Search with Learning for Differentiated Products: Evidence from E-Commerce*

Babur De los Santos†
Ali Hortaçsu‡
Matthijs R. Wildenbeest§

November 2015

Abstract

This paper provides a method to estimate search costs in a differentiated product environment in which consumers are uncertain about the utility distribution. Consumers learn about the utility distribution by Bayesian updating their Dirichlet process prior beliefs. The model provides expressions for bounds on the search costs that can rationalize observed search and purchasing behavior. Using individual-specific data on web browsing and purchasing behavior for MP3 players sold online we show how to use these bounds to estimate search costs as well as the parameters of the utility distribution. Our estimates indicate that search costs are sizable. We show that ignoring consumer learning while searching can lead to severely biased search cost and elasticity estimates.

Key words: consumer behavior, consumer search, discrete choice models of demand

*We are grateful to the Editor, an Associate Editor, and three anonymous referees for their very useful comments and suggestions. In addition we thank Matt Backus, Mitsukuni Nishida, and Benjamin Shiller for their useful comments and suggestions. This paper has also benefited from presentations at the 2014 IIOC in Chicago, the 29th Annual Congress of the European Economic Association in Toulouse, the 2014 Workshop on Search and Switching Costs at Indiana University, the 2014 MaCCI Workshop on Consumer Search in Bad Homburg, and the 2015 EARIE Annual Conference in Munich. Ali Hortaçsu acknowledges financial support of the NSF (SES 1426823 and SES 1124073). This paper was previously circulated under the title "Search with Learning."

†Clemson University, E-mail: `babur@clemson.edu`.
‡University of Chicago and NBER, E-mail: `hortacsu@uchicago.edu`.
§Kelley School of Business, Indiana University, E-mail: `mwildenb@indiana.edu`.

1 Introduction

Over the past fifty years, a large literature on search has tried to explain markets that are characterized by imperfectly competitive behavior. In a typical search model, consumers make a tradeoff between the gains from searching and the cost of searching to determine whether to continue searching or, alternatively, how many times to search. The gains from search are typically derived using the assumption that consumers "know" the distribution of prices or wages (Stigler, 1961; McCall, 1970; Mortensen, 1970). Several papers diverge from this view and have analyzed optimal search behavior when consumers are not only uncertain about actual draws but also about the distribution (Rothschild, 1974; Rosenfield and Shapiro, 1981; Chou and Talmain, 1993; Dana, 1994; Bikhchandani and Sharma, 1996). This is especially important since search behavior has been shown to be sensitive to the assumed distribution (see, e.g., Gastwirth, 1976).

In this paper we develop a method to estimate consumer search costs for differentiated products when consumers have only partial information about the distribution from which is being sampled. In the next section we present a model, based on the works of Rothschild (1974), Rosenfield and Shapiro (1981), and Bikhchandani and Sharma (1996), which relaxes the assumption that consumers "know" the distribution of offerings while deciding on their search strategy, and allows for learning of the utility distribution. More specifically, consumers learn about the utility distribution by Bayesian updating their Dirichlet process priors while sampling information about products and retailers. We use information on a consumer's sequence of searches to derive expressions for bounds on the search cost that rationalizes the consumer's observed search behavior. The intuition behind this approach is straightforward: if a consumer stops searching this means she found an alternative that has a higher utility than her reservation utility. Since reservation utility is a function of both search costs and the expected gains from searching, this inequality can be inverted to obtain a lower bound on the consumer's search cost. Similarly, if a consumer continues searching after having searched an alternative, this means her search cost should have been lower than the expected gains from search, which can be used to obtain an upper bound on her search cost.

If products are homogeneous and consumers are only searching for the lowest price, the search cost bounds can be obtained directly from the observed gains from search. In the case of differentiated products it is more difficult to obtain search cost bounds because the bounds are conditional on the parameters of the utility function as well as an unobserved utility component. We show how to map a differentiated product utility framework into the learning model and derive an estimation strategy for the parameters of both the utility function and search cost distribution. As such, these bounds allow us to estimate the relationship between observed consumer characteristics and search costs using simulated maximum likelihood.

An important feature of the learning model is that reservation utilities are decreasing in the number of alternatives sampled, but constant in a model in which there is no learning. We show that if consumers have correct prior beliefs, i.e., the base distribution of the Dirichlet process equals the true utility distribution, the decreasing reservation utility property will result in less overall search activity than in the no-learning model. The decreasing reservation utility property may also trigger recall: a sampled alternative that was not good enough initially might pass the bar after a few searches, once more alternatives have been sampled and the reservation utility has gone down. This is useful for explaining actual search data: the search patterns in the data reveal that in close to a third of transactions consumers recall a previously visited firm. As argued in De los Santos, Hortaçsu, and Wildenbeest (2012), this violates optimal behavior in the most basic version of the sequential search model. Although recall can also be rationalized if sequential search is directed (as in Weitzman, 1979) or if consumers use fixed sample size search rules (De los Santos, Hortaçsu, and Wildenbeest, 2012), learning provides an additional mechanism—in many markets it is plausible that consumers update their beliefs while searching and as such learning may be an important driver of recall behavior.

Our estimation strategy applies to settings in which purchase decisions as well as search histories are observed. In Section 3, we present an application of the model using data on the web browsing and purchasing behavior of a large panel of consumers. We focus on purchases of MP3 players. Our data not only allow us to observe online transactions for all consumers in the panel, but also which online stores have been visited shortly before a transaction. We let utility be a function of retailer and product characteristics as well as an idiosyncratic unobserved component. Our estimates indicate that median search costs are close to \$28. Our model gives a better fit to the data than a sequential search model in which consumers do not update their priors. Moreover, we find search costs to be uniformly lower in the learning model.

In Section 4, we test the identification properties of our model using Monte Carlo experiments. The experiments confirm that our estimation method can recover the parameters of the utility function and the distribution of search costs. Furthermore, we show that ignoring learning leads to biased search cost and elasticity estimates. As predicted by the model, if consumers have true beliefs about the prior distribution the bias will be towards higher search costs. Elasticity estimates are biased towards zero for all retailers. In this section we also study how measurement errors in the choice sets and prices affect the estimation.

Our paper relates to several recent papers that provide empirical methods to estimate search costs. The vast majority of these paper assume consumers do not update their beliefs about the distribution of prices when searching for homogenous goods (Hong and Shum, 2006; Moraga-González and Wildenbeest, 2008) or, when searching for differentiated goods, other components of the utility function (Hortaçsu and Syverson, 2004; Kim, Albuquerque, and Bronnenberg, 2010; Wildenbeest, 2011; De los Santos, Hortaçsu, and Wildenbeest, 2012; Honka, 2014; Koulayev, 2014). An exception is Koulayev (2013), who estimates a model of search with Dirichlet priors for homogenous products using aggregate data on prices and market shares. Unlike our dataset, Koulayev's data does not contain information on search sequences—to be able to estimate the model he derives closed-form ex-ante buying probabilities. Although less general than the Dirichlet *process* priors we use in our paper, Dirichlet priors also imply search decisions at any given time can be characterized by the identity of the best alternative observed so far and the number of searches to date, which greatly simplifies integrating out the unobserved search histories. Another related paper is Häubl, Dellaert, and Donkers (2010), who estimate the parameters of a product differentiation and learning model in an experimental setting. Their experimental data allows them to observe the identity of the alternative with the highest utility for each subject at any point in the search sequence, which substantially simplifies the estimation of especially the parameters of the utility function. Our paper is also related to a literature that estimates Bayesian learning models (Erdem and Keane, 1996; Ackerberg, 2003; Crawford and Shum, 2005; Chernew, Gowrisankaran, and Scanlon, 2008). An important difference with our paper is that while we explicitly model consumers' joint search and learning decisions, search is not explicitly modeled in the literature on the estimation of Bayesian learning models.

The product differentiation model we use is similar to those used in several recent papers on search and product differentiation (Kim, Albuquerque, and Bronnenberg, 2010; Honka, 2014; Koulayev, 2014; Moraga-González, Sándor, and Wildenbeest, 2015). However, unlike these papers, we assume that consumers have no prior knowledge on the realization of any of the characteristics that appear in the utility function. This means that consumers in our model search randomly across firms, while consumers would search in a directed way if some attributes would be observed by consumers before searching (see, e.g., Weitzman, 1979). Note that our assumption that none of the characteristics are observed prior to searching can be relaxed, although doing so would make the estimation of the model less tractable.

We believe that our study makes several contributions relative to existing papers. First of all, we provide a methodology to estimate search costs in environments in which learning is important, using individual-specific search data. The use of Dirichlet process priors allows us to build our model around a discrete choice product differentiation model of demand, and is therefore sufficiently flexible to allow for both horizontal and vertical differentiation. Our paper therefore brings together two separate themes in the recent search literature: search for differentiated products (Kim, Albuquerque, and Bronnenberg, 2010; Honka, 2014; Koulayev, 2014; Moraga-González, Sándor, and Wildenbeest, 2015) and search with learning (Koulayev, 2013). Our paper also shows that modeling learning is important: ignoring learning may lead to biased search cost and elasticity estimates.

2 Model

In this section we present a model in which consumers learn about the utility distribution while searching. Our framework, which is based on the learning models of Rothschild (1974), Rosenfield and Shapiro (1981), and Bikhchandani and Sharma (1996), demonstrates how to estimate search costs in a learning context. The data set in our application consists of visits and purchases by a large number of households at online retailers, which allows us to reconstruct the sequence of online retailers that have been visited shortly before a transaction. We use these search histories to obtain search cost bounds that rationalize consumers' observed behavior in a learning environment. In our application both retailers and products are differentiated—to capture this, we show how to map a product differentiation model into the search and learning model and discuss how to estimate the model using a simulated maximum likelihood procedure.

2.1 Consumer Learning

Consumers learn about the different options available to them while searching by Bayesian updating their priors on a utility distribution. We assume consumers are searching sequentially, which means that after each observation they have to determine whether to continue searching or not, making a tradeoff between the additional cost of another search and the potential gains of observing a better offer. Note that even though non-sequential search can be optimal in situations in which information has to be gathered quickly, the optimal search strategy is often a mix of non-sequential and sequential search strategies (Morgan and Manning, 1985). Nevertheless, search strategies need to be sequential for the updating to have any meaningful implications on consumers' search behavior. We assume consumers can recall previous offers at no cost. The cost of each search, including the first, is specific to consumer i, and denoted by c_i. Consumers have imperfect information about the utility distribution; each new search provides the consumer with additional information, which is used to update priors on the utility distribution. In the next subsection we discuss the learning environment. We begin by introducing our learning concept using a Dirichlet distribution over a discrete set of alternatives. We then generalize the setup by modeling learning over a continuous distribution of alternatives using a Dirichlet process.

Dirichlet Setup

Suppose first that there are N options available in the market. Let $u = \{u_1, u_2, \ldots, u_N\}$ denote the utility values of the alternatives, where the subscript indicates the index of the alternative in terms of utility. The probability of sampling each utility is given by a vector $\rho = (\rho_1, \rho_2, \ldots, \rho_N)$, where $\sum_{n=1}^{N} \rho_n = 1$. The utility values are known to consumers, while the probabilities of sampling each utility value are not. Instead, consumers consider the probabilities to be random variables that are distributed according to a Dirichlet distribution of order N with density

$$f(\rho_1, \ldots, \rho_N) = \frac{\Gamma\left(\sum_{n=1}^{N} a_n\right)}{\prod_{n=1}^{N} \Gamma(a_n)} \prod_{n=1}^{N} \rho_n^{a_n - 1},$$

where Γ is the gamma function and $a = (a_1, a_2, \ldots, a_N)$ are concentration parameters. The prior expected value of probability ρ_n is given by

$$E[\rho_n] = \frac{a_n}{\sum_{n=1}^{N} a_n},$$

where $\sum_{n=1}^{N} a_n$ can be interpreted as the weight put on the initial prior.

The prior is updated as consumers start searching and evaluating new alternatives. Since the Dirichlet distribution is the conjugate prior of the multinomial distribution, the posterior distribution will be Dirichlet as well. Specifically, the posterior expected value of ρ_n after the consumer has sampled an alternative is

$$E[\rho_n] = \begin{cases} \dfrac{a_n}{\sum_{n=1}^{N} a_n + 1} & \text{if alternative } n \text{ is not sampled;} \\[2ex] \dfrac{a_n + 1}{\sum_{n=1}^{N} a_n + 1} & \text{if alternative } n \text{ is sampled.} \end{cases}$$

A simple example illustrates the updating process. Suppose there are three options with utility values $u = (1, 2, 3)$ and consumers have a diffuse, non-informative prior, i.e., $a = (1, 1, 1)$, so the prior expected values of the probability of re-sampling each option are given by $E[\rho] = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. If a consumer starts searching and randomly draws option 2, the updated concentration parameter of option 2 is obtained by adding 1 to a_2. Hence, the posterior distribution has updated concentration parameters $a = (1, 2, 1)$ and posterior expected values $E[\rho] = \left(\frac{1}{4}, \frac{2}{4}, \frac{1}{4}\right)$.

When deciding whether to continue searching, consumers make a tradeoff between the cost of an additional search and the expected gains from search, where the latter is a function of the expected probability of finding a better alternative. The posterior probability of re-sampling an alternative with a higher utility than the best alternative observed so far only depends on the weight put on the initial prior and how many alternatives have been evaluated to date. In our example, if after drawing option 2 a consumer did not stop the search, the gains of searching are calculated using the utility value of the better alternative not yet drawn (option 3 in the example) which has a posterior probability of $\frac{1}{4}$. In general, if k is the alternative with the highest utility observed after t searches, the posterior probability of an alternative m such that $u_m > u_k$ is given by $a_m/(\sum_{n=1}^{N} a_n + t)$. This is a useful feature of this learning environment as it will allow us to characterize the gains of search using only the prior distribution and the observed sequence of searches.

Dirichlet Process

In the setting we are studying, a continuous distribution of utilities is more applicable. The Dirichlet process (DP) is a stochastic process whose realizations are probability distributions. A DP can be interpreted as an infinite dimensional generalization of a Dirichlet distribution where its marginal distributions are Dirichlet distributed. Bikhchandani and Sharma (1996) provide a useful discussion of the DP (see also Häubl, Dellaert, and Donkers, 2010) and describe a DP as a continuous distribution over probability distributions.

Following Ferguson's (1973) definition, let probability distribution D be distributed according to a DP with base distribution H over probability space Θ, and let concentration parameter W be a positive real number. Hence, D is a DP with concentration parameter W and base distribution H, $D \sim DP(W, H)$, if for any finite measurable partition $T_1, T_2, \ldots, T_N$ of Θ,

$$(D(T_1), \ldots, D(T_N)) \sim \text{Dir}\left(WH(T_1), \ldots, WH(T_N)\right),$$

where the marginal distribution $D(T_i)$ is a Dirichlet distribution with parameters $WH(T_i)$.

Posterior Distribution. Let $u_1, u_2, \ldots, u_t$ be a sequence of independent draws from D. We are interested in the posterior distribution of D given the observed draws $u_1, u_2, \ldots, u_t$. Let $\delta_{u_i} = \mathbb{1}(u_i \in T_i)$ be an indicator of observed values of u_i and let $t_i = \sum_{i=1}^{t} \delta_{u_i}$ be the number of observed values in T_i. We can write the updated marginal distributions as

$$(D(T_1), \ldots, D(T_N)) \mid u_1, u_2, \ldots, u_t \sim \text{Dir}\left(WH(T_1) + t_1, \ldots, WH(T_N) + t_N\right).$$

Since the Dirichlet process is the conjugate prior for any arbitrary distribution, the posterior distribution will be a Dirichlet process as well and can be written as

$$D \sim DP\left(W + t, \frac{W}{W+t}H + \frac{t}{W+t}\hat{H}_{u_1,\ldots,u_t}\right),$$

where $\hat{H}_{u_1,\ldots,u_t} \equiv \frac{1}{t}\sum_{i=1}^{t}\delta_{u_i}$ can be interpreted as the empirical distribution of observed utilities. In other words, the posterior of D is a Dirichlet process with updated concentration parameter $W + t$ and posterior distribution $\frac{W}{W+t}H + \frac{t}{W+t}\hat{H}_{u_1,\ldots,u_t}$. This means the updated distribution is just the weighted average of the base distribution H and the empirical distribution of observed utilities $\hat{H}$.

The posterior base distribution is also the predictive distribution of u_{t+1} and can be written as

$$u_{t+1}|u_1, u_2, \ldots, u_t \sim \frac{W}{W+t}H + \frac{t}{W+t}\hat{H}_{u_1,\ldots,u_t}. \tag{1}$$

The concentration parameter W, which can be interpreted as the weight put on the initial prior, determines how quickly searchers update: the smaller W, the faster the weight is shifted from the base distribution to the empirical distribution of observed utilities.

Pólya Urn Example. In order to better understand the predictive distribution of utilities in equation (1) it is useful to review the Pólya urn scheme developed by Blackwell and MacQueen (1973). In the basic Pólya urn scheme, the urn contains two colors and every time a color is drawn the ball is replaced and a ball of the same color is added to the urn. Blackwell and MacQueen (1973) extend the basic Pólya urn scheme by allowing for a continuum of colors and show that the posterior distribution of colors after t draws converges to a DP. Each u value in Θ represent a unique color and the sequence of draws $u_1, u_2, \ldots, u_t$ represent the results of successive draws from an urn. We start with an urn filled with W white balls. If we draw a white ball from the urn, we draw a non-white color from H, initially $u_1 \sim H$, create a new ball of the color drawn and add it to the urn. In the next steps we will either draw a white ball and hence draw a color from H or draw a non-white ball, in which case we will paint a new ball with the same color and drop it into the urn. Specifically, in step $t + 1$ we will either draw a white ball from the urn with probability $\frac{W}{W+t}$ and pick a color $u_{t+1} \sim H$, or with probability $\frac{t}{W+t}$ draw a random color from the urn, which implies a draw from the empirical distribution $u_{t+1} \sim \hat{H}$. The parameter W determines the speed of the weight updating process. A smaller W implies that weight is shifted from the base distribution to the empirical distribution at a faster rate.

Consumer Search Process. Let $\hat{u}_{it}$ denote consumer i's highest observed utility after having observed t utility draws. Notice that by definition the upper bound of the empirical distribution of observed utilities corresponds to the highest observed utility $\hat{u}_{it}$, hence the gains from search calculated using the empirical distribution of observed utilities are zero at $\hat{u}_{it}$. In other words, if utilities are drawn from the empirical distribution of observed utilities, consumer i cannot do better than the highest utility observed so far. This also means that when calculating the gains from search using the predictive distribution of utilities in equation (1), only the first term of the equation is relevant since only this part reflects the option value of searching. Hence, the gains from search are only a function of the base distribution H with updated weight $W/(W+t)$ and can be written as:

$$G(\hat{u}_{it}) = \frac{W}{W+t} \int_{\hat{u}_{it}}^{\infty} (u - \hat{u}_{it}) \cdot h(u)\ du, \tag{2}$$

where $h(u)$ is the density of the base distribution. Intuitively, the gain is equal to the expected utility when searching minus the offer at hand, considering that a consumer will keep the current offer in the event that a utility lower than $\hat{u}_{it}$ is sampled. The term $W/(W+t)$ reflects consumers' updating process: less weight is put on offers that exceed $\hat{u}_{it}$ every time a utility is drawn that is lower than $\hat{u}_{it}$. In the case of t equal to zero this expression describes the gains from search for the basic (non-learning) sequential search model.

2.2 Product Differentiation Model

In our application consumers are searching for differentiated products that are sold by differentiated retailers, which means it is important to model both product and retailer differentiation. Let there be K retailers and J products. Denote consumer i's indirect utility for product j, sold by retailer k, as

$$u_{ijk} = \alpha p_{jk} + X_j\beta + X_k\gamma + \varepsilon_{ijk}, \tag{3}$$

where p_{jk} is the price of product j sold by retailer k, X_j are product characteristics, X_k are retailer characteristics, and ε_{ijk} is a utility shock from a Type I Extreme Value (EV) distribution. We assume consumers do not know the random error ε_{ijk} before searching. The actual values of the attributes are not observed either and have to be learned. The parameters in the utility function, on the other hand, are known to the consumer, but unknown to the econometrician. Once a consumer has visited a retailer then the value of the attributes are known to the consumer and the random error term is realized. We simplify matters by assuming consumers know the joint distribution of the attributes. Since utility parameters are known to the consumer, this implies consumers know the available variety of mean utilities δ_{jk}, where $\delta_{jk} = \alpha p_{jk} + X_j\beta + X_k\gamma$. However, consumers do not know which firm is offering which mean utility until they start visiting retailers, hence alternatives are ex-ante identical to a consumer and only ex-post differentiated (see Hortaçsu and Syverson, 2004, for a similar assumption in the context of a vertical product differentiation model of search). From a consumer's perspective search is therefore random, i.e., the indirect utility for an alternative is a random draw from a mixture distribution of $J \times K$ Type I EV distributions with location parameter δ_{jk} and scale parameter 1. This means that before having searched, a consumer considers a product's utility to be a random draw from the density

$$f(u) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{J} \sum_{j=1}^{J} \exp(-(u - \delta_{jk} + \exp(-(u - \delta_{jk})))).$$

We assume that by visiting a retailer a consumer observes prices and characteristics of *all* products sold by retailer k. This simplifying assumption is necessary due to the structure of the data we use in our application: although we observe the product that was purchased by a consumer as well as which domains are visited, we do not observe consumers' browsing behavior within a retailer. Hence we cannot directly identify the products viewed by a consumer during their search. This simplifying assumption means a consumer samples J times from the utility distribution when visiting a firm k, which modifies the gains from search in equation (2) to

$$G(\hat{u}_{it}) = \frac{W}{W + t \cdot J} \sum_{k=1}^{K} \frac{1}{K} \int_{\hat{u}_{it}}^{\overline{u}} (u - \hat{u}_{it}) \cdot \frac{d}{du} \left(\prod_{J} H_{jk}(u - \delta_{jk}) \right) du,$$

where $H_{jk}(u)$ is the CDF of the initial prior of product j's utility sold by retailer k. The integral reflects the expected maximum utility of J draws from a firm's utility distribution, taking into account that each firm is selling products with different mean utilities δ_{jk}; to get the overall gains from search from randomly sampling the retailers, we can take the average over the gains from search from visiting each individual firm.

Most existing papers on search in differentiated goods markets (Kim, Albuquerque, and Bronnenberg, 2010; Honka, 2014; Koulayev, 2014; Moraga-González, Sándor, and Wildenbeest, 2015) assume that part of the alternative-specific utility function is known, while consumers have to search to resolve uncertainty about the remaining parts of the utility function. Although our assumption of no prior knowledge on the realization of any of the characteristics that appear in the utility function can be relaxed, doing so is non-trivial. For instance, if some of the characteristics would be observed before searching, the sampling process would be different: instead of searching randomly, consumers would rank alternatives according to the observed part of utility and would search in a directed way (see Weitzman, 1979). If only some of the vertical characteristics are observed, the utility function in equation (3) implies that the ex-ante ranking of alternatives is the same across consumers, which means consumers visit firms in the same order. Although this can be relaxed by allowing for random coefficients on the observed characteristics or by letting ε_{ijk} be observed as well, doing so would make the model much more difficult to estimate. Moreover, when some characteristics are observed before searching, it will be difficult to make the case that priors on the unobserved characteristics are uncorrelated with the observed characteristics. For instance, if prices are correlated with retailer quality, and quality is observed but prices are not, it will be much harder to make the case that by visiting one retailer, a consumer learns something useful about the price distribution at another retailer.

2.3 Priors

So far we have left unspecified the density of the initial prior, which corresponds to the density $h(u)$ of the base distribution of the Dirichlet process. The data we use in our application does not allow us to identify consumers' priors. However, our model is flexible enough to capture different assumptions about the initial prior. In our main specification we use an informative prior in which consumers have correct initial beliefs, i.e., consumers' initial priors correspond to the overall utility distribution such that $h(u) = f(u)$. When consumers have such an informative prior, the prior is a reflection of the search technology available to consumers; what makes it different from a search model in which there is no learning is that uncertainty about the search technology makes consumers update their beliefs throughout the search process. Following the discussion in the previous subsection, note that consumers form priors about the joint distribution of attributes and prices, which means that consumers do not have knowledge about attributes or prices at specific retailers before searching. For instance, if retailer A sells a specific MP3 player for \$100 and retailer B for \$120, having an informative prior in our model means that consumers know that there is one store that sells the MP3 player for \$100 while there is another store that sells it for \$120. Consumers do not know whether the price at store A is lower than the price at store B. In Section 3 we will also estimate the model assuming a non-informative prior.

If we assume an informative prior, the Type I EV assumption allows us to write the gains from search equation as

$$G(\hat{u}_{it}) = \frac{W}{W + t \cdot J} \sum_{k=1}^{K} \frac{1}{K} \left(\gamma + \log \left[\sum_{J} \exp(\delta_{jk}) \right] - \hat{u}_{it} + \int_{\sum_J \exp[\delta_{jk} - \hat{u}_{it}]}^{\infty} e^{-x}/x \; dx \right), \quad (4)$$

where γ is the Euler constant. In this expression, which is derived in the Appendix, the term between brackets can be broken down into three parts. The first part, given by $\gamma + \log\left[\sum_J \exp(\delta_{jk})\right]$, is the expected maximum utility when taking J draws from the utility distribution of retailer k. The second part consists of the reference utility $\hat{u}_{it}$, which needs to be subtracted since we are calculating the gains from search in comparison to this reference utility. The last part is the exponential integral function evaluated at $\sum_J \exp[\delta_{jk} - \hat{u}_{it}]$ and captures the option value of $\hat{u}_{it}$ in case the maximum utility out of J draws is less than $\hat{u}_{it}$.

A consumer with search cost c_i will continue searching as long as the gains from an additional search more than offset the cost of searching once more, i.e., $G(\hat{u}_{it}) \geq c_i$. Once this is no longer the case, the consumer will stop and buy from the store selling at the highest utility observed so far. Let the reservation utility r_{it} be the utility at which the consumer is indifferent between searching and not searching, i.e., r_{it} solves $G(r_{it}) = c_i$. In general, reservation utilities are decreasing in search costs. Figure 1 shows this for the gains from search equation (4), assuming two retailers each selling a single product, with $W = 2$, $\delta_1 = 2$, and $\delta_2 = 3$.

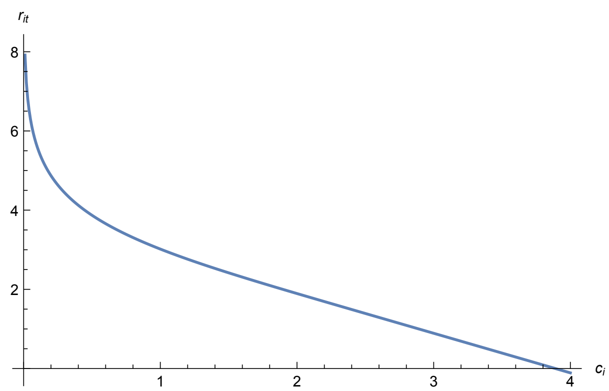


Figure 1: Reservation utility r_{it} for $t = 0$

Rosenfield and Shapiro (1981) have shown that in the related setting of a multinomial distribution with a Dirichlet prior, consumers' optimal search policy is myopic and can be characterized by a reservation utility that is non-increasing in the number of alternatives sampled. Figure 2 gives an example of such non-increasing reservation utilities for our model using the same parameter values as used in Figure 1. Also plotted is the reservation utility for the no-learning model (dashed line), using the same parameter values. Since the initial prior (i.e., the prior at $t = 0$) is the same in both models, the decreasing reservation utility for the learning model means consumers are more likely to accept a sampled alternative over time in the learning model than in the model without learning. The decreasing reservation utility is also the reason that consumers may recall in the learning model. An alternative that was below the reservation utility when it was sampled can be recalled if the reservation utility decreases below the utility of that alternative. Since the reservation utility is constant over time in the no-learning model, consumers will not recall in that model. In Section 3 we show that there are non-trivial amounts of recall in the data we use in our application, and even though there are other models that can rationalize recall (Weitzman, 1979; De los Santos, Hortaçsu, and Wildenbeest, 2012), learning provides an additional mechanism.

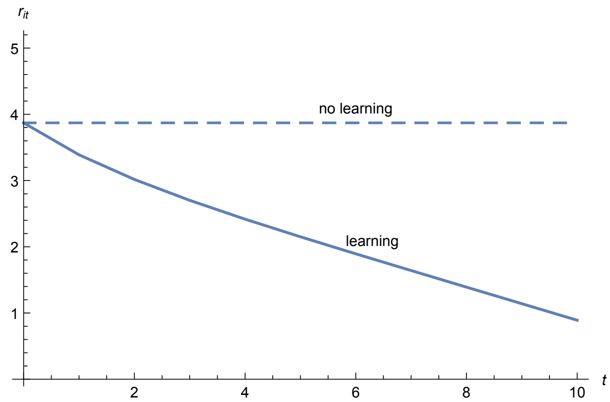


Figure 2: Reservation utility r_{it} for $c = 0.5$

Next, we show how to estimate consumers' search costs when we observe the sequence of retailers visited by a consumer prior to a transaction. More specifically, we explain how we can use this search history to obtain bounds on a consumer's search cost while simultaneously estimating the parameters of the utility function.

2.4 Estimation Strategy

A consumer will keep searching until the gains from search no longer exceed the cost of searching. We can obtain bounds on a consumer's search cost from observed search patterns. Equation (4) gives an expression for the gains from search as a function of the highest utility offer observed so far, $\hat{u}_{it}$. To obtain the lower bound of a consumer's search cost, consider a transaction as the outcome of the search process: buying a product corresponds to a decision not to continue searching, hence, the search cost for the consumer should have exceeded the gains from search, i.e., $c_i > G(\hat{u}_{it})$. The lower bound on the search cost for such a consumer, denoted by $\underline{c}_i$, is therefore

$$\underline{c}_i = G(\hat{u}_{it}). \tag{5}$$

The upper bound for the search cost is not identified for those consumers that only sample one firm. For those consumers who sampled multiple firms, the search cost upper bound, denoted by $\overline{c}_i$, corresponds to the gains from search at the best offer observed before the last search. Since the consumer found it optimal to sample more than once, the gains from search in the last period, $t-1$, were higher than her search cost, hence $c_i \leq G(\hat{u}_{it-1})$. Therefore the search cost upper bound for consumers sampling more than once is

$$\overline{c}_i = G(\hat{u}_{it-1}). \tag{6}$$

For each consumer in our sample, we use equation (5) to obtain a lower bound on the consumer's search cost, conditional on utility parameters. In addition we can use equation (6) to obtain an upper bound for those consumers that searched more than once. Since the gains from search will depend on actual utility values that have been observed while visiting stores, we simulate consumers by drawing from the distribution of ε_{ijk} conditional on the parameters of the utility function in equation (3). The gains from search are obtained using equation (4) and the highest simulated utility draw at time t.

In order to derive the probability that a consumer's search cost is within the observed search cost bounds, we assume search costs follow a log-normal distribution, and relate a consumer's search cost c_i to both demographic-related covariates as well as a stochastic term in the following way:

$$\ln c_i = \mu Z_i + \eta_i, \tag{7}$$

where Z_i is a vector of consumer demographics and η_i is a standard normally distributed error term. The probability that consumer i's search cost is within the relevant search cost bounds is then given by

$$P\big(\underline{c}_i < c_i \le \overline{c}_i\big) = \Phi\big(\ln \overline{c}_i - \mu Z_i\big) - \Phi\big(\ln \underline{c}_i - \mu Z_i\big). \tag{8}$$

where $\Phi(\cdot)$ is the standard normal CDF. Note that if a consumer searches once, the search cost upper bound is not identified and we set $\Phi\big(\ln \overline{c}_i - \mu Z_i\big) = 1$.

Although the model can be estimated using only equation (8), in our application we only observe search behavior at the domain level, which makes it difficult to estimate product-specific utility parameters with much precision. We include in the likelihood function the probability that the chosen product is the preferred product among the set of products sold by the retailer the consumer buys from. Taking both the stopping decision and the conditional product choice decision into account, the likelihood function is

$$\prod_i \int P\big(\underline{c}_i < c_i \le \overline{c}_i\big) \cdot P\big(u_{i\ell} \ge u_{ij} \quad \forall\ j \ne \ell \in \mathcal{J}_k\big) f(\varepsilon) d\varepsilon, \tag{9}$$

where subscript ℓ denotes the purchased product and $\mathcal{J}_k$ denotes the set of products sold by retailer k. Note that consumers are more likely to stop searching and buy if a product is observed with a relatively high ε draw, which means that the distribution of ε is not Type I EV, even conditional on a specific seller. We therefore use a logit-smoothed AR simulator to smooth the conditional product choice decision for a given set of ε draws. We estimate the parameters of the search cost distribution and utility function using simulated maximum likelihood. Specifically, for a given observation in the data we simulate a large number of consumers (each represented by a different

ε draw). We obtain search cost bounds for each of these simulated consumers by calculating the gains from search at the highest simulated utility draw among the retailers visited up to time t and $t-1$ and use this to calculate the probability that the consumer represented by the observation has search costs within these simulated bounds. We next multiply this by the probability that the purchased product has the highest simulated utility among all products sold by the retailer the consumer is buying from and take averages across all simulated consumers to obtain the likelihood contribution for the observation.

A special case of the model is when there is no vertical or horizontal product differentiation, so the utility function simplifies to $u_{ijk} = u_{jk} = -p_{jk}$. In this case, since prices are observed by the econometrician, the search cost bounds used in equation (8) can be calculated directly from price data.

An important question is to what extent consumers are behaving as if they know the distribution of utilities. Intuitively, the weight on the initial prior may be identified from recall patterns in the search data: as discussed in Section 2.3, recall is a result of a declining reservation utility, so the rate of "decline" should reflect how the prior departs from the sampled utility distribution. To investigate to what extent the weight on the initial prior can be estimated in practice, we parameterize the weight on the initial prior as $W = \omega JK$, where JK corresponds to the number of product-retailer combinations and ω is a parameter to be estimated. Inspection of the likelihood function reveals that even though this parameter should be identified due the parametric form of the gains from search equations, identification is likely to be weak. More specifically, it will be difficult to separately identify ω from the search cost constant. To see this, when taking logs of equation (2), ω appears twice: by itself and as $\omega JK + t$. Notice that only the second term will allow us to identify ω (the first term will be collinear with the search cost constant).

3 Application

3.1 Data

The dataset was constructed from the 2007 Comscore Web-Behavior Panel, which includes detailed online browsing and transaction data from 91,689 users. The users in the sample are randomly chosen from a universe of 1.5 million global users. Comscore is a leading provider of information on consumers' online behavior and supplies Fortune 500 companies and large news organizations with market research on e-commerce sales trends, website traffic, and online advertising campaigns. Each user's online activity is channeled through Comscore proxy servers that record all Internet traffic, including information on visits to a website or domain (browsing), as well as secure online transactions. The servers only capture web-browsing on one computer per household and therefore may not capture browsing and transactions at work. The data include date, time, and duration of website visits, as well as price, quantity, and description of each product purchased during a session. Unfortunately, the data does not allow us to observe website visits at the URL (sub-domain) level. This means we observe a consumer visiting a specific domain, but do not observe the specific webpages visited within a domain.

The dataset is for 2007 and contains 731 transactions that led to sales of 10 different MP3 players. Table 1 summarizes the transactions of the products in the sample. There are three different brands of MP3 players in our sample: Apple iPod, Microsoft Zune, and Sandisk Sansa. The iPod Nano 4Gb is with 192 transactions the most popular product. Prices range from a $28 Sandisk Sansa Shaker MP3 player to up to $349 for the iPod 80Gb, which is the most expensive item in our sample. The concentration ratio of the largest retailer in terms of sales (CR1) ranges from 33 to 74 percent of the market.

Table 1: Transaction characteristics by product

	Obs.	Number	CR1	Transaction price			
Product		of firms		Mean	Std. Dev.	Min	Max
iPod Nano 4Gb	192	9	0.71	170.05	25.43	129.00	249.99
iPod Shuffle 1Gb	147	10	0.60	78.28	2.69	58.95	129.00
iPod Nano 8Gb	110	9	0.74	203.08	18.06	179.94	249.00
iPod 80Gb	65	8	0.68	282.23	48.66	229.00	349.99
iPod 30Gb	60	6	0.65	242.74	13.62	184.99	299.99
iPod Nano 2Gb	47	5	0.45	141.98	13.78	99.99	199.99
Zune 30Gb	46	8	0.42	174.94	52.80	89.99	250.00
iPod Touch 8Gb	43	5	0.73	296.64	5.93	276.99	299.99
Sandisk Sansa Shaker MP3	13	4	0.37	30.87	3.47	28.28	39.99
Sandisk Sansa E250 MP3	8	5	0.33	73.83	23.78	38.99	139.98
Total	731	8	0.74	174.93	73.94	137.68	233.29

In order to identify a user's visit to a website as search behavior related to a particular transaction, we link the browsing history up to seven days before a transaction. Whereas there is no evidence to guide the definition of a search time span in relation to a transaction, one week is long enough to capture all search behavior related to a transaction; any longer intervals are likely to also capture unrelated website visits. A search history could be less than seven days if another transaction has occurred within seven days. The browsing activity of users consists of 9,742 visits to the retailers in the sample. Although some user search activity may not be linked to the transaction that immediately follows, but to a subsequent one, there is no clear way to link this intervening search to a later transaction. Another limitation of the Comscore data is that although we observe consumer visits to different online stores, we only observe the price of the transaction. To recover missing prices for all visited retailers we use the most recent transaction prices at those retailers with missing values. The average difference between a transaction and the transaction from which the price was imputed is 16 days. For a popular product like the iPod Nano 4Gb, the average difference is as low as less than one day, but for a less popular product like the Sandisk Sansa E250 the average difference is 31 days.

Table 2 presents descriptive statistics of the search patterns for a particular transaction per product. Overall, consumers visit on average 2.82 retailers before buying an MP3 player. Price dispersion of the searched firms is small for most products, ranging from 1 to 11 percent. The iPod 80Gb is the product with the most dispersed prices, as measured by the coefficient of variation. The average price difference between the transaction price and the lowest price across all transactions and stores is \$37.25. Most consumers will not visit all stores and encounter the full range of prices. The difference between the transaction price and the lowest price of the stores visited during the customer's search is on average \$30.05.

Table 2: Descriptive characteristics of retailers searched by product

	Mean number of	Prices of searched firms		
Product	firms searched	Mean	Std. Dev.	CV
iPod Nano 4Gb	2.79	161.81	11.29	0.07
iPod Shuffle 1Gb	2.62	75.86	0.62	0.01
iPod Nano 8Gb	2.88	194.29	3.36	0.02
iPod 80Gb	2.91	262.30	28.21	0.11
iPod 30Gb	2.60	244.78	2.68	0.01
iPod Nano 2Gb	2.47	134.96	6.49	0.05
Zune 30Gb	3.46	168.25	13.84	0.08
iPod Touch 8Gb	3.23	294.82	3.39	0.01
Sandisk Sansa Shaker MP3	4.15	31.60	0.49	0.02
Sandisk Sansa E250 MP3	1.75	86.57	3.18	0.04
Total	2.82	168.52	68.13	0.40

As shown in Table 3, when counting all electronic store visits in the seven days prior to a purchase as searches, close to fifty percent of consumers recalled a previously visited firm conditional on having searched at least twice. We have conditioned our recall statistic on "awareness," by not including recalls for consumers that visited all stores they are aware of (as measured by having previously visited the store)—as noted by De los Santos, Hortaçsu, and Wildenbeest (2012), the standard sequential search model predicts a consumer always buys from the last store she visited, unless she has visited all stores she is aware of. Since almost seventy percent of consumers searched at least twice, this means that in almost a third of transactions consumers recalled a previously visited firm. Although some of these recalls could be an artifact of the relatively large search window of seven days, as shown in Table 3, most recalls remain if the search window is made smaller. As argued in the Section 2, these non-trivial amounts of recall violate optimal behavior in the standard sequential search model, but can be rationalized by a learning model and provide an additional motivation for estimating a learning model.

Table 3: Consumer recall patterns by search window

	Two or more stores visited (as percentage of total)			Recalled (as percentage of those having visited two or more stores)		
Product	One-day	Three-day	Seven-day	One-day	Three-day	Seven-day
iPod Nano 4Gb	42.7	56.3	69.8	35.4	38.9	44.8
iPod Shuffle 1Gb	42.2	51.0	61.2	14.5	34.7	42.2
iPod Nano 8Gb	49.1	60.0	74.5	24.1	48.5	46.3
iPod 80Gb	43.1	55.4	70.8	39.3	44.4	45.7
iPod 30Gb	50.0	56.7	63.3	33.3	35.3	50.0
iPod Nano 2Gb	40.4	53.2	70.2	10.5	20.0	24.2
Zune 30Gb	56.5	63.0	76.1	42.3	48.3	51.4
iPod Touch 8Gb	39.5	60.5	72.1	64.7	57.7	74.2
Sandisk Sansa Shaker MP3	61.5	84.6	84.6	37.5	54.5	54.5
Sandisk Sansa E250 MP3	12.5	25.0	25.0	100.0	100.0	100.0
Total	44.7	56.4	68.7	30.6	41.3	46.4

Notice that recall may also be rationalized by a more general sequential search model in which consumers search in a directed way or by a fixed sample size search model. De los Santos, Hortaçsu, and Wildenbeest (2012) look at several testable predictions of the sequential search model versus the fixed sample size search model and argue that a robust prediction of sequential search (with and without learning) is that the decision to continue searching depends upon the outcome of the previous search, while it does not with fixed sample size search. Although this is relatively straightforward to test for in a homogenous product setting, testing between our learning model and a fixed sample size search model is more complicated in a differentiated product setting, especially since we allow for multi-product search. De los Santos, Hortaçsu, and Wildenbeest (2012) develop a test for sequential search versus fixed sample size search that accounts for differentiation across stores. However, their test cannot be directly applied to this model as consumers are also searching for a right product match.

Firms and Consumers

Table 4 presents market shares, the number of consumer search visits, and general information for the main online retailers in the sample. The products included in the sample were bought from 18 different retailers. Given that there are few transactions for MP3 players for several retailers we show the largest retailers in terms of MP3 players and group smaller retailers into two categories. The market for these goods is less concentrated than many other online markets. Given the inclusion of popular Apple products in the sample, Apple has the largest number of transactions, followed by Amazon. The presence of other large retailers in our sample, such as Walmart, BestBuy, and Circuit City, means Amazon is less dominant in the market for electronics than it is for books: even if we exclude Apple products from the sample, Amazon's market share for the remaining products is still significantly lower than Amazon's share of the book market (which is 66 percent according to De los Santos, Hortaçsu, and Wildenbeest, 2012).

Table 4: Firms' market shares and shares of search visits

Firms	Market Share %	Search visits %
apple.com	59.10	13.56
amazon.com	16.69	19.91
circuitcity.com	9.71	9.80
target.com	3.15	9.10
bestbuy.com	2.33	9.96
walmart.com	2.33	12.55
overstock.com	1.50	4.54
Other electronic stores	3.69	9.77
Other retailers	1.50	10.80
Observations	731	9,742

Notes: Other electronic stores include bhphotovideo.com, compusa.com, dell.com, macmall.com, newegg.com, and tigerdirect.com. Other Retailers include buy.com, costco.com, officedepot.com, sears.com, and staples.com.

Table 5 presents household characteristics in the sample. The overall Comscore sample is representative of online buyers in the U.S. in terms of age, education, income, household composition, and other observable characteristics. De los Santos (2010) shows that the Comscore sample is representative of the U.S. by comparing it with users that have bought a product online from the sample with the Internet and Computer Use Supplement of the Current Population Survey (CPS) and the Forrester Technographics Survey. The main differences between the CPS and Comscore samples are that in the Comscore sample, Internet users are older, have higher income, and are more likely to have some college but no degree. The racial composition is similar across samples—online users are predominantly white. However, compared with CPS, Comscore over samples Hispanics and Forrester over samples whites. The geographic distribution of users is similar to CPS population estimates at the regional and state levels. As this data is composed of electronics consumers, the sample has a larger fraction of consumers with high income. Consumers with income above $75,000 represent 41 percent of the sample compared to 14 percent of the U.S. population that has bought a product online.

Table 5: Descriptive statistics of consumer characteristics

	Mean	Std. Dev.
Broadband connection	0.95	0.21
Household size	3.31	1.35
Children present	0.75	0.43
Age		
18-20	0.00	0.06
21-24	0.02	0.15
25-29	0.04	0.20
30-34	0.09	0.29
35-39	0.13	0.34
40-44	0.18	0.39
45-49	0.18	0.39
50-54	0.14	0.35
55-59	0.07	0.25
60-64	0.06	0.23
65 and over	0.08	0.28
Household income		
Less than $15,000	0.09	0.29
$15,000 - $25,000	0.05	0.22
$25,000 - $35,000	0.08	0.28
$35,000 - $50,000	0.13	0.33
$50,000 - $75,000	0.23	0.42
$75,000 - $100,000	0.16	0.37
More than $100,000	0.25	0.44
Race		
White	0.94	0.24
Black	0.05	0.21
Hispanic	0.01	0.10
Other	0.00	0.04

Notes: Number of consumers is 637.

3.2 Estimation results

We estimate the model using transactions for MP3 players as well as the corresponding search history before a transaction. In all our empirical specifications we allow for retailer fixed effects. In addition, we include several product attributes in the utility function—besides price we include storage per weight and brand fixed effects. Our learning model assumes that the retailer fixed effects as well as the actual values of the attributes of the MP3 players (including the utility shock) are unknown. Consumers have to search for these unknown attributes. Once a retailer is visited then the values of the attributes are known and the utility shock is realized. Notice that the domain-level browsing data we use does not allow us to directly observe which characteristics of the MP3 players consumers are searching for. However, Bronnenberg, Kim, and Mela (2015) observe website visits at the URL-level for consumers shopping online for digital cameras, which allows them to better analyze the characteristics consumers are searching for when shopping online. One of their

Table 6: Estimation results

	Learning					
	(1) No Control Function		(2) Control Function		(3) No Learning	
Variable	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Search Cost						
Constant	-1.086	(0.186)***	-1.103	(0.169)***	-0.777	(0.154)***
Broadband	-0.171	(0.166)	-0.183	(0.156)	-0.133	(0.132)
Age 60+	-0.086	(0.118)	-0.076	(0.108)	-0.054	(0.099)
Income >75k	-0.134	(0.085)	-0.132	(0.080)*	-0.084	(0.072)
Household size	-0.005	(0.027)	0.002	(0.026)	-0.003	(0.024)
Retailer Fixed Effects						
Apple Store	1.121	(0.201)***	1.337	(0.189)***	0.914	(0.146)***
Circuit City	-0.156	(0.219)	-0.061	(0.231)	-0.123	(0.164)
Overstock	-1.128	(2.945)	-1.201	(1.221)	-0.458	(0.273)*
Target	-0.929	(0.290)***	-0.905	(0.354)**	-0.620	(0.199)***
Walmart	-1.040	(0.403)**	-0.940	(0.382)**	-0.294	(0.168)*
Other electronics	-0.045	(0.314)	-0.137	(0.339)	0.126	(0.220)
Other general merchandise	-1.145	(0.488)**	-1.165	(0.466)**	-0.638	(0.241)***
Product Attributes						
Zune	0.207	(0.194)	0.142	(0.198)	-0.173	(0.190)
Sansa	-1.418	(0.313)***	-1.960	(0.285)***	-1.383	(0.349)***
Storage/weight	0.012	(0.017)	0.066	(0.015)***	0.061	(0.017)***
Price	-0.400	(0.090)***	-0.942	(0.093)***	-0.981	(0.093)***
Control Function						
Unobserved attributes			0.915	(0.150)***	0.891	(0.139)***
Log-likelihood	-2,655.05		-2,623.01		-2,763.86	

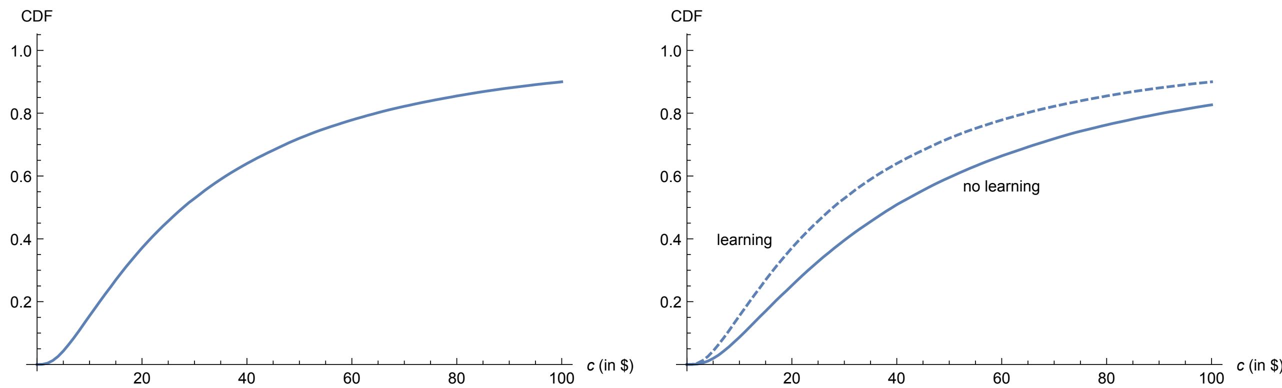
Notes: *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Bootstrapped standard errors within parentheses. The number of observations is 731. The number of simulated consumers is 200 per observation. The value of the smoothing parameter of the logit-smoothed AR simulator is 0.10. Weight on the initial prior is $W = 62$. Price is measured in hundreds of dollars.

main findings is that consumers search more extensively than commonly reported. Moreover, their findings imply that consumers search for more attributes than price, which offers some justification for our approach.

Table 6 presents the results for various specifications of model. Following Koulayev (2013), we set the weight on the initial prior equal to the total number of product-retailer combinations ($W = 62$). Unobserved product and retailer characteristics may be correlated with prices, which, if not corrected for, may lead to biased estimates. We deal with this in two ways; firstly, all our specifications include retailer and brand fixed effects, which will accommodate most of the correlation between prices and unobserved characteristics that is retailer and brand specific; and secondly, in columns (2) and (3) we use a control function approach (Petrin and Train, 2010) to correct for any remaining endogeneity bias. The approach consists of two stages. In the first stage we regress prices on all exogenous variables as well as instruments. In a second stage we estimate our model adding the residuals of the first-stage regression as an additional control. We use two instruments for the control function, which are inspired by the instruments used in Berry, Levinsohn, and Pakes (1995). As each observation reflects the sale of a particular MP3 player at a particular retailer, one instrument is the number of retailers that sell the MP3 player and the second is the total number of other MP3 players sold at all other retailers. A comparison of the log-likelihood values of the specification in column (1), which does not use the control function, and column (2) indicates that accounting for price endogeneity using the control function approach results in a slightly better fit. Both instruments are significantly different from zero at the one-percent level in the first-stage regression (not reported), and the F statistic for joint exclusion of the two instruments from the first-stage regression is 11,056. Moreover, the parameter that reflects unobserved heterogeneity in the second-stage regression is significantly different from zero, which suggests price endogeneity is indeed an issue in the data. Eight parameter estimates increase in magnitude when using the control function approach, including the price coefficient.

The Sansa brand dummy is negative and highly significant, which indicates that the marginal utility of Sansa MP3 players is lower in comparison to the base group, which consists of Apple branded MP3 players. The Zune dummy is not significantly different from zero. Storage per weight has a positive marginal utility and is highly significant when controlling for price endogeneity, which reflects that consumers place a premium on smaller MP3 players that can still carry a large number of songs. The estimated retailer fixed effects show that consumer preferences for stores are not uniform—although most estimated retailer dummies reported in Table 6 indicate a lower marginal utility in comparison to Amazon, the positive and significant Apple Store dummy suggests that after controlling for differences in pricing, the Apple Store is the most preferred online store when shopping for MP3 players.

Except for the coefficient on household size, all estimated search cost parameters have a negative sign, although only the constant is significantly different from zero at the 1 percent level. Figure 3(a) plots the estimated search cost CDF in dollar values using the estimates in column (2) of Table 6. Search costs are sizable: median search costs are $27.86 and 25 percent of households have search costs that exceed $54.63. Although these figures seem relatively large, recall that in our setting consumers retrieve information on all MP3 players sold by a retailer (on average 7 different products) during a search, which means that search costs per sampled product are substantially lower. Also note that in our model there is not only uncertainty about prices but also about all other characteristics of the products, including the matching term, which will increase the time needed to process all information necessary to make search and purchase decisions. As a result, the potentially large gains from search in combination with the relatively low number of searches we observe in the data (on average 2.82) can only be rationalized by relatively large search cost estimates. Indeed, according to our estimates the median difference (in dollars) between the maximum observed utility at time t and the overall maximum utility is \$33.77. Even though around 43 percent of the simulated consumers do manage to find the retailer-product combination that generates the highest utility, for the consumers that did not succeed in finding the highest utility alternative the median difference is \$150, which indicates that search frictions have substantial implications for the majority of consumers in this market.



(a) Learning

(b) Model comparison

Figure 3: Estimated search costs

In our final specification in Table 6, we fit a standard sequential search model to the data by assuming there is no learning. This is equivalent to setting $t = 0$ in equation (2) and means consumers know the utility distribution with certainty. As shown in the last column of Table 6 the model does worse in terms of fitting the data. Part of this is explained by the no-learning model being unable to account for recall patterns. The estimates are also quite different: according to the learning model median search costs are \$27.86, while in the no-learning model median search costs are with \$39.15 more than 40 percent higher. The difference in search cost estimates can also be seen in Figure 3(b), which plots the search cost CDFs for both specifications. As we showed in Section 2.3, reservation utilities in the learning model are decreasing in the number of alternatives sampled. This means that everything else equal the no-learning specification requires higher search costs to rationalize the observed search patterns, which explains the difference in search cost estimates.

Robustness

Table 7 gives parameter estimates for various alternative specifications. As mentioned before, our data only allows us to observe visits to retailers at the domain level—for all estimates so far we have linked visits to a retailer for up to seven days prior to a MP3-player transaction as being related to the purchase of the MP3-player. To analyze how sensitive our estimates are to assumptions about the length of this window, we have also estimated the model assuming only visits to a retailer in the three days prior to a purchase are related. The results assuming this different window are shown in column (1) of Table 7. Except for the search cost constant, which is higher for the three-day window, all parameter estimates are very similar to the results when assuming a seven-day window, as reported in column (2) of Table 6. The difference in search cost estimates is intuitive: a lower number of retailers is visited when assuming a smaller window, which can only be rationalized by having higher search cost estimates.

Table 7: Alternative specifications

	(1) Three-Day Window		(2) Uniform Prior		(3) $W = 31$	
Variable	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Search Cost						
Constant	-0.663	(0.199)***	-0.699	(0.310)**	-1.275	(0.170)***
Broadband	-0.230	(0.179)	-0.273	(0.252)	-0.220	(0.163)
Age 60+	-0.168	(0.117)	-0.131	(0.263)	-0.060	(0.120)
Income >75k	-0.120	(0.103)	-0.477	(0.228)**	-0.142	(0.074)*
Household size	-0.022	(0.030)	-0.001	(0.057)	-0.002	(0.030)
Retailer Fixed Effects						
Apple Store	1.467	(0.216)***	1.173	(0.178)***	1.522	(0.198)***
Circuit City	0.055	(0.253)	-0.277	(0.177)	0.014	(0.214)
Overstock	-1.009	(0.439)**	-1.535	(0.342)***	-1.322	(0.523)**
Target	-0.642	(0.293)**	-0.759	(0.195)***	-1.222	(0.427)***
Walmart	-1.745	(0.409)***	-0.367	(0.251)	-1.223	(0.417)***
Other electronics	0.153	(0.380)	-0.547	(0.258)**	-0.306	(0.491)
Other general merchandise	-1.419	(0.581)**	-1.013	(0.275)***	-1.384	(0.431)***
Product Attributes						
Zune	0.258	(0.180)	0.083	(0.278)	0.178	(0.193)
Sansa	-2.138	(0.243)***	-1.717	(0.354)***	-2.075	(0.266)***
Storage/weight	0.065	(0.016)***	0.058	(0.024)**	0.071	(0.015)***
Price	-0.971	(0.118)***	-0.938	(0.120)***	-0.982	(0.103)***
Control Function						
Unobserved attributes	0.949	(0.177)***	0.916	(0.137)***	0.979	(0.151)***
Log-likelihood	-2,405.92		-2,601.56		-2,574.47	

Notes: *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Bootstrapped standard errors within parentheses. The number of observations is 731. The number of simulated consumers is 200 per observation. The value of the smoothing parameter of the logit-smoothed AR simulator is 0.10. Price is measured in hundreds of dollars.

In the second specification in Table 7 we use a uniform distribution instead of Type I EV distribution as the initial prior. If we assume consumers have uniform priors with support $[\underline{u}, \overline{u}]$, the gains from search equation is

$$G(\hat{u}_{it}) = \frac{W}{W + t \cdot J} \sum_{k=1}^{K} \frac{1}{K} \left(\frac{1}{J+1} \left(\underline{u} + \overline{u}J - (J+1)\hat{u}_{it} + (\hat{u}_{it} - \underline{u}) \left(\frac{\hat{u}_{it} - \underline{u}}{\overline{u} - \underline{u}} \right)^{J} \right) \right). \qquad (10)$$

Here we assume the lower and upper bound of the uniform are known to the consumer and correspond to the lowest and highest utility in the market. Estimated search costs are higher than when assuming consumers have true beliefs about the initial prior. Since a uniform distribution puts equal weight to all utility values in excess of the highest utility observed so far, the gains from search are uniformly higher, which means search costs need to be higher in order to rationalize observed search patterns.

In the third specification in Table 7 we set $W = 31$, which means the weight on the initial prior is half of what we assume in our main specification. The estimated search cost constant is slightly lower; the rest of the estimated parameters are very similar to those reported in column (2) of Table 6. Non-reported estimates assuming $W = 93$ are also similar to the main specification.

We use a log-normal distribution as the parametric form of the search cost distribution. In order to analyze the sensitiveness of the estimates to the log-normal distribution assumption we estimate the model using a semi-nonparametric (SNP) approach. The search cost density distribution is based on flexible Hermite polynomial functions that approximate arbitrarily closely a large class of sufficiently smooth density functions (Gallant and Nychka, 1987). We use an SNP estimator that is equivalent to that in Moraga-González, Sándor, and Wildenbeest (2013), except that we constrain the standard deviation parameter of the associated normal distribution to be 1. Although this approach allows for more flexibility, the estimated search cost distribution is very similar to the log-normal distribution.

4 Monte Carlo Experiments

In this section we investigate the performance of our estimation procedure using a number of Monte Carlo experiments. Our first objective is to confirm that our estimation procedure is able to recover the unknown parameters of the search cost distribution and the utility function. Next, we study to what extent estimates will be biased if we do not take learning into account when estimating data generated by a learning model. Finally, we look at possible measurement error in the composition of choice sets, as well as measurement error in prices.

The setup of the experiments is as follows. We randomly generate 1,000 observations, where each observation corresponds to one household. An observation includes data on the search sequence prior to a transaction, as well as the product bought, its price, and the identity of the retailer. We simulate 5 different retailers, each selling 3 different products.

We assume consumers' search costs are drawn from a log-normal search cost distribution with the standard deviation of the associated normal distribution set to 1. We let the mean of the associated normal distribution depend on a constant as well as an indicator for whether the household has a broadband connection (randomly drawn from a Bernoulli distribution with $p = 0.3$).

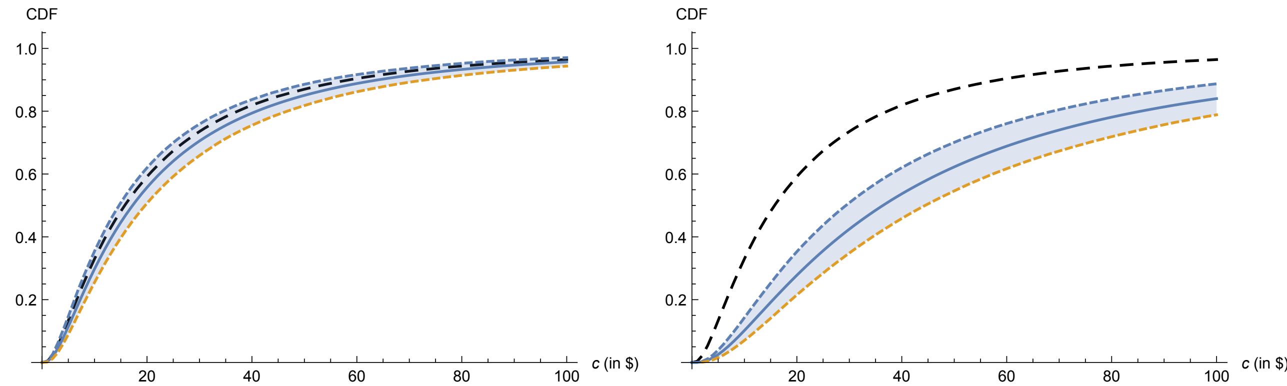
Household i's utility for product j sold at retailer k is given by equation (3). The stochastic term in the utility specification is randomly drawn from a standard Type I EV distribution and each household has a different draw for each product-retailer combination. Prices are randomly drawn from a uniform distribution with product-specific parameters.

Table 8: Monte Carlo simulations

	(1) True	(2) Learning		(3) No Learning	
Variable	Coeff.	Coeff.	Std. Dev.	Coeff.	Std. Dev.
Search Cost					
Constant	-1.000	-0.931	(0.052)	-0.279	(0.044)
Broadband	-0.500	-0.439	(0.084)	-0.278	(0.078)
Utility					
Firm 1	-2.000	-2.049	(0.179)	-1.773	(0.160)
Firm 2	-1.500	-1.553	(0.162)	-1.404	(0.154)
Firm 3	-1.000	-1.048	(0.152)	-0.972	(0.143)
Firm 4	-0.500	-0.526	(0.133)	-0.498	(0.126)
Product 2	-1.000	-0.984	(0.105)	-0.947	(0.120)
Product 3	1.000	1.009	(0.108)	0.979	(0.124)
Price	-2.000	-2.003	(0.193)	-1.918	(0.222)

Notes: Number of observations is 1,000. The number of simulated consumers is 200 per observation. Weight on the initial prior $W = 15$. Simulated prices for product 1 are uniform $U(100, 175)$, prices for product 2 are uniform $U(75, 125)$, and prices for product 3 are uniform $U(125, 225)$. Price is measured in hundreds of dollars.

Table 8 presents the results of the Monte Carlo simulations. Column (1) provides the search cost and utility parameters used to generate the data. For the simulations, we assume that consumers have correct beliefs about the prior, corresponding to the true joint utility distribution, which means that consumers use equation (4) to calculate the gains from search. As in the application, we set the weight on the initial prior equal to the total number of product-retailer combinations, so $W = 15$. Column (2) presents the mean and standard deviation of the parameter estimates of the main learning specification across 100 replications. The estimates are relatively close to the true parameter values and have low standard deviations. This can also be seen in Figure 4(a), which plots the actual search cost CDF (black dashed curve) as well as the estimated CDF (solid curve) and corresponding 90 percent confidence interval for the main specification.



(a) Learning specification

(b) No-learning specification

Figure 4: Estimated search costs

Column (3) of Table 8 presents parameter estimates for the standard sequential search model (no-learning). Although estimation of the price parameter and product dummies appear unaffected by this, the search cost parameters are biased upward and both parameters are no longer within two standard deviations of their true values. Figure 4(b) plots the true search cost CDF (black dashed curve) as well as the estimated CDF (solid curve) and corresponding 90 percent confidence interval for the no-learning model, and shows that search cost estimates will be severely biased when the data is generated from a learning model and learning is unaccounted for in the estimation. As we argued in Section 2.3, reservation utilities in the learning model are decreasing in the number of alternatives sampled. Since we are assuming a correct initial prior, the distribution will be the same in both learning and no-learning models. In this case consumers will be searching less in the learning model (which is used to generate the data) than they would in the no-learning model. This means that the no-learning model can only rationalize the observed search patterns in the Monte Carlo experiment by having higher search costs. (Koulayev, 2013) finds the direction of the bias to be similar for his learning model.

Elasticity estimates reflect the bias in search costs when consumer learning is ignored. Table 9 presents the true own-price elasticities for each retailer as well as the own-price elasticities based on the estimates from the no-learning model. The elasticities are calculated by simulation. For the learning model, we simulate the percentage change in demand as a result of a 10 percent increase in price for 100,000 consumers, using the true modeling parameters. We use a similar approach for the no-learning model, using the estimated parameters in column (3) of Table 8. The table shows that elasticity estimates are biased towards zero for all retailers, and that the bias is most severe for the firms with the lowest market shares.

Table 9: Own-price elasticity estimates

	Learning	No Learning
Firm 1	-1.735	-1.385
Firm 2	-1.631	-1.343
Firm 3	-1.474	-1.241
Firm 4	-1.271	-1.104
Firm 5	-1.022	-0.932

Notes: Firms are ordered by increasing market shares.

Robustness

In our application as well as in most online settings, visits to retailers are observed at the domain level, which makes it difficult to infer as to whether a particular product was viewed by the consumer during her visit. The common strategy is to treat prior visits to retailers—during a set search window—as being related to a subsequent purchase. This means we may wrongfully attribute a visit to a retailer as a consumer sampling a particular product sold by this retailer. On the other hand, consumers may get some price and product information from non-retailer websites such as price comparison sites, which means we may fail to account for some retailers in consumers' choice sets. To see to what extent our estimates are affected by the potential error that could arise as a result of these assumptions, we add noise to the sets of searched retailers in the data used for estimating the learning model. We let 50 percent of the observations be affected by this noise—to half of these observations we add a retailer at a random position in the search sequence, while for the other half we take out a randomly selected retailer from the sequence. We only delete a randomly selected retailer if the consumer has searched more than one firm and if this selected retailer is not the seller. The estimation results are shown in column (2) of Table 10. Although the mean parameter estimates are somewhat affected by the noise, standard deviations do not change that much in comparison to those in column (2) of Table 8 and all parameter estimates are within two standard deviations of the true parameter values.

Another potential concern is that in the actual application, prices may be measured with error. Prices are typically obtained from transactions, and hence we infer prices at the other retailers from transaction prices of other consumers. Since prices may change over time, this means there is potential for measurement error in prices. To study how this will affect the estimation of the learning model, we add noise to the simulated prices by allowing for a multiplicative noise term that is drawn from a log-normal distribution with an associated mean-zero normal distribution with 0.1 standard deviation. The estimation results shown in column (3) of Table 10 indicate that the

Table 10: Monte Carlo simulations (robustness)

Variable	(1) True Coeff.	(2) Noisy Choice Set Coeff.	(2) Noisy Choice Set Std. Dev.	(3) Noisy Prices Coeff.	(3) Noisy Prices Std. Dev.
Search Cost					
Constant	-1.000	-1.101	(0.049)	-0.942	(0.055)
Broadband	-0.500	-0.368	(0.083)	-0.425	(0.093)
Utility					
Firm 1	-2.000	-1.781	(0.185)	-2.035	(0.173)
Firm 2	-1.500	-1.341	(0.153)	-1.520	(0.190)
Firm 3	-1.000	-0.903	(0.155)	-1.006	(0.144)
Firm 4	-0.500	-0.444	(0.126)	-0.493	(0.130)
Product 2	-1.000	-0.967	(0.108)	-0.795	(0.110)
Product 3	1.000	0.992	(0.107)	0.800	(0.098)
Price	-2.000	-1.960	(0.195)	-1.436	(0.152)

Notes: Number of observations is 1,000. The number of simulated consumers is 200 per observation. Weight on the initial prior $W = 15$. Price is measured in hundreds of dollars. In column (3), the noise term has a mean of 1.005 and a variance of 0.010, which implies that 95 percent of draws is between 0.8 and 1.2.

noise does not affect the estimation of search cost parameters. However, the price noise does affect some of the utility parameters: even though the firm dummies appear unaffected, there is a bias in the estimation of the product dummies and the price coefficient. The direction of the bias is as expected: the effect of prices on search and purchase decisions is less pronounced when prices are measured with error, which leads to a price coefficient that is biased towards zero.

5 Conclusions

In this paper we have presented a methodology to estimate a model of search for differentiated products with learning. The distribution of utilities is assumed to be unknown to consumers and is learned in the search process by Bayesian updating Dirichlet process priors. We have shown how to use information on the sequence of searches as well as prices to derive expressions for the bounds on a consumer's search costs. We relate search costs to the characteristics of households in our sample and find search costs to be sizable: our estimates indicate that median search costs are close to \$28. Estimated search costs are uniformly lower than in a sequential search model with a known distribution of utilities. Moreover, the learning model gives a better fit to the data than the model in which there is no updating.

Despite the richness of the browsing and transaction data we have used in this paper, the data does have some weaknesses. For instance, we only observe transaction prices, which means that we have to use recent transactions to impute prices on days when no transaction was made. Also, we only observe website visits at the domain level, which means we do not know with certainty what consumers were searching for when visiting a website. Finally, we do not know what knowledge consumers have about products and retailers. Some of these limitations can be dealt with by having more detailed data. In a recent paper, Bronnenberg, Kim, and Mela (2015) observe browsing at a much more detailed level, including sites visited at the URL level. With richer data like this, it is likely that some of the more restrictive assumptions we have made can be relaxed.

Another limitation of our study is that we have to take the learning process as given. Although the Monte Carlo experiments have shown that our estimation method is able to recover the parameters of both the search cost and utility distributions, to be able to do so, we have to take the initial prior as given, including the weight consumers put on it. We have shown that to estimate our model, only the number of searches and information from the last search are needed: the search cost bounds in our model are derived from the simulated gains from search when actually purchasing the product and the period before that. We are hopeful that using the entire sequence of searches will help in identifying the importance of learning in a product differentiation setting, and leave this for future research.

References

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: An empirical examination. *International Economic Review 44*, 1007–1040.

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

Bikhchandani, S. and S. Sharma (1996). Optimal search with learning. *Journal of Economic Dynamics and Control 20*, 333–359.

Blackwell, D. and J. B. MacQueen (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics 1*, 353–355.

Bronnenberg, B. J., J. B. Kim, and C. F. Mela (2015). Zooming in on choice: How do consumers search for cameras online? Forthcoming in *Marketing Science*.

Chernew, M., G. Gowrisankaran, and D. P. Scanlon (2008). Learning and the value of information: Evidence from health plan report cards. *Journal of Econometrics 144*, 156–174.

Chou, C. and G. Talmain (1993). Nonparametric search. *Journal of Economic Dynamics and Control 17*, 771–784.

Crawford, G. S. and M. Shum (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica 73*, 1137–1173.

Dana, J. D. (1994). Learning in an equilibrium search model. *International Economic Review 35*, 745–771.

De los Santos, B. (2010). Consumer search on the Internet. *Mimeo*.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review 102*, 2955–2980.

Erdem, T. and M. P. Keane (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science 15*, 1–20.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics 1*, 209–230.

Gallant, A. R. and D. W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica 55*, 363–390.

Gastwirth, J. L. (1976). On probabilistic models of consumer search for information. *Quarterly Journal of Economics*, 38–50.

Häubl, G., B. G. C. Dellaert, and B. Donkers (2010). Tunnel vision: Local behavioral influences on consumer decisions in product search. *Marketing Science 29*, 438–455.

Hong, H. and M. Shum (2006). Using price distributions to estimate search costs. *RAND Journal of Economics 37*, 257–275.

Honka, E. (2014). Quantifying search and switching costs in the U.S. auto insurance industry. *RAND Journal of Economics 45*, 847–884.

Hortaçsu, A. and C. Syverson (2004). 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.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010). Online demand under limited consumer search. *Marketing Science 29*, 1001–1023.

Koulayev, S. (2013). Search with Dirichlet priors: Estimation and implications for consumer demand. *Journal of Business & Economic Statistics 31*, 226–239.

Koulayev, S. (2014). Search for differentiated products: Identification and estimation. *RAND Journal of Economics 45*, 553–575.

McCall, J. J. (1970). Economics of information and job search. *Quarterly Journal of Economics 84*, 113–126.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2013). Semi-nonparametric estimation of consumer search costs. *Journal of Applied Econometrics 28*, 1205–1223.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2015). Consumer search and prices in the automobile market. *Mimeo*.

Moraga-González, J. L. and M. R. Wildenbeest (2008). Maximum likelihood estimation of search costs. *European Economic Review 52*, 820–848.

Morgan, P. and R. Manning (1985). Optimal search. *Econometrica 53*, 923–944.

Mortensen, D. T. (1970). Job search, the duration of unemployment and the Phillips curve. *American Economic Review 60*, 847–862.

Petrin, A. and K. Train (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research 47*, 3–13.

Rosenfield, D. B. and R. D. Shapiro (1981). Optimal adaptive price search. *Journal of Economic Theory 25*, 1–20.

Rothschild, M. (1974). Searching for the lowest price when the distribution of prices is unknown. *Journal of Political Economy 82*, 689–711.

Stigler, G. (1961). The economics of information. *Journal of Political Economy 69*, 213–225.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica 47*, 641–654.

Wildenbeest, M. R. (2011). An empirical model of search with vertically differentiated products. *RAND Journal of Economics 42*, 729–757.

Appendix

This Appendix provides additional details on the derivation of equation (4). In general, the expected maximum utility out of J draws from distribution $H(u)$ is given by

$$\begin{aligned}
E\left[\max_{j\in J}\{u_{ij}\}\right] &= \int_{-\infty}^{\infty} u\frac{d}{du}\left(\prod_{j\in J} H(u-\delta_j)\right)du; \\
&= \int u\sum_{j\in J}\left(h(u-\delta_j)\prod_{r\in J\backslash j} H(u-\delta_r)\right)du; \\
&= \int u\sum_{j\in J}\left(\exp[-((u-\delta_j)+\exp[-(u-\delta_j)])]\prod_{r\in J\backslash j}\exp\left[-\exp\left[-(u-\delta_r)\right]\right]\right)du; \\
&= \gamma+\log\left(\sum_{j\in J}\exp[\delta_j]\right).
\end{aligned}$$

Note that $E\left[\max_{j\in J}\{u_{ij}\}\right]$ is relevant when searching using a fixed sample size search strategy, for instance when the decision is whether to search J or $J+1$ times. In our setting the decision is whether to sample J alternatives (ignoring the retailer dimension) given that the utility at hand is $\hat{u}_{it}$. Just focusing on this, and ignoring the $W/(W+t)$ term, this means the gains from search equation is

$$G(\hat{u}_{it}) = \int_{\hat{u}_{it}}^{\infty}(u-\hat{u}_{it})\frac{d}{du}\left(\prod_{J} H(u-\delta_j)\right)du.$$

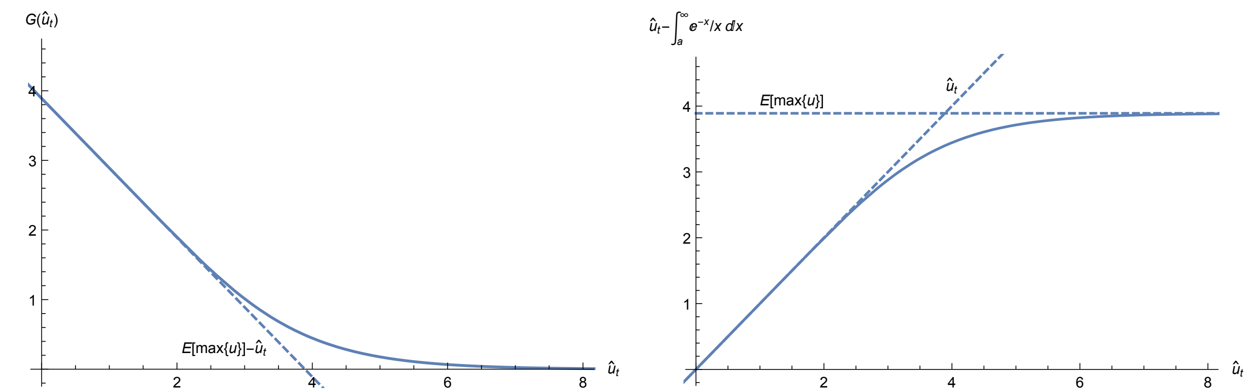
We can rewrite this integral as to sum of three separate integrals:

$$\begin{aligned}
G(\hat{u}_{it}) &= \int_{-\infty}^{\infty} u\frac{d}{du}\left(\prod_{J} H(u-\delta_j)\right)du - \int_{-\infty}^{\infty}\hat{u}_{it}\frac{d}{du}\left(\prod_{J} H(u-\delta_j)\right)du \\
&\quad - \int_{-\infty}^{\hat{u}_{it}}(u-\hat{u}_{it})\frac{d}{du}\left(\prod_{J} H(u-\delta_j)\right)du.
\end{aligned}$$

The first integral is the expected maximum utility out of J draws from $H(u)$, which, using the results from above, has the closed-form solution $\gamma+\log\left(\sum_{j\in J}\exp[\delta_j]\right)$. The second integral is just the reference utility $\hat{u}_{it}$. Finally, the last integral reflects a situation in which the maximum utility of J draws is less than $\hat{u}_{it}$, so the consumer will find it optimal to stick to $\hat{u}_{it}$. This integral does not have a closed-form solution, but can be described by the exponential integral function evaluated at $a=\sum_{j\in J}\exp\left[\delta_j-\hat{u}_{it}\right]$, i.e., the integral equals $-\int_a^{\infty} e^{-x}/x\ dx$. Therefore, the gains from search equation simplifies to

$$G(\hat{u}_{it}) = \gamma+\log\left(\sum_{J}\exp[\delta_j]\right)-\left(\hat{u}_{it}-\int_a^{\infty} e^{-x}/x\ dx\right). \tag{A1}$$

As an example, Figure A1(a) gives $G(\hat{u}_{it})$ as a function of $\hat{u}_{it}$ for $J = 2$ with $\delta_1 = 2$ and $\delta_2 = 3$. For relatively low values of $\hat{u}_{it}$, the gains from search correspond to $E[\max\{u\}] - \hat{u}_{it}$ (the dashed line in the figure), since the maximum utility of (in this case) two draws from the utility distribution is almost always larger than $\hat{u}_{it}$. For higher values of $\hat{u}_{it}$, the probability of finding a higher utility than $\hat{u}_{it}$ becomes smaller and smaller, and in the limit the gains from search are zero. The exponential integral in equation (A1) captures the option value of sticking to $\hat{u}_{it}$ in case the maximum utility drawn is less than $\hat{u}_{it}$. The difference between $\hat{u}_{it}$ and the exponential integral (the second term between brackets in equation (A1)) is plotted in Figure A1(b). The graph shows that while for relatively small values of $\hat{u}_{it}$ the exponential integral is close to zero, for larger values of $\hat{u}_{it}$ that term is substantial, and $\hat{u}_{it} - \int_a^\infty e^{-x}/x\ dx$ converges to the expected maximum utility.



(a) Gains from search $G(\hat{u}_{it})$

(b) Reference utility minus exponential integral

Figure A1: Gains from search for $\delta_1 = 2$ and $\delta_2 = 3$