates the weight parameter. The estimated weight parameter of 0.964 suggests consumers care mostly about expected utility when making decisions about which hotels to consider. Specification (3) adds utility variables as search cost shifters. Although these additional variables are not significantly different from zero, this specification is useful for comparing the fit of the search model to the ASC model in specification (5).

Columns (4) and (5) give estimates for the ASC model with logit consideration. The ASC model can be obtained by setting the weight parameter to zero. Since this means that $\hat{D}_i$ is 1 for all possible draws of $\mathbf{u}^r$, these specifications can be estimated without using the Monte Carlo estimator. Even though the position parameter estimate is similar to that that for the search model,

[19]The Monte Carlo estimator can be easily adjusted for this by deducting the expected utility of the consideration set that does not contain any firms.

[20]For $w = 0.5$ the probability that consumer i searches S and buys j simplifies to

$$s_{iSj} = \frac{\exp[\delta_{ij} - \bar{c}_{is}]}{\Pi_i \left(1 + \sum_{k=1}^{J} \frac{\exp[\delta_{ik}]}{1+\exp[\bar{c}_{i\{g\}}]}\right)},$$

where $\Pi_i = \prod_{g \in F} \left(1 + \exp[\bar{c}_{i\{g\}}]\right)$. Although specification (1) in Table 4 is estimated using the Monte Carlo estimator, using the closed form expression instead gives virtually identical results.

Table 4: Estimation Results Expedia Data

	Search and consideration model			ASC model	
	(1)	(2)	(3)	(4)	(5)
Utility					
Constant	-5.259	-4.257	-4.392	-4.772	-4.772
	(2.425)	(0.361)	(0.769)	(2.485)	(2.497)
Price ($100)	-1.385	-0.703	-0.701	-0.787	-0.787
	(0.309)	(0.071)	(0.099)	(0.312)	(0.314)
Promotion	0.419	0.329	0.134	-0.162	-0.162
	(0.284)	(0.058)	(0.113)	(0.292)	(0.293)
Stars	0.958	0.563	0.354	-0.031	-0.031
	(0.281)	(0.055)	(0.093)	(0.293)	(0.294)
Review score	-0.562	-0.148	0.010	0.009	-0.009
	(0.462)	(0.070)	(0.125)	(0.478)	(0.480)
Chain	0.133	-0.048	0.122	0.235	0.235
	(0.309)	(0.043)	(0.107)	(0.320)	(0.321)
Location score	0.578	0.109	0.193	0.757	0.757
	(0.517)	(0.075)	(0.170)	(0.533)	(0.534)
Search costs					
Constant	4.227	5.519	5.116	4.156	5.510
	(0.051)	(0.089)	(1.121)	(0.051)	(0.521)
Position	0.020	0.016	0.013	0.020	0.019
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Price ($100)			-0.016		0.890
			(0.098)		(0.065)
Promotion			-0.299		-0.547
			(0.171)		(0.076)
Stars			-0.371		-1.038
			(0.136)		(0.077)
Review score			0.290		0.584
			(0.201)		(0.116)
Chain			0.298		0.151
			(0.175)		(0.076)
Location score			0.201		-0.100
			(0.245)		(0.113)
Weight parameter	0.500	0.964	0.966	0.000	0.000
		(0.004)	(0.004)		
Log-likelihood	-5016.14	-4824.57	-4800.23	-5035.04	-4806.10

Notes: Standard errors in parentheses. Estimates are obtained using the data for Destination 1 from Ursu (2018). The number of (Owen-randomized) quasi-random draws for the Monte Carlo estimator is 10,201. Total number of observations is 34,311.

the estimated utility parameters are quite different. Moreover, the search model in specification (2) gives a much better fit. Of course, in the ASC model it is important to allow the consideration set probabilities to depend on utility shifters as well. Specification (5) includes all utility shifters in the consideration set probabilities. Note that this specification does worse in terms of fit than the equivalent specification of the search model—a log-likelihood ratio test of the comparison between columns 3 and 5 gives a test statistic of $2 \times (-4800.23 - (-4806.10)) = 11.74$, so we can conclude (at a 1% level) that the non-sequential search model gives a better fit than the ASC model with logit consideration, even in the most favorable case for the ASC model in which all utility shifters also act as consideration set shifters.

6 Conclusions

This paper has proposed a framework for the estimation of demand for differentiated products in which consideration sets are determined endogenously through non-sequential search. In our model consumers are initially unaware of whether a given product is a good match or not. Consumer decision making consists of a search stage and a purchase stage. In the search stage, consumers optimally determine which sellers to visit in order to maximize expected utility. In making this decision, consumers take into account their preferences for the various alternatives as well as the costs of searching them. In the purchase stage, after the matching parameters of all products in their choice sets are revealed, consumers either pick the good with the highest realized utility among the products searched, or else go for the outside option. A novel feature of our model is that we allow for a parameter that determines how much weight consumers put on expected utility versus search costs when making search decisions and we have shown that the ASC model with logit consideration arises as a special case of the model in which consumers put zero weight on expected utility.

Modeling demand for differentiated products when consideration sets are endogenously determined through non-sequential search poses a few challenges for the researcher. First, finding a consumer's optimal search set is a very complicated task when the number of alternatives is large and the existing methods do not generalize to common choice situations. We have proposed a fix to this problem that consists of adding a TIEV distributed shock to the costs of searching a choice set. We have shown that our model belongs to the GEV class of discrete choice models, which implies that it is consistent with utility maximization and that search data are necessary to distinguish between the search model and the equivalent full information model. Second, estimation of product demand suffers from a dimensionality problem because the product of a firm may be part of a very large number of search sets. To deal with this problem, we have provided a novel Monte Carlo approach and have shown that our estimator of the purchase probabilities is accurate and computationally fast. Finally, we have proposed a maximum likelihood approach to estimate demand and search costs using individual-level data on search and purchases. We have illustrated using several existing data sets that our approach works well in practice and can also distinguish between data that has consideration sets determined endogenously through search or through other means, as in the ASC model.

Appendix

A Derivation Equation (15) for Individual Purchase Probabilities

Based on the discussion from Section 2 and using the notation from Section 2.2, we have $\bar{c}_{i\varnothing} = 0$. Recall from Section 4 that

$$\bar{\delta}_{iS} \equiv \sum_{j \in S} \exp\left[\delta_{ij}\right].$$

and notice that it can be decomposed as follows:

$$\bar{\delta}_{iS} = \bar{\delta}_{i\{f\}} + \bar{\delta}_{iS\setminus\{f\}},$$

where $\bar{\delta}_{i\{f\}} = \sum_{j \in \mathcal{G}_f} \exp\left[\delta_{ij}\right]$ and $\bar{\delta}_{iS\setminus\{f\}} = \sum_{j \in S\setminus\{f\}} \exp\left[\delta_{ij}\right]$. Notice that for $S = \varnothing$, we have $\bar{\delta}_{i\varnothing} = 0$. Using this notation, we write P_{iS} and $P_{ij|S}$ as:[21]

$$P_{iS} = \frac{\left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]}, \tag{A26}$$

$$P_{ij|S} = \frac{\exp[\delta_{ij}]}{1+\bar{\delta}_{iS}}. \tag{A27}$$

Therefore:

$$\begin{aligned}
s_{ij} &= \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S} \\
&= \sum_{S \in \boldsymbol{S}_f} \frac{\left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]} \frac{\exp[\delta_{ij}]}{1+\bar{\delta}_{iS}} \\
&= \exp\left[\delta_{ij}\right] \frac{\sum_{S \in \boldsymbol{S}_f} \left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1} \exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]} && \text{(A28)} \\
&= \exp\left[\delta_{ij}\right] \exp\left[-\bar{c}_{i\{f\}}\right] \frac{\sum_{S \in \boldsymbol{S}_{-f}} \left(1+\bar{\delta}_{i\{f\}}+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1} \exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right]}, && \text{(A29)}
\end{aligned}$$

[21] The denominator of the fraction in P_{iS} in equation (7) is

$$1 + \sum_{S' \in \boldsymbol{S}\setminus\varnothing} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right] = \sum_{S' \in \boldsymbol{S}} \left(1+\bar{\delta}_{iS'}\right)^{\frac{w}{1-w}} \exp\left[-\bar{c}_{iS'}\right].$$

where f is the firm producing j and $\boldsymbol{S}_f$ is the set of all choice sets containing firm f. The last equality follows from the fact that the numerator of the fraction in s_{ij} in equation (A28) is

$$\sum_{S\in\boldsymbol{S}_f}\left(1+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1}\exp\left[-\bar{c}_{iS}\right]=\exp\left[-\bar{c}_{i\{f\}}\right]\sum_{S\in\boldsymbol{S}_{-f}}\left(1+\bar{\delta}_{i\{f\}}+\bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1}\exp\left[-\bar{c}_{iS}\right],$$

where $\boldsymbol{S}_{-f}$ is the set of all choice sets that do not contain firm f.

Now divide both the numerator and denominator of equation (A29) by $\sum_{S'\in\boldsymbol{S}}\exp\left[-\bar{c}_{iS'}\right]$. Noticing that

$$\sum_{S\in\boldsymbol{S}}\exp\left[-\bar{c}_{iS}\right]=\prod_{f\in\mathcal{F}}\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)=\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)\prod_{g\in\mathcal{F}\backslash\{f\}}\left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right),$$

and

$$\sum_{S\in\boldsymbol{S}_{-f}}\exp\left[-\bar{c}_{iS}\right]=\prod_{g\in\mathcal{F}\backslash\{f\}}\left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)$$

gives expression (15) for the individual purchase probabilities.

B Derivation of Buying Probabilities under Normalization $\mathbf{w}=\frac{1}{2}$

Using the normalization $w=\frac{1}{2}$ in equation (8) gives

$$s_{ij}=\exp\left[\delta_{ij}\right]\frac{\sum_{S\in\boldsymbol{S}_f}\exp\left[-\bar{c}_{iS}\right]}{\sum_{S'\in\boldsymbol{S}}(1+\bar{\delta}_{iS'})\exp[-\bar{c}_{iS'}]}.$$

We get

$$\begin{aligned}s_{ij}&=\exp[\delta_{ij}]\frac{\sum_{S\in\boldsymbol{S}_{-f}}\exp\left[-\bar{c}_{i\{f\}}-\bar{c}_{iS}\right]}{\sum_{S'\in\boldsymbol{S}}(1+\bar{\delta}_{iS'})\exp[-\bar{c}_{iS'}]}\\&=\exp[\delta_{ij}]\frac{\exp\left[-\bar{c}_{i\{f\}}\right]\sum_{S\in\boldsymbol{S}_{-f}}\exp\left[-\bar{c}_{iS}\right]}{\sum_{S'\in\boldsymbol{S}}(1+\bar{\delta}_{iS'})\exp[-\bar{c}_{iS'}]}\\&=\exp[\delta_{ij}]\frac{\exp\left[-\bar{c}_{i\{f\}}\right]\sum_{S\in\boldsymbol{S}_{-f}}\prod_{f\in S}\exp\left[-\bar{c}_{i\{f\}}\right]}{\sum_{S'\in\boldsymbol{S}}(1+\bar{\delta}_{iS'})\exp[-\bar{c}_{iS'}]}.\end{aligned}$$

Since

$$\sum_{S\in\boldsymbol{S}_{-f}}\prod_{f\in S}\exp\left[-\bar{c}_{i\{f\}}\right]=\prod_{g\in F\backslash\{f\}}\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)=\frac{\prod_{g\in F}\left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)}{\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)},$$

we have that

$$s_{ij} = \exp[\delta_{ij}] \frac{\exp\left[-\bar{c}_{i\{f\}}\right] \frac{\prod\limits_{g \in F} \left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)}{\left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)}}{\sum_{S' \in \boldsymbol{S}} (1+\bar{\delta}_{iS'}) \exp[-\bar{c}_{iS'}]} = \frac{\exp\left[\delta_{ij} - \ln\left(1+\exp\left[\bar{c}_{i\{f\}}\right]\right)\right] \Pi_i}{\sum_{S' \in \boldsymbol{S}} (1+\bar{\delta}_{iS'}) \exp[-\bar{c}_{iS'}]},$$

where $\Pi_i = \prod\limits_{g \in F} \left(1+\exp\left[-\bar{c}_{i\{g\}}\right]\right)$.

Now note that s_{i0} is given by

$$\begin{aligned} s_{i0} &= \frac{1+\sum_{S \in \boldsymbol{S} \setminus \{\varnothing\}} \exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} (1+\bar{\delta}_{iS'}) \exp[-\bar{c}_{iS'}]} = \frac{\sum_{S \in \boldsymbol{S}} \prod\limits_{f \in S} \exp\left[-\bar{c}_{i\{f\}}\right]}{\sum_{S' \in \boldsymbol{S}} (1+\bar{\delta}_{iS'}) \exp[-\bar{c}_{iS'}]} \\ &= \frac{\prod\limits_{f \in F} \left(1+\exp\left[-\bar{c}_{i\{f\}}\right]\right)}{\sum_{S' \in \boldsymbol{S}} (1+\bar{\delta}_{iS'}) \exp[-\bar{c}_{iS'}]} = \frac{\Pi_i}{\sum_{S' \in \boldsymbol{S}} (1+\bar{\delta}_{iS'}) \exp[-\bar{c}_{iS'}]}. \end{aligned}$$

Since $\sum_{j=0}^{J} s_{ij} = 1$, it has to be that

$$\begin{aligned} \sum_{S' \in \boldsymbol{S}} (1+\bar{\delta}_{iS'}) \exp[-\bar{c}_{iS'}] &= \Pi_i + \sum_{k=1}^{J} \exp\left[\delta_{ik} - \ln\left(1+\exp\left[\bar{c}_{i\{g\}}\right]\right)\right] \Pi_i \\ &= \Pi_i \left(1 + \sum_{k=1}^{J} \exp\left[\delta_{ik} - \ln\left(1+\exp\left[\bar{c}_{i\{g\}}\right]\right)\right]\right). \end{aligned}$$

Consequently,

$$\begin{aligned} s_{ij} &= \frac{\exp\left[\delta_{ij} - \ln\left(1+\exp\left[\bar{c}_{i\{f\}}\right]\right)\right]}{1+\sum_{k=1}^{J} \exp\left[\delta_{ik} - \ln\left(1+\exp\left[\bar{c}_{i\{g\}}\right]\right)\right]} \\ &= \frac{\exp\left[\delta_{ij} - \ln\left(1+\exp\left[\kappa\left(t'_{i\{f\}}\gamma_i\right)\right]\right)\right]}{1+\sum_{k=1}^{J} \exp\left[\delta_{ik} - \ln\left(1+\exp\left[\kappa\left(t'_{i\{g\}}\gamma_i\right)\right]\right)\right]}. \end{aligned} \tag{A30}$$

and

$$s_{i0} = \frac{1}{1+\sum_{k=1}^{J} \exp\left[\delta_{ik} - \ln\left(1+\exp\left[\kappa\left(t'_{i\{g\}}\gamma_i\right)\right]\right)\right]}.$$

C Monte Carlo Estimation of Individual Purchase Probabilities

To complete the explanation on how to estimate the individual purchase probabilities by Monte Carlo, we now provide details on how to estimate the numerator of the second fraction in equation

(15):

$$N_{if} \equiv \sum_{S \in \boldsymbol{S}_{-f}} \left(1 + \bar{\delta}_{i\{f\}} + \bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1} \frac{\exp\left[-\bar{c}_{iS}\right]}{\prod\limits_{g \in \mathcal{F} \setminus \{f\}} \left(1 + \exp\left[-\bar{c}_{i\{g\}}\right]\right)}.$$

Again, because $\exp\left[-\bar{c}_{iS}\right] = \prod\limits_{f \in S} \exp\left[-\bar{c}_{i\{f\}}\right]$, for S not containing f we have:

$$Q_{ifS} \equiv \frac{\exp\left[-\bar{c}_{iS}\right]}{\prod\limits_{g \in \mathcal{F} \setminus \{f\}} \left(1 + \exp\left[-\bar{c}_{i\{g\}}\right]\right)} = \prod_{g \in S} \frac{\exp\left[-\bar{c}_{i\{g\}}\right]}{1 + \exp\left[-\bar{c}_{i\{g\}}\right]} \prod_{\substack{g \notin S, \\ g \neq f}} \frac{1}{1 + \exp\left[-\bar{c}_{i\{g\}}\right]} = \prod_{g \in S} \phi_{ig} \prod_{\substack{g \notin S, \\ g \neq f}} \left(1 - \phi_{ig}\right),$$

where

$$\phi_{ig} \equiv \frac{\exp\left[-\bar{c}_{i\{g\}}\right]}{1 + \exp\left[-\bar{c}_{i\{g\}}\right]}.$$

Hence, we can write N_{if} as:

$$N_{if} = \sum_{S \in \boldsymbol{S}_{-f}} \left(1 + \bar{\delta}_{i\{f\}} + \bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1} Q_{ifS}.$$

Like it was the case for the denominator, $\sum_{S \in \boldsymbol{S}_{-f}} Q_{ifS} = 1$ so $\left(Q_{ifS}\right)_{S \in \boldsymbol{S}_{-f}}$ can be regarded as a probability mass function. Therefore, N_{if} can be interpreted as the expected value of a discrete random variable taking on values $\left[\left(1 + \bar{\delta}_{i\{f\}} + \bar{\delta}_{iS}\right)^{\frac{w}{1-w}-1}\right]_{S \in \boldsymbol{S}_{-f}}$ with probabilities $\left(Q_{ifS}\right)_{S \in \boldsymbol{S}_{-f}}$. Hence, similar to equation (17), by letting

$$\bar{\delta}_{if}\left(\mathbf{u}_{-f}^{r}\right) = \sum_{g \in \mathcal{F} \setminus \{f\}} \mathbf{1}\left(u_g \leq \phi_{ig}\right) \exp\left[\delta_{ig}\right],$$

we can then rewrite N_{if} as:

$$N_{if} = \int_{[0,1]^{F-1}} \left(1 + \bar{\delta}_{i\{f\}} + \bar{\delta}_{if}\left(\mathbf{u}_{-f}\right)\right)^{\frac{w}{1-w}-1} d\mathbf{u}_{-f}.$$

Note that we can also write N_{if} as

$$N_{if} = \int_{[0,1]^{F}} \left(1 + \bar{\delta}_{i\{f\}} + \bar{\delta}_{if}\left(\mathbf{u}_{-f}\right)\right)^{\frac{w}{1-w}-1} d\mathbf{u},$$

which holds due to the fact that $\int_{[0,1]} 1 du = 1$. This way of writing N_{if} is useful because it allows the Monte Carlo estimates of the market shares to depend on the same draws. Like before, to obtain a smooth-in-parameters estimator of N_{if}, we use $\Phi\left(\dfrac{\phi_{ig} - u_g}{h}\right)$ rather than $\mathbf{1}\left(u_g \leq \phi_{ig}\right)$, where Φ is standard normal CDF and h is a smoothing parameter (see Section 3.1). Let

$$\widetilde{\delta}_{if}\left(\mathbf{u}_{-f}\right)=\sum_{g\in\mathcal{F}\backslash\{f\}}\Phi\left(\frac{\phi_{ig}-u_g}{h}\right)\exp\left[\delta_{ig}\right]$$

and the corresponding smooth-in-parameters version of N_{if}

$$\widetilde{N}_{if}=\int_{[0,1]^F}\left(1+\bar{\delta}_{i\{f\}}+\widetilde{\delta}_{if}\left(\mathbf{u}_{-f}\right)\right)^{\frac{w}{1-w}-1}d\mathbf{u}.$$

D Supply Side

We assume manufacturers maximize profits in a pricing game. Assuming a Nash equilibrium exists for this game, any product sold should have prices that satisfy the first order conditions

$$s_j(p)+\sum_{r\in\mathcal{G}_f}(p_r-mc_r)\frac{\partial s_r(p)}{\partial p_j}=0.$$

We assume deviation prices are *not* not observable. Since this implies $\partial P_{iS}/\partial p_k=0$, we get

$$\frac{\partial s_j}{\partial p_k}=\int\sum_{S\in\mathcal{S}_j}\left(P_{iS}\frac{\partial P_{ij|S}}{\partial p_k}\right)f(\tau_i)d\tau_i.$$

This means

$$\frac{\partial s_j}{\partial p_k}=\begin{cases}\int-\alpha_i\left(s_{ij}-\sum_{S\in\mathcal{S}_j}P_{iS}P_{ij|S}^2\right)f(\tau_i)d\tau_i & \text{if } k=j;\\ \int-\alpha_i\left(-\sum_{S\in\mathcal{S}_j}P_{iS}P_{ij|S}P_{ik|S}\right)f(\tau_i)d\tau_i & \text{if } k\neq j.\end{cases}\tag{A31}$$

To estimate the marginal costs, the first order conditions can be rewritten as in BLP

$$p-\Delta(p)^{-1}s(p)=mc,\tag{A32}$$

where the element of $\Delta(p)$ in row j column r is denoted by Δ_{jr} and

$$\Delta_{jr}=\begin{cases}-\dfrac{\partial s_r}{\partial p_j}, & \text{if } r \text{ and } j \text{ are produced by the same firm;}\\ 0, & \text{otherwise.}\end{cases}$$

References

Abaluck, Jason and Abi Adams-Prassl: "What Do Consumers Consider Before they Choose? Identification from Asymmetric Demand Responses," *The Quarterly Journal of Economics* 136, 1611–1663, 2021.

Amemiya, Takeshi: "Qualitative Response Models: A Survey," *Journal of Economic Literature* 19, 1483–1536, 1981.

Anderson, Simon P., André de Palma, and Jacques-François Thisse: "Discrete Choice Theory of Product Differentiation," The MIT Press, Cambridge, MA, 1992.

Anderson, Simon P. and André de Palma: "Multiproduct Firms: A Nested Logit Approach," *The Journal of Industrial Economics* 40, 261–176, 1992.

Anderson, Simon P. and Régis Renault: "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *RAND Journal of Economics* 30, 719–735, 1999.

Armstrong, Mark: "Ordered Consumer Search," *Journal of European Economic Association* 15, 989–1024, 2017.

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum: "Heterogeneous Choice Sets and Preferences," *Econometrica* 89, 2015–2048, 2021.

Ben-Akiva, Moshe and Bruno Boccara: "Discrete Choice Models with Latent Choice Sets," *International Journal of Research in Marketing* 12, 9–24, 1995.

Berger, Jack: "The Impact of Limited Consideration on Market Power: Evidence from the Hospital Industry," Mimeo, 2024.

Berry, Steven: "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics* 25, 242–262, 1994.

Berry, Steven, James Levinsohn, and Ariel Pakes: "Automobile Prices in Market Equilibrium," *Econometrica* 63, 841–890, 1995.

Berry, Steven, James Levinsohn, and Ariel Pakes: "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New product Market," *Journal of Political Economy* 112, 68–105, 2004.

Bratley, Paul, Bennett L. Fox, and Harald Niederreiter: "Implementation and Tests of Low-Discrepancy Sequences," *ACM Transactions on Modeling and Computer Simulation* 2, 195–2013, 1992.

Bronnenberg, Bart and Wilfried Vanhonacker: "Limited Choice Sets, Local Price Response, and Implied Measures of Price Competition," *Journal of Marketing Research* 33, 163–173. 1996.

Burdett, Kenneth and Kenneth L. Judd: "Equilibrium Price Dispersion," *Econometrica* 51, 955–969, 1983.

Chade, Hector and Lones Smith: "Non-Sequential Search," *Econometrica* 74, 1293–1307, 2006.

Chib, Siddhartha and Kenichi Shimizu: “Scalable Estimation of Multinomial Response Models with Uncertain Consideration Sets,” Mimeo, 2023.

Ching, Andrew, Tülin Erdem, and Michael Keane: “The Price Consideration Model of Brand Choice,” *Journal of Applied Econometrics* 24, 393–420, 2009.

Ching, Andrew, Tülin Erdem, and Michael Keane: “A Simple Method to Estimate the Roles of Learning, Inventories, and Category Consideration in Consumer Choice,” *The Journal of Choice Modelling* 13, 60–72, 2014.

Choi, Michael, Anovia Yifan Dai, and Kyungmin Kim: “Consumer Search and Price Competition,” *Econometrica* 86, 1257–1281, 2018.

Crawford, Gregory S., Rachel Griffith, and Alessandro Iaria: “A Survey of Preference Estimation with Unobserved Choice Set Heterogeneity,” *Journal of Econometrics* 222, 4–43, 2021.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest: “Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior,” *American Economic Review* 102, 2955–2980, 2012.

Donna, Javier D., Pedro Pereira, Tiago Pires, and André Trindade: “Measuring the Welfare of Intermediaries,” *Management Science* 68, 8083–8115, 2022.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan: “Consumer Price Search and Platform Design in Internet Commerce,” *American Economic Review* 108, 1820–1859, 2018.

Dressler, Luisa and Stefan Weiergraeber: “Alert the Inert? Switching Costs and Limited Awareness in Retail Electricity Markets,” *American Economic Journal: Microeconomics* 15, 74–116, 2023.

Ershov, Daniel: “The Effects of Consumer Search Costs on Entry and Quality in the Mobile App Market,” Mimeo, 2018.

Goolsbee, Austan and Amil Petrin: “The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV,” *Econometrica* 72, 351–381, 2004.

Grieco, Paul L. E., Charles Murry, Joris Pinkse, and Stephan Sagl: “Conformant and Efficient Estimation of Discrete Choice Demand Models,” Mimeo, 2023.

Hauser, John and Birger Wernerfelt: “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research* 16, 393–408. 1990.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton: “The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program,” *RAND Journal of Economics* 48 877–905, 2017.

Hong, Han and Matthew Shum: “Using Price Distributions to Estimate Search Costs,” *RAND Journal of Economics* 37, 257–275, 2006.

Honka, Elisabeth: “Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry,” *RAND Journal of Economics* 45, 847–884, 2014.

Honka, Elisabeth and Pradeep Chintagunta: “Non-Sequential or Sequential? Search Strategies in the US Auto Insurance Industry,” *Marketing Science* 36, 21–40, 2017.

Honka, Elisabeth, Ali Hortaçsu, and Matthijs R. Wildenbeest: "Empirical Search and Consideration Sets." In Jean-Pierre Dubé and Peter Rossi (editors), Handbook of the Economics of Marketing, Elsevier, 193–257, 2019.

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L. Puller: "Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market," *American Economic Journal: Economic Policy* 9 192–226, 2017.

Hortaçsu, Ali and Chad Syverson: "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *Quarterly Journal of Economics* 119, 403–456, 2004.

Kim, Jun, Paulo Albuquerque, and Bart J. Bronnenberg: "Online Demand Under Limited Consumer Search," *Marketing Science* 29, 1001–1023, 2010.

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

Lee, Younghwan: "Fast Computation Algorithm for the Random Consideration Set Model," *Economics Letters* 179, 38–41, 2019.

Leslie, Philip: "Price Discrimination in Broadway Theater," *RAND Journal of Economics* 35, 520–541, 2004.

Lin, Haizhen and Matthijs Wildenbeest: "Non-Parametric Estimation of Search Costs for Differentiated Products: Evidence from Medigap," *Journal of Business & Economic Statistics* 38, 754–770, 2020.

Lu, Zhentong: "Estimating Multinomial Choice Models with Unobserved Choice Sets," *Journal of Econometrics* 226, 368–398, 2022.

Manski, Charles: "The Structure of Random Utility Models," *Theory and Decision* 8, 229–254, 1977.

McFadden, Daniel: "Modeling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 75–96, 1978.

Moraga-González, José Luis and Matthijs R. Wildenbeest: "Maximum Likelihood Estimation of Search Costs," *European Economic Review* 52, 820–848, 2008.

Moraga-González, José Luis, Zsolt Sándor, and Matthijs R. Wildenbeest: "Consumer Search and Prices in the Automobile Market," CEPR Discussion Paper No. 10487, 2015 (Accessible at https://cepr.org/publications/dp10487).

Moraga-González, José Luis, Zsolt Sándor, and Matthijs R. Wildenbeest: "Prices and Heterogeneous Search Costs," *RAND Journal of Economics* 48, 125–146, 2017.

Moraga-González, José Luis, Zsolt Sándor, and Matthijs R. Wildenbeest: "Non-Sequential Search for Differentiated Products: the Impact of Search Costs and Firm Prominence," *The Economic Journal* 131, 1308–1330, 2021.

Moraga-González, José Luis, Zsolt Sándor, and Matthijs R. Wildenbeest: "Consumer Search and Prices in the Automobile Market," *Review of Economic Studies* 90, 1394–1440, 2023.

Morgan, Peter and Richard Manning: “Optimal Search,” *Econometrica* 53, 923–944, 1985.

Murry, Charles and Yiyi Zhou: “Consumer Search and Automobile Dealer Co-Location,” *Management Science* 66, 1909–1934, 2020.

Nurski, Laura and Frank Verboven: “Exclusive Dealing as a Barrier to Entry? Evidence from Automobiles,” *Review of Economic Studies* 83, 1156–1188, 2016.

Pires, Tiago: “Costly Search and Consideration Sets in Storable Goods Markets,” *Quantitative Marketing and Economics* 14, 157–193, 2016.

Pires, Tiago: “Measuring the Effects of Search Costs on Equilibrium Prices and Profits,” *International Journal of Industrial Organization* 60, 179–205, 2018.

Roberts, John H. and James M. Lattin: “Development and Testing of a Model of Consideration Set Composition,” *Journal of Marketing Research* 28, 429–440, 1991.

Sándor, Zsolt and Péter András: “Alternative Sampling Methods for Estimating Multivariate Normal Probabilities,” *Journal of Econometrics* 120, 207–234, 2004.

Seiler, Stephan: “The Impact of Search Costs on Consumer Behavior: A Dynamic Approach,” *Quantitative Marketing and Economics* 11, 155–203, 2013.

Sovinsky Goeree, Michelle: “Limited Information and Advertising in the U.S. Personal Computer Industry,” *Econometrica* 76, 1017–1074, 2008.

Stigler, George: “The Economics of Information,” *Journal of Political Economy* 69, 213–225, 1961.

Swait, Joffre and Moshe Ben-Akiva: “Incorporating Random Constraints in Discrete Models of Choice Set Generation,” *Transportation Research Part B* 21, 91-102, 1987.

Swait, Joffre: “Choice Set Generation within the Generalized Extreme Value Family of Discrete Choice Models,” *Transportation Research Part B* 35, 643–666, 2001.

Train, Kenneth E. and Clifford Winston: “Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers,” *International Economic Review* 48, 1469–1496, 2007.

Ursu, Raluca M.: “The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions,” *Marketing Science* 37, 530–552, 2018.

Van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip Hans Franses: “Retrieving Unobserved Consideration Sets from Household Panel Data,” *Journal of Marketing Research* 47, 63–74, 2010.

Yavorsky, Dan, Elisabeth Honka, and Keith Chen: “Consumer Search in the U.S. Auto Industry: The Role of Dealership Visits,” *Quantitative Marketing and Econonomics* 19, 1–52, 2021.

Weitzman, Martin L.: “Optimal Search for the Best Alternative,” *Econometrica* 47, 641–654, 1979.

Wolinsky, Asher: “True Monopolistic Competition as a Result of Imperfect Competition,” *Quarterly Journal of Economics* 101, 493–511, 1986.

Supplementary Material

S1 The Contraction Property

Here we present a version of the Contraction Theorem from BLP also used by Moraga-González, Sándor, and Wildenbeest (2023). This Contraction Theorem is essentially the same as BLP's but here the conditions are specified on the market share function instead of the contraction mapping itself. From equation (2) of the main text, the first order derivatives of s_{ij} with respect to $\delta_j = \alpha p_j + x_j'\beta + \xi_j$ and δ_h when $j, h \in \mathcal{G}_f$ and when $h \in \mathcal{G}_g,\ g \neq f$, are:

$$\frac{\partial s_{ij}}{\partial \delta_j} = \left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S}^2 + \left(1 - \frac{w}{1-w} s_{ij}\right) s_{ij}, \tag{S33}$$

$$\frac{\partial s_{ij}}{\partial \delta_h} = \left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S} P_{ih|S} - \frac{w}{1-w} s_{ih} s_{ij} \quad \text{for} \quad j, h \in \mathcal{G}_f, \tag{S34}$$

$$\frac{\partial s_{ij}}{\partial \delta_h} = \left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}_f \cap \boldsymbol{S}_g} P_{iS} P_{ij|S} P_{ih|S} - \frac{w}{1-w} s_{ih} s_{ij} \quad \text{for} \quad h \in \mathcal{G}_g,\ g \neq f. \tag{S35}$$

The first order derivatives of the choice probability of the outside option s_{i0} with respect to δ_j' is

$$\frac{\partial s_{i0}}{\partial \delta_j} = \left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}} P_{iS} P_{i0|S} P_{ij|S} - \frac{w}{1-w} s_{ij} s_{i0}, \tag{S36}$$

where

$$s_{i0} = \sum_{S \in \boldsymbol{S}} P_{iS} P_{i0|S} = \sum_{S \in \boldsymbol{S}} \frac{\exp[m_{iS}/(1-w)]}{\sum_{S' \in \boldsymbol{S}} \exp[m_{iS'}/(1-w)]} \frac{1}{1 + \sum_{r \in S} \exp[\delta_{ir}]}. \tag{S37}$$

S1.1 Contraction theorem

Contraction Theorem (BLP). Let $f : \mathbb{R}^J \to \mathbb{R}^J$ be defined as

$$f_j(\delta) = \delta_j + \log s_j^o - \log s_j(\delta), \quad j = 1, \ldots, J,$$

where $s^o = (s_1^o, \ldots, s_J^o)$ is the vector of observed market shares and suppose that the market share vector $s(\delta)$ as a function of $\delta = (\delta_1, \ldots, \delta_J) \in \mathbb{R}^J$ satisfies the following conditions.

1. s is continuously differentiable in δ and

$$\frac{\partial s_j}{\partial \delta_j}(\delta) \le s_j(\delta), \quad \frac{\partial s_j}{\partial \delta_k}(\delta) \le 0 \text{ for any } j,\ k \ne j \text{ and } \delta \in \mathbb{R}^J,$$

(the former is equivalent to the fact that the function $\overline{s}_j : \mathbb{R}^J \to \mathbb{R}$, $\overline{s}_j(\delta) = s_j(\delta)\exp(-\delta_j)$ is decreasing in δ_j) and

$$\sum_{k=1}^{J} \frac{\partial s_j}{\partial \delta_k}(\delta) > 0 \text{ for any } \delta \in \mathbb{R}^J.$$

2. The function $\overline{s}_j$ defined in Condition 1 satisfies

$$\lim_{\delta \to -\infty} \overline{s}_j(\delta) > 0.$$

3. The share of the outside alternative $s_0(\delta) = 1 - \sum_{j=1}^{J} s_j(\delta)$ is decreasing in all its arguments and it satisfies that for any j and $x \in \mathbb{R}$ the limit

$$\lim_{\delta_{-j} \to -\infty} s_0(\delta_1, \ldots, \delta_{j-1}, x, \delta_{j+1}, \ldots, \delta_J) \equiv \widetilde{s}_0^j(x)$$

is finite and the function $\widetilde{s}_0^j : \mathbb{R} \to \mathbb{R}$ obtained as the limit satisfies that

$$\lim_{x \to -\infty} \widetilde{s}_0^j(x) = 1 \text{ and } \lim_{x \to \infty} \widetilde{s}_0^j(x) = 0,$$

where $\delta_{-j} \to -\infty$ means that $\delta_1 \to -\infty, \ldots, \delta_{j-1} \to -\infty, \delta_{j+1} \to -\infty, \ldots, \delta_J \to -\infty$.

Then there are values $\underline{\delta}, \overline{\delta} \in \mathbb{R}$ such that the function $\overline{f} : [\underline{\delta}, \overline{\delta}]^J \to \mathbb{R}^J$ defined by $\overline{f}_j(\delta) = \min[\overline{\delta}, f_j(\delta)]$ has the property that $\overline{f}\left([\underline{\delta}, \overline{\delta}]^J\right) \subseteq [\underline{\delta}, \overline{\delta}]^J$, is a contraction with modulus less than 1 with respect to the sup norm $\|(x_1, \ldots, x_J)\| = \max_j |x_j|$, and, in addition, f has no fixed point outside $[\underline{\delta}, \overline{\delta}]^J$.

S1.2 Verifying the contraction theorem conditions

We next verify the conditions for our market share vector function $s = (s_1, \ldots, s_J)$ for $w \le \frac{1}{2}$, where $s_j = \int s_{ij} f_\tau(\tau_i) d\tau_i$, $j = 0, 1, \ldots, J$.

Condition 1. The market share vector s is obviously continuously differentiable in δ.

The inequality $\frac{\partial s_j}{\partial \delta_j} \leq s_j$ holds if (see equation (S33))

$$\left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S}^2 + \left(1 - \frac{w}{1-w} s_{ij}\right) s_{ij} - s_{ij} \leq 0,$$

i.e.,

$$\left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S}^2 - \frac{w}{1-w} s_{ij}^2 \leq 0.$$

This is clearly true for $w \leq \frac{1}{2}$.

The inequality $\frac{\partial s_j}{\partial \delta_k} < 0$ holds if (see equations (S34) and (S35))

$$\left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}_f} P_{iS} P_{ij|S} P_{ih|S} - \frac{w}{1-w} s_{ih} s_{ij} < 0 \quad \text{and}$$
$$\left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}_f \cap \boldsymbol{S}_g} P_{iS} P_{ij|S} P_{ih|S} - \frac{w}{1-w} s_{ih} s_{ij} < 0 \quad \text{for} \quad h \in \mathcal{G}_g, \; g \neq f.$$

These clearly hold for $w \leq \frac{1}{2}$.

Finally, to prove $\sum_{k=1}^{J} \frac{\partial s_j}{\partial \delta_k} > 0$, first note that $\frac{\partial s_j}{\partial \delta_k} = \frac{\partial s_k}{\partial \delta_j}$ (see equations (S34) and (S35)). Then

$$\sum_{k=1}^{J} \frac{\partial s_j}{\partial \delta_k} = \sum_{k=1}^{J} \frac{\partial s_k}{\partial \delta_j} = -\frac{\partial s_0}{\partial \delta_j}.$$

If $w \leq \frac{1}{2}$ then (see equation (S36))

$$\frac{\partial s_{i0}}{\partial \delta_j} = \left(\frac{w}{1-w} - 1\right) \sum_{S \in \boldsymbol{S}} P_{iS} P_{i0|S} P_{ij|S} - \frac{w}{1-w} s_{ij} s_{i0} < 0 \tag{S38}$$

and therefore $-\frac{\partial s_0}{\partial \delta_j} > 0$. This completes the proof of Condition 1.

Condition 2. We have

$$\lim_{\delta \to -\infty} s_j \exp\left[-\delta_j\right] = \int \lim_{\delta \to -\infty} s_{ij} \exp\left[-\delta_j\right] f_\tau(\tau_i) d\tau_i.$$

Further,

$$s_{ij} \exp\left[-\delta_j\right] = \sum_{S \in \boldsymbol{S}_f} \frac{\exp[m_{iS}/(1-w)]}{\sum_{S' \in \boldsymbol{S}} \exp[m_{iS'}/(1-w)]} \frac{\exp[d_{ij}]}{1 + \sum_{g \in S} \exp[\delta_{ig}]},$$

where $d_{ij} = \delta_{ij} - \delta_j$ does not depend on δ_j. Since by equation (4) of the main text

$$\lim_{\delta \to -\infty} \exp[m_{iS}/(1-w)] = \exp\left(-\bar{c}_{iS}\right), \tag{S39}$$

$$\lim_{\delta \to -\infty} \frac{1}{1 + \sum_{g \in S} \sum_{h \in \mathcal{G}_g} \exp[\delta_{ih}]} = 1, \tag{S40}$$

we get that

$$\lim_{\delta \to -\infty} s_{ij} \exp\left[-\delta_j\right] = \sum_{S \in \boldsymbol{S}_f} \frac{\exp\left[-\bar{c}_{iS}\right]}{\sum_{S' \in \boldsymbol{S}} \exp\left[-\bar{c}_{iS'}\right]} \exp[d_{ij}] > 0.$$

So Condition 2 holds.

Condition 3. The fact that the share of the outside alternative $s_0 = 1 - \sum_{j=1}^{J} s_j$ is decreasing in all its arguments follows from (S38) for $w \leq \frac{1}{2}$.

In order to see that $\widetilde{s}_0^j(x)$ finite, first notice that the limit $\lim_{\delta_{-j} \to -\infty} s_0$ is (see equation (S37))

$$\int \lim_{\delta_{-j} \to -\infty} \sum_{S \in \boldsymbol{S}} \frac{\exp[m_{iS}/(1-w)]}{\sum_{S' \in \boldsymbol{S}} \exp[m_{iS'}/(1-w)]} \frac{1}{1 + \sum_{g \in S} \sum_{h \in \mathcal{G}_g} \exp[\delta_{ih}]} f_\tau(\tau_i) d\tau_i.$$

If $j \in \mathcal{G}_f$,

$$\lim_{\delta_{-j} \to -\infty} \exp[m_{iS}/(1-w)] = \begin{cases} \exp\left[w \log\left(1 + \exp[\delta_{ij}]\right)/(1-w) - \bar{c}_{iS}\right] & \text{if } f \in S \\ \exp\left(-\bar{c}_{iS}\right) & \text{if } f \notin S \end{cases}$$

and

$$\lim_{\delta_{-j} \to -\infty} \frac{1}{1 + \sum_{g \in S} \sum_{h \in \mathcal{G}_g} \exp[\delta_{ih}]} = \begin{cases} \dfrac{1}{1 + \exp[\delta_{ij}]} & \text{if } f \in S \\ 1 & \text{if } f \notin S. \end{cases}$$

So

$$\lim_{\delta_{-j} \to -\infty} \sum_{S \in \boldsymbol{S}} \exp[m_{iS}/(1-w)] = \sum_{S \in \boldsymbol{S}_f} \exp\left[w \log\left(1 + \exp[\delta_{ij}]\right)/(1-w) - \bar{c}_{iS}\right] + \sum_{S \notin \boldsymbol{S}_f} \exp\left(-\bar{c}_{iS}\right).$$

Since all these limits exist and are strictly positive, $\lim_{\delta_{-j} \to -\infty} s_0$ will be finite.

The limit $\lim_{x \to -\infty} \widetilde{s}_0^j(x)$ is

$$\int \lim_{\delta \to -\infty} \sum_{S \in \boldsymbol{S}} \frac{\exp[m_{iS}/(1-w)]}{\sum_{S' \in \boldsymbol{S}} \exp[m_{iS'}/(1-w)]} \frac{1}{1 + \sum_{g \in S} \sum_{h \in \mathcal{G}_g} \exp[\delta_{ih}]} f_\tau(\tau_i) d\tau_i.$$

From equations (S39) and (S40)

$$\lim_{\delta\to-\infty}\sum_{S\in\boldsymbol{S}}\frac{\exp[m_{iS}/(1-w)]}{\sum_{S'\in\boldsymbol{S}}\exp[m_{iS'}/(1-w)]}\frac{1}{1+\sum_{g\in S}\sum_{h\in\mathcal{G}_g}\exp[\delta_{ih}]}=\lim_{\delta\to-\infty}\sum_{S\in\boldsymbol{S}}\frac{\exp[m_{iS}/(1-w)]}{\sum_{S'\in\boldsymbol{S}}\exp[m_{iS'}/(1-w)]}=1.$$

The limit $\lim_{x\to\infty}\widetilde{s}_0^j(x)$ is

$$\int\lim_{\delta_j\to\infty}\left(\lim_{\delta_{-j}\to-\infty}\sum_{S\in\boldsymbol{S}}\frac{\exp[m_{iS}/(1-w)]}{\sum_{S'\in\boldsymbol{S}}\exp[m_{iS'}/(1-w)]}\frac{1}{1+\sum_{g\in S}\sum_{h\in\mathcal{G}_g}\exp[\delta_{ih}]}\right)f_\tau(\tau_i)d\tau_i.$$

Note that with $j\in\mathcal{G}_f$

$$\lim_{\delta_j\to\infty}\left(\lim_{\delta_{-j}\to-\infty}\exp[m_{iS}/(1-w)]\right)=\begin{cases}\infty & \text{if } f\in S\\ \exp\left(-\bar{c}_{iS}\right) & \text{if } f\notin S,\end{cases}$$

so

$$\lim_{\delta_j\to\infty}\left(\lim_{\delta_{-j}\to-\infty}\sum_{S\in\boldsymbol{S}}\exp[m_{iS}/(1-w)]\right)=\infty.$$

Also

$$\lim_{\delta_j\to\infty}\left(\lim_{\delta_{-j}\to-\infty}\frac{1}{1+\sum_{g\in S}\sum_{h\in\mathcal{G}_g}\exp[\delta_{ih}]}\right)=\begin{cases}0 & \text{if } f\in S\\ 1 & \text{if } f\notin S,\end{cases}$$

so

$$\lim_{\delta_{-j}\to-\infty}\sum_{S\in\boldsymbol{S}}\frac{\exp[m_{iS}/(1-w)]}{\sum_{S'\in\boldsymbol{S}}\exp[m_{iS'}/(1-w)]}\lim_{\delta_j\to\infty}\frac{1}{1+\sum_{g\in S}\sum_{h\in\mathcal{G}_g}\exp[\delta_{ih}]}=\lim_{\delta_{-j}\to-\infty}\sum_{S\notin\boldsymbol{S}_f}\frac{\exp[m_{iS}/(1-w)]}{\sum_{S'\in\boldsymbol{S}}\exp[m_{iS'}/(1-w)]}.$$

Therefore,

$$\begin{aligned}&\lim_{\delta_j\to\infty}\left(\lim_{\delta_{-j}\to-\infty}\sum_{S\in\boldsymbol{S}}\frac{\exp[m_{iS}/(1-w)]}{\sum_{S'\in\boldsymbol{S}}\exp[m_{iS'}]}\frac{1}{1+\sum_{g\in S}\sum_{h\in\mathcal{G}_g}\exp[\delta_{ih}]}\right)\\&=\lim_{\delta_j\to\infty}\left(\lim_{\delta_{-j}\to-\infty}\sum_{S\notin\boldsymbol{S}_f}\frac{\exp[m_{iS}/(1-w)]}{\sum_{S'\in\boldsymbol{S}}\exp[m_{iS'}]}\right)\\&=\sum_{S\notin\boldsymbol{S}_f}\frac{\exp\left(-\bar{c}_{iS}\right)}{\infty}=0.\end{aligned}$$

So Condition 3 is satisfied. □

S2 The Smoothing Parameter for the Monte Carlo Estimator

In this section we present Monte Carlo evidence on how the Monte Carlo estimation error depends on the smoothing parameter. One of the main questions here is whether the performance of the Monte Carlo estimator is sensitive to the choice of the smoothing parameter. To answer this, we study the performance of the estimator in equation (18) of the main text for estimating sums of the type in equation (17) of the main text. Specifically, we estimate

$$D = \sum_{S \in \boldsymbol{S}} \left(1 + \sum_{f \in S} \delta_f\right)^{\frac{w}{1-w}} \prod_{f \in S} \phi_f \prod_{f \notin S} (1 - \phi_f) = \int_{[0,1]^F} \left(1 + \sum_{f=1}^{F} \mathbf{1}\left(u_f \leq \phi_f\right) \exp\left[\delta_f\right]\right)^{\frac{w}{1-w}} d\mathbf{u},$$

where $\phi_f = \frac{\exp[-c_f]}{1+\exp[-c_f]}$, by

$$\widetilde{D} = \frac{1}{R} \sum_{r=1}^{R} \left(1 + \sum_{f \in \mathcal{F}} \Phi\left(\frac{\phi_f - u_f^r}{h}\right) \exp\left[\delta_f\right]\right)^{\frac{w}{1-w}}, \tag{S41}$$

where Φ is the standard normal CDF, h is a smoothing parameter, and $\{(u_1^r, u_2^r, \ldots, u_F^r)\}_{r=1}^{R} \subset [0,1]^F$ is a quasi-random sample of size R of type $(0, 2, s)$-net (e.g., Sándor and András, 2004).[22]

For different values of F (reported in Tables S1 and S2) we draw vectors $(\delta_1, \ldots, \delta_F)$ and $(c_1, \ldots, c_F)$ randomly such that $\delta_f \sim N(0, 25)$ and $c_f \sim N(0, 1)$. We compute (S41) 100 times by using different randomized versions of the quasi-random sample; we also calculate the actual value of D and we multiply the ratio of the two by 1000. This way we normalize D to 1000, so the Monte Carlo estimation errors can be compared across the different cases. Based on the estimates computed 100 times we calculate the root-mean-squared-error (hereafter RMSE). Finally, we replicate this procedure 10 times and report the means and standard deviations (Stds) of the 10 RMSE values in Tables S1 and S2.

We do these computations for various values of w, h, and R (see Tables S1 and S2; the h values appear in the first line called "Bandwidths"). In the tables we can see that for cases in which the smoothing parameter is lower than 10^{-3} the RMSE's are relatively lower than those for larger smoothing parameter values. This implies that both the bias and the standard deviation is small in these cases. In both tables we can notice that smoothing parameter values $h = 3 \cdot 10^{-4}$ or lower

[22] Using this type of quasi-random sample turns out to reduce computing time at least by a factor of 40 with respect to a pseudo-random sample because even the lowest sample (of size 256) yields more precise estimates than a pseudo-random sample of size $10,000$.

Table S1: Means and Standard Deviations of RMSE's

w	R	Bandwidths	$3 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$3 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
					Number of firms $F = 5$					
0.17	256	**Means**	**30.59**	**10.47**	**3.68**	**2.11**	**2.00**	**2.21**	**2.16**	**2.19**
		Stds	25.42	8.55	2.74	1.24	1.01	1.33	1.23	1.31
	529	**Means**	**26.65**	**9.06**	**2.91**	**1.40**	**1.07**	**1.04**	**1.00**	**0.99**
		Stds	24.79	8.24	2.43	0.86	0.63	0.65	0.69	0.61
	1024	**Means**	**28.71**	**9.72**	**2.97**	**1.16**	**0.65**	**0.60**	**0.61**	**0.66**
		Stds	21.00	7.08	2.14	0.80	0.51	0.41	0.42	0.45
0.33	256	**Means**	**17.15**	**5.79**	**2.05**	**1.53**	**1.64**	**1.69**	**1.75**	**1.85**
		Stds	11.26	3.71	1.07	0.68	0.86	1.01	1.03	1.06
	529	**Means**	**17.31**	**5.90**	**1.99**	**1.16**	**1.04**	**1.09**	**1.24**	**1.23**
		Stds	9.31	3.07	0.86	0.38	0.37	0.34	0.38	0.42
	1024	**Means**	**26.58**	**8.88**	**2.71**	**0.96**	**0.55**	**0.63**	**0.68**	**0.69**
		Stds	19.17	6.39	1.90	0.59	0.25	0.38	0.44	0.44
0.44	256	**Means**	**9.55**	**3.35**	**1.69**	**2.03**	**2.16**	**2.29**	**2.33**	**2.37**
		Stds	9.63	3.31	1.35	1.48	1.68	1.94	1.90	1.82
	529	**Means**	**11.19**	**3.77**	**1.36**	**1.05**	**1.45**	**1.73**	**1.79**	**1.80**
		Stds	6.52	2.16	0.70	0.46	0.67	0.81	0.86	0.82
	1024	**Means**	**8.78**	**2.95**	**0.91**	**0.46**	**0.50**	**0.60**	**0.70**	**0.69**
		Stds	5.93	2.00	0.59	0.29	0.25	0.28	0.34	0.33
					Number of firms $F = 10$					
0.17	256	**Means**	**9.29**	**3.19**	**1.16**	**0.81**	**0.81**	**0.75**	**0.72**	**0.76**
		Stds	12.02	4.10	1.21	0.64	0.50	0.43	0.40	0.42
	529	**Means**	**9.60**	**3.28**	**1.10**	**0.59**	**0.52**	**0.55**	**0.57**	**0.57**
		Stds	12.17	4.12	1.25	0.47	0.30	0.35	0.35	0.33
	1024	**Means**	**14.81**	**4.95**	**1.51**	**0.55**	**0.29**	**0.28**	**0.29**	**0.28**
		Stds	15.03	5.01	1.51	0.51	0.24	0.23	0.24	0.24
0.33	256	**Means**	**22.91**	**7.84**	**2.80**	**2.25**	**2.63**	**2.79**	**2.90**	**2.86**
		Stds	23.13	7.98	2.80	2.17	2.58	2.56	2.69	2.52
	529	**Means**	**16.45**	**5.51**	**1.68**	**0.77**	**0.65**	**0.62**	**0.65**	**0.60**
		Stds	26.10	8.72	2.56	0.99	0.65	0.60	0.57	0.59
	1024	**Means**	**10.86**	**3.63**	**1.12**	**0.44**	**0.34**	**0.36**	**0.35**	**0.38**
		Stds	14.78	4.93	1.48	0.53	0.33	0.34	0.32	0.33
0.44	256	**Means**	**5.96**	**2.10**	**1.21**	**1.62**	**1.93**	**2.00**	**2.07**	**2.14**
		Stds	9.02	3.07	1.49	1.97	2.33	2.26	2.59	2.65
	529	**Means**	**4.17**	**1.37**	**0.50**	**0.47**	**0.65**	**0.82**	**0.90**	**0.90**
		Stds	4.60	1.56	0.49	0.35	0.47	0.58	0.65	0.72
	1024	**Means**	**4.77**	**1.60**	**0.51**	**0.28**	**0.31**	**0.39**	**0.43**	**0.45**
		Stds	6.37	2.12	0.62	0.26	0.26	0.35	0.40	0.40
					Number of firms $F = 15$					
0.17	256	**Means**	**20.05**	**7.08**	**3.04**	**2.36**	**2.29**	**2.33**	**2.33**	**2.46**
		Stds	10.44	3.48	1.18	0.65	0.74	0.78	0.80	0.81
	529	**Means**	**29.40**	**10.16**	**3.50**	**1.84**	**1.67**	**1.63**	**1.65**	**1.66**
		Stds	21.73	7.48	2.32	1.05	1.11	1.00	1.07	1.01
	1024	**Means**	**14.99**	**5.24**	**1.76**	**1.00**	**0.85**	**0.83**	**0.89**	**0.82**
		Stds	8.84	2.98	0.86	0.36	0.35	0.39	0.43	0.32
0.33	256	**Means**	**23.28**	**8.06**	**3.09**	**2.22**	**2.68**	**2.85**	**3.11**	**3.02**
		Stds	15.55	5.02	1.23	0.70	1.16	1.19	1.24	1.22
	529	**Means**	**22.35**	**7.63**	**2.59**	**1.36**	**1.23**	**1.33**	**1.43**	**1.50**
		Stds	11.46	3.67	0.91	0.39	0.51	0.54	0.46	0.49
	1024	**Means**	**16.32**	**5.56**	**1.91**	**1.00**	**0.77**	**0.82**	**0.84**	**0.85**
		Stds	14.32	4.74	1.31	0.36	0.30	0.31	0.34	0.32
0.44	256	**Means**	**5.45**	**1.92**	**1.17**	**1.50**	**1.81**	**1.93**	**2.00**	**2.11**
		Stds	3.21	1.07	0.48	0.63	1.01	1.04	1.04	0.96
	529	**Means**	**10.69**	**3.62**	**1.31**	**0.95**	**1.19**	**1.28**	**1.34**	**1.37**
		Stds	8.73	2.95	1.03	0.59	0.67	0.75	0.95	0.86
	1024	**Means**	**5.25**	**1.78**	**0.62**	**0.39**	**0.45**	**0.53**	**0.60**	**0.60**
		Stds	3.15	1.05	0.30	0.13	0.13	0.14	0.17	0.16

Table S2: Means and Standard Deviations of RMSE's

w	R	Bandwidths	$3 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$3 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
				Number of firms $F = 5$						
0.63	529	**Means**	**26.32**	**8.89**	**3.07**	**1.92**	**2.00**	**2.26**	**2.34**	**2.32**
		Stds	12.86	4.33	1.33	1.19	1.28	1.44	1.33	1.42
	1024	**Means**	**30.05**	**10.02**	**3.06**	**1.18**	**0.87**	**1.07**	**1.15**	**1.16**
		Stds	16.25	5.41	1.62	0.54	0.39	0.55	0.60	0.63
	2209	**Means**	**23.53**	**7.85**	**2.38**	**0.85**	**0.45**	**0.42**	**0.43**	**0.48**
		Stds	11.20	3.73	1.10	0.34	0.17	0.13	0.13	0.16
0.83	529	**Means**	**70.27**	**23.62**	**7.91**	**3.97**	**3.24**	**3.27**	**3.31**	**3.36**
		Stds	43.03	14.34	4.54	2.70	2.54	2.65	2.68	2.70
	1024	**Means**	**65.20**	**22.02**	**7.49**	**4.18**	**3.56**	**3.34**	**3.27**	**3.59**
		Stds	48.81	16.90	6.74	6.07	6.97	6.16	6.21	7.22
	2209	**Means**	**94.33**	**31.50**	**9.55**	**3.73**	**1.90**	**1.52**	**1.58**	**1.57**
		Stds	71.36	23.77	7.13	2.72	1.78	1.76	1.86	1.84
0.95	529	**Means**	**202.98**	**85.00**	**50.05**	**41.32**	**38.72**	**38.52**	**44.40**	**40.08**
		Stds	89.24	62.36	70.31	68.51	69.58	67.53	82.24	70.89
	1024	**Means**	**201.68**	**68.64**	**22.48**	**10.76**	**8.45**	**7.63**	**7.53**	**7.40**
		Stds	108.28	36.40	11.68	6.96	8.27	7.65	7.75	7.54
	2209	**Means**	**213.60**	**72.83**	**23.36**	**10.23**	**6.60**	**5.78**	**6.24**	**6.34**
		Stds	121.33	40.78	12.90	7.38	7.83	6.92	8.58	9.09
				Number of firms $F = 10$						
0.63	529	**Means**	**18.17**	**6.15**	**2.10**	**1.38**	**1.46**	**1.52**	**1.71**	**1.70**
		Stds	7.09	2.35	0.88	0.85	1.10	0.99	1.23	1.19
	1024	**Means**	**28.96**	**9.67**	**2.94**	**1.15**	**0.82**	**0.97**	**1.07**	**1.12**
		Stds	23.44	7.82	2.33	0.77	0.43	0.56	0.69	0.75
	2209	**Means**	**24.51**	**8.18**	**2.48**	**0.91**	**0.48**	**0.49**	**0.52**	**0.53**
		Stds	9.43	3.15	0.94	0.34	0.21	0.21	0.27	0.29
0.83	529	**Means**	**101.83**	**34.62**	**11.84**	**6.37**	**5.02**	**5.34**	**4.94**	**5.16**
		Stds	80.72	27.14	9.23	6.53	6.10	7.23	6.22	6.71
	1024	**Means**	**82.80**	**28.01**	**9.12**	**4.39**	**3.00**	**3.00**	**2.97**	**2.99**
		Stds	25.30	8.35	2.89	3.07	3.70	4.02	3.91	3.61
	2209	**Means**	**98.16**	**32.79**	**9.97**	**3.66**	**1.83**	**1.51**	**1.49**	**1.49**
		Stds	80.21	26.74	8.03	2.67	1.20	1.03	1.08	1.11
0.95	529	**Means**	**150.73**	**59.55**	**29.70**	**24.32**	**23.01**	**21.51**	**23.56**	**23.72**
		Stds	82.95	44.53	37.50	39.76	38.13	34.34	38.52	40.17
	1024	**Means**	**193.87**	**69.06**	**29.14**	**19.22**	**17.00**	**17.53**	**17.33**	**17.19**
		Stds	62.31	25.65	21.32	23.22	22.63	24.97	24.88	25.03
	2209	**Means**	**224.70**	**77.70**	**26.06**	**14.06**	**11.96**	**12.36**	**11.47**	**11.49**
		Stds	95.69	33.43	12.34	9.24	9.77	11.00	10.00	9.97
				Number of firms $F = 15$						
0.63	529	**Means**	**24.80**	**8.33**	**2.88**	**1.82**	**2.04**	**2.25**	**2.35**	**2.30**
		Stds	14.82	4.94	1.43	0.90	1.02	1.06	1.09	1.15
	1024	**Means**	**18.29**	**6.13**	**1.96**	**1.10**	**1.03**	**1.10**	**1.14**	**1.10**
		Stds	8.60	2.86	0.88	0.56	0.56	0.62	0.64	0.54
	2209	**Means**	**22.40**	**7.47**	**2.29**	**0.90**	**0.56**	**0.52**	**0.53**	**0.57**
		Stds	17.18	5.72	1.69	0.52	0.22	0.17	0.16	0.16
0.83	529	**Means**	**93.78**	**31.55**	**10.88**	**6.36**	**5.53**	**5.66**	**5.66**	**5.69**
		Stds	68.11	22.72	7.07	4.83	4.56	4.76	4.51	4.71
	1024	**Means**	**74.16**	**25.40**	**9.18**	**5.79**	**4.70**	**4.87**	**4.82**	**4.61**
		Stds	40.05	13.59	5.15	4.93	4.87	4.94	5.03	4.66
	2209	**Means**	**100.91**	**34.18**	**12.15**	**7.73**	**7.09**	**6.93**	**6.75**	**6.99**
		Stds	70.22	23.99	10.16	9.43	9.78	10.18	9.79	10.02
0.95	529	**Means**	**212.03**	**89.31**	**58.85**	**52.51**	**50.42**	**49.58**	**49.97**	**50.56**
		Stds	86.01	51.29	56.44	61.37	58.73	59.31	58.12	58.28
	1024	**Means**	**180.61**	**67.45**	**33.31**	**25.94**	**25.37**	**25.96**	**24.78**	**25.68**
		Stds	71.21	29.16	23.09	23.79	26.25	26.45	24.71	25.95
	2209	**Means**	**187.63**	**66.89**	**28.92**	**19.65**	**17.52**	**18.58**	**17.26**	**18.19**
		Stds	64.95	24.37	16.16	16.07	16.65	17.72	16.80	17.86

yield rather similar mean RMSE's.

We can also notice that in most cases the RMSE's are lower for higher sample sizes R, which is something we would expect. Changing the number of firms F does not seem to alter significantly the magnitudes of the RMSE's. We can notice a similar phenomenon when we look at changes in the number of firms F. Based on the results from Table S1 we can conclude that, when we take a smoothing parameter lower than or equal to 10^{-3} and sample size $R = 529$, the Monte Carlo estimator (S41) is expected to have both bias and standard deviation lower than 0.2% of the true value of D. This remarkable precision explains why in Table 1 of the main text the parameter estimates obtained when using the Monte Carlo estimator are almost identical to those when using the actual probabilities. Comparing the results from Tables S1 and S2 we can see that when the weight parameter w is larger than 0.5 the RMSE's will also be larger. Nevertheless, the Monte Carlo estimator (S41) is still rather precise. Specifically, based on the results from Table S2 we can conclude that, when we take a smoothing parameter lower than or equal to $3 \cdot 10^{-4}$ and sample size 2209, the Monte Carlo estimator (S41) is expected to have both bias and standard deviation lower than 2% of the true value of D.