Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior

By Babur De los Santos, Ali Hortaçsu, and Matthijs R. Wildenbeest[*]

FORTHCOMING IN THE *AMERICAN ECONOMIC REVIEW*

Using a large data set on web browsing and purchasing behavior we test to what extent consumers are searching in accordance to various classical search models. We find that the benchmark model of sequential search with an a priori known distribution of prices can be rejected based on both the recall patterns we observe in the data as well as the absence of dependence of search decisions on observed prices. Our findings suggest that fixed sample size search provides a more accurate description of observed consumer search behavior. We then utilize the fixed sample size search model to estimate demand elasticities of online book stores in an environment where consumers' store preferences are heterogeneous.
JEL: D43, D83, L13
Keywords: consumer search, electronic commerce, consumer behavior

Since Stigler's (1961) seminal paper, models of costly search have been at the heart of many economic models trying to explain imperfectly competitive behavior in product and labor markets. The theoretical literature typically models consumer search in two ways: following Stigler's original model, a strand of literature assumes *fixed sample size* search behavior, where consumers sample a fixed number of stores and choose to buy the lowest priced alternative.[1] A much larger strand of the literature, starting with McCall (1970) and Mortensen (1970), points out that consumers cannot commit to a fixed sample size search strategy in instances where the expected marginal benefit of an extra search exceeds the

[1]See also Burdett and Judd (1983) and Janssen and Moraga-González (2004).

marginal cost. Thus, this literature argues that a *sequential* search model provides a better description of actual consumer search.[2]

Unfortunately, beyond the a priori reasons put forth by the literature, there have been few empirical studies of whether actual consumers follow sequential or fixed sample size search strategies. This is, no doubt, due to the difficulty of collecting data on individual search behavior. Therefore, most of what we know about individual-level search behavior is from laboratory experiments. The majority of the experimental literature on search has focused on sequential search.[3] Schotter and Braunstein (1981) have reported that on average subjects tend to search in a fashion that is consistent with sequential search strategies, although subjects tend to engage in too little search to be searching optimally. Kogut (1990) and Sonnemans (1998) find evidence that individuals make decisions based on the total return from searching instead of on the marginal return from another draw as they would do if searching sequentially, resulting in too little search. Moreover, Kogut (1990) finds that about one third of the time individuals accepted old offers, which violates optimal policy. More recently, Brown, Flinn and Schotter (2011) find declining reservation wages in the duration of search in their baseline experiment, which is not in line with the standard sequential search model. Harrison and Morgan (1990) directly compare fixed sample size and sequential strategies to so-called variable-sample-size strategies. The latter strategy is described in Morgan and Manning (1985) and is a generalization of both fixed sample size and sequential search since it allows individuals to choose both sample size and number of times to search. Harrison and Morgan (1990) report that experimental subjects indeed employ the least restrictive strategy if they are allowed to do so.

Aside from experimental studies, Hong and Shum (2006) and Chen, Hong and Shum (2007) are the only papers that we are aware of which have attempted to discriminate between sequential and fixed sample size search models using data from a real-world market. Hong and Shum (2006) collect data on textbook prices, and estimate structural parameters of search cost distributions (i.e. the demand parameters) that rationalize the prices set by competing stores. They find larger search-cost magnitudes for the parametrically estimated sequential search model than for the nonparametrically estimated fixed sample size search model. Similar data is used in Chen, Hong and Shum (2007) to conduct a nonparametric likelihood ratio test for choosing among the nonparametrically estimated moment-based fixed sample size and parametrically estimated sequential search models. Although certain parameterizations of the sequential search model are found to be inferior, they conclude that it is difficult to distinguish between the fixed sample size search model and the log-normal parameterization of the sequential search model in terms of fit.

This paper utilizes novel data on the web browsing and purchasing behavior of

---

[2]Examples of sequential search models in the consumer search literature are Axell (1977), Reinganum (1979), Carlson and McAfee (1983), Rob (1985), and Stahl (1989).

[3]See Camerer (1995, pp. 670–73) for a review of this literature.

a large panel of consumers to test classical models of consumer search. Our data, described in detail in Section I, allows us to observe the online stores consumers visited while shopping for a particular item, and which store the consumer decided to buy from. As pointed out by Kogut (1990), and as we will argue in more detail in Section II below, under the reservation price rule prescribed by the "benchmark" model of sequential search, a consumer always buys from the last store she visited, unless she has visited all stores she is aware of. Using data on consumers shopping for books online, we find that this prediction is rejected by a large number of consumers in our data set.

Even though the online book market has limitations when it comes to studying search behavior, in particular the dominance of Amazon and Barnes and Noble and the relatively small price differences between stores, our focus on online book sellers is intentional. Books are relatively homogeneous, and online book retailing is a relatively mature and predominant online industry. Many individuals in our data set have bought books online, and more importantly, a substantial number of book titles have been purchased by more than one individual, something which is not necessarily the case in other online industries. Nevertheless, it is important to keep in mind that our conclusions may not generalize to other (online) markets, especially markets in which price differences are more significant and markets that are not characterized by one or two dominant players.

In Section II we also look at other testable predictions of the sequential model versus the fixed sample size search model. Even when consumers use a fixed sample size search strategy they will visit the websites of bookstores sequentially. Therefore, the crucial distinction between sequential and fixed sample size search models is in how consumers select the number of stores to visit. If consumers are using a fixed sample size search strategy they will decide in advance how many stores to sample; if searching sequentially the number of stores visited is a random variable which depends on the outcome of search. A robust prediction of sequential search is therefore that the decision to search again depends upon the outcome of the previous search, while it does not with fixed sample size search. Using the observed browsing patterns as well as the book prices in our dataset, we do not find any dependence of the decision to continue searching on observed prices, which one would expect if consumers were searching according to one of the sequential search variants.

According to our data around 36 percent of consumers do not buy from the lowest priced store in their sample. Moreover, consumers are more likely to sample certain stores (Amazon in this context) when searching. Both observations suggest differences between bookstores other than price (e.g. perceived reliability/expediency of shipping, ease of using the online interface) might be important. To test for sequential search in a way that accounts for differentiation across stores, we study the effect of relative price position *within* bookstores on the decision to search again. Again we do not find any correlation between the decision to search further and sampled prices, ruling out sequential search.

Finally, in Section III we use the favored fixed sample size search model to estimate the price elasticities faced by online stores. To do this, we derive expressions for demand elasticities implied by the fixed sample size search model. One important feature of this model is that we allow for product differentiation and *asymmetric sampling*: due to for instance advertising or prior shopping experience, consumers' first draw may be skewed toward some online stores (e.g. Amazon) over others. Consumers are assumed to have full knowledge of their utility from buying the book from a particular bookstore except for the price they have to pay, which they learn from costly sampling. We also estimate a specification in which consumers observe their utility from shopping from a given store except for price and a stochastic component that is observable only upon visiting the store (for example, some stores may have the book in stock and ship it right away, and others may do so with some delay).

Our estimates for own-price elasticities of demand are larger in absolute value than in a standard discrete choice differentiated products model that incorrectly assumes consumers have full information about their choices (see also Draganska and Klapper, 2011, for a similar finding in an advertising model). This is intuitive: the price changes we—as econometricians—observe in the data are not observed by consumers who sample only a subset of the stores. A full information logit model assumes that all prices are observed, thus ascribing unresponsiveness to price changes to low price elasticity. Our results also indicate mostly higher price elasticities for Amazon as well as lower price elasticities for Barnes and Noble than reported by Chevalier and Goolsbee (2003). A further discussion of our results vis-à-vis prior findings is in Section III.D.

Our demand model in Section III is related to several papers that embed discrete choice models of differentiated product demand in costly search models—particularly the "consideration set" models in the marketing literature that relate the formation of consideration sets to fixed sample size search (see, for example, Roberts and Lattin, 1991; Mehta, Rajiv and Srinivasan, 2003). We should note, however, that ours is the only paper that bases its specification of the search process on empirical tests. Our model is closest to Moraga-González, Sándor and Wildenbeest (2011) and Honka (2010) who develop discrete choice models of demand with fixed sample size consumer search in which consumers engage in search to determine whether a product is a good match. While Moraga-González, Sándor and Wildenbeest (2011) estimate their model using aggregate data in which consumers' search behavior and choice sets are not observed, our application, like Honka's, uses data on individual consumers' choice sets. However, Honka's data do not contain information on the sequence of searches, which is key to testing between sequential and fixed sample size search. Kim, Albuquerque and Bronnenberg (2010) and Koulayev (2009) estimate sequential search models of demand. While Kim, Albuquerque and Bronnenberg (2010) utilize aggregate information on search behavior at Amazon, Koulayev (2009) utilizes individual search histories from an online hotel price comparison engine. Koulayev (2009) models the

decision to go to the next page of the comparison site as a function of the utility of the hotels observed so far as well as search costs and finds lower price elasticities than in a full information setting.

I. Data

We construct the dataset using two sources of data. The main data comes from the ComScore Web-Behavior Panel and includes detailed online browsing and transaction data from 100,000 Internet users in 2002 and 52,028 users in 2004. The users were chosen at random by ComScore 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.[4] The data include date, time, and duration of visit, as well as price, quantity, and description of each product purchased during the session.

We find that individuals in the ComScore sample are representative of online buyers in the United States. Comparing Internet users that have bought a product online from our sample with the Internet and Computer Use Supplement of the Current Population Survey (CPS) and the Forrester Technographics Survey, we find that all three are similar in terms of age, education, income, household composition, and other observable characteristics. 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. Using the Com-Score sample, we find that book buyers, those who purchased at least one book online, are slightly older, have greater income, and more education than those who had any online transaction. We refer to De los Santos (2008) for a more detailed description of the sample.

The dataset contains users' transactions for products and services from June to December of 2002 and for the full year of 2004. We excluded observations from stores that could not be identified as online bookstores, such as unidentified domains and auction sites. In total, 18 percent of the transactions were excluded; most of these were from Ebay.com (15 percent of transactions). Although the excluded transactions represent a large number of observations, they cannot be considered sales from an online bookstore because they are auctions of potentially

---

[4]The data only captures web-browsing on one computer, hence it might not capture browsing and transactions at work.

different books, for example used books or autographed volumes. A small number of transactions from international Amazon websites (United Kingdom, Canada, and Germany) were also dropped. Given that Borders transactions were handled by Amazon in 2002 and 2004, we excluded browsing activity from Borders.com to avoid double counting.[5] Approximately 38 percent of the users realized a product transaction in 2002 (48 percent of users in 2004), and 7 percent of users bought at least one book online in 2002 (10 percent in 2004). This results in transactions from 15 online bookstores with 7,558 observations in 2002 and 8,020 observations in 2004.[6]

TABLE 1—TRANSACTIONS AND VISITS BY BOOKSTORE

	Transactions		Visits	
Bookstore	Number	Percentage	Number	Percentage
Amazon	10,197	65.5	249,593	76.3
Barnes and Noble	3,042	19.6	25,758	7.9
Book Clubs				
christianbook.com	615	3.9	3,968	1.2
doubledaybookclub.com	468	3.0	4,001	1.2
eharlequin.com	61	0.4	3,647	1.1
literaryguild.com	322	2.1	3,500	1.1
mysteryguild.com	187	1.2	2,095	0.6
Other Bookstore				
1bookstreet.com	10	0.1	120	0.0
allbooks4less.com	5	0.0	199	0.1
alldirect.com	27	0.2	490	0.1
ecampus.com	114	0.7	1,206	0.4
powells.com	68	0.4	1,326	0.4
varsitybooks.com	16	0.1	218	0.1
walmart.com	183	1.2	28,663	8.8
booksamillion.com	246	1.6	2,290	0.7
Total	15,561	100.0	327,074	100.0

In order to analyze consumer search of online bookstores, we grouped small bookstores into two categories. In total we have four stores: Amazon (66 percent of transactions), Barnes and Noble (20 percent), Book Clubs (11 percent), and

[5]Although initially Borders operated Borders.com, in April 2001 it signed a commercial agreement giving Amazon control of customer service, fulfillment, and inventory operations. As a result all visits to Borders.com were redirected to Amazon. In 2008 Borders relaunched Borders.com as an independent online bookstore.

[6]Each observation represents a single book purchased during one transaction; if multiple copies of the book are purchased in the same transaction, it is recorded as one observation.

Other Bookstores (4 percent). "Book Clubs" include the following stores (.com): Christianbook, Doubledaybookclub, Eharlequin, Literaryguild, and Mysteryguild. Other Bookstores include (.com): 1bookstreet, Allbooks4less, Alldirect, Book-samillion, Ecampus, Powells, Varsitybooks, and Walmart. Table 1 displays the number of transactions and visits per bookstore for the groups of stores.

The browsing activity of all users consists of 112,361 visits to the websites of online bookstores in 2002 and 214,713 visits in 2004.[7] 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 7 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 7 days if another transaction has occurred within 7 days. Limiting browsing to search occurring 7 days prior to a purchase reduces the sample to 18,350 observations in 2002 and 25,556 in 2004. Although some user search may not be linked to the next transaction, but to a subsequent one, there is no clear way to link this intervening search to a later transaction. For example, if a user searches prices for book A but buys book B first, the search for book A is linked to book B. In the case where multiple books are acquired in the same transaction, browsing is linked to all books purchased.[8] In the empirical analysis we use several definitions of the relevant search period, from 7 days to the same day of the transaction. Table 2 gives descriptive statistics of the sample.[9]

Despite the relatively large number of online bookstores in 2002 and 2004, the market is highly concentrated, with the two dominant stores capturing 83 percent of the market: Amazon (66 percent of book sales) and Barnes and Noble (17 percent).[10] Amazon was visited in 74 percent of book transactions, and in only 17 percent of transactions did Amazon buyers browse any other bookstore. Also, these stores capture most of the searching activity online. Of the 234 online bookstores listed on the Yahoo directory, 15 bookstores in the sample capture 98.4 percent of all consumer visits to an online bookstore. The dominance of Amazon and Barnes and Noble in the market might explain the low levels of consumer search: users on average searched 1.2 bookstores in 2002 and 1.3 in 2004 (De los Santos, 2008).

A limitation of the ComScore data is that although we observe consumer visits to different stores, we only observe the price of the transaction. To recover missing prices for all visited bookstores we use the most recent transaction prices at those

---

[7]This large increase was the result of a more than twofold increase in the number of visits to Amazon, which is the largest online bookstore and had 80 percent of website visits in 2004.

[8]We treat consumers who visited multiple book clubs as having visited only one store. However, there were only 193 instances of consumers visiting multiple book clubs or multiple other bookstores, comprising a tiny fraction (1.2 percent) of our data set. Therefore we do not expect this to affect our results.

[9]There is only limited use of price search engines or shopbots. For instance, in our 2004 sample only 12 transactions were referrals from mySimon.com (around that time the leading price comparison site for books).

[10]Books sales in dollars for 2004 from the ComScore data sample.

Table 2—Descriptive Statistics of ComScore Book Sample

	2002		2004	
	Mean	Std. Dev.	Mean	Std. Dev.
Duration of each website visit (in minutes)				
Visits not within 7 days of transaction	8.89	13.03	7.69	12.36
Visits within 7 days, excluding transactions	12.72	15.83	11.02	15.00
Visits within 7 days, including transactions	19.04	18.26	15.74	17.37
Transactions only	28.06	17.69	26.08	17.71
Total duration, excluding transaction visits	32.47	49.80	38.41	78.33
Total duration, including transaction visits	43.88	43.27	47.43	66.11
Number of stores searched	1.27	0.54	1.30	0.56
Number of books per transaction	2.38	2.10	2.20	1.95
Transaction expenditures (books only)	36.67	40.64	32.21	35.68
Number of books purchased	17,956		17,631	
Number of transaction sessions	7,559		8,002	
Number of visits within 7 days	18,350		25,556	
Number of visits not within 7 days	94,011		189,157	

bookstores with missing values. The average time lapse between the most recent available transaction price and the transaction for which we are imputing a price is 10.8 days.[11] Given the breadth of books purchased, the imputation process resulted in prices for more than one store in only 13 percent of the observations. However, as consumers visited more than one store in only 25 percent of the transactions, we have all the prices for the stores visited by the consumer in 77 percent of the observations.

Consumers do not always buy from the lowest priced store: only in 63 percent of the observations do they buy from a store with the lowest price (or equally lowest price as often multiple stores have the same price). The average price difference between the transaction price and the lowest price across all stores is \$2.60. 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 \$1.99. The potential gains of search are further diminished by the fact that consumers are willing to pay higher prices to buy at their preferred store. Interestingly, if we estimate this price difference for transactions occurring at the two largest stores, the average consumer left \$1.19

---

[11]Unfortunately we do not know how frequently stores changed book prices during the sampling period. However, using daily price data collected more recently from various online bookstores we find that during April 2010 Amazon changed prices on average 4.1 times for the ten highest ranked hardcover fiction and non-fiction books on the New York Times best seller list of March 28, 2010, while Barnes and Noble changed prices on average 2.9 times. This means our approach will likely lead to some measurement error, although the above numbers suggest this error will not be very large.

on the table when shopping at Amazon compared to \$2.45 for consumers buying at Barnes and Noble.
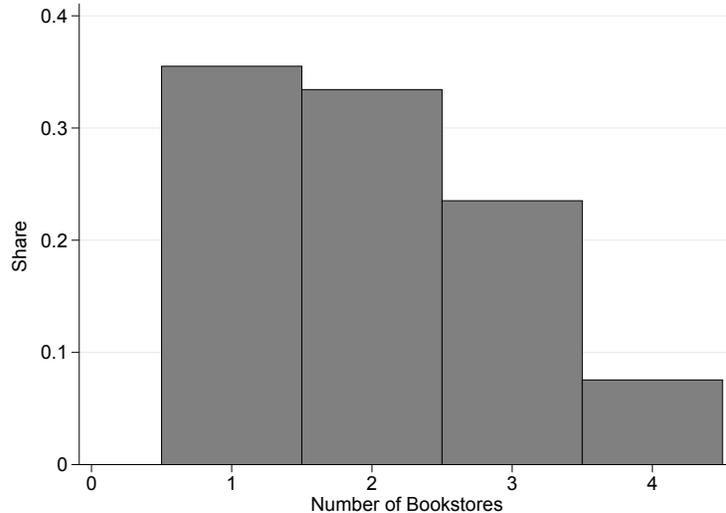

FIGURE 1. CONSUMER BOOKSTORE AWARENESS

Given the large number of online bookstores relative to the low number of bookstores actually visited, we need to define which bookstores are relevant in the consumer search process as consumers might not be aware of all the online bookstores in the market. We construct consumers' awareness of different bookstores by analyzing the consumer's browsing history within the dataset. For each transaction, a consumer is aware of a given bookstore if she has previously visited the bookstore. For a given search sequence the number of bookstores N is defined as the number of bookstores a consumer is aware at the time of the transaction. Figure I displays the distribution of consumer bookstore awareness.

II. Empirical Implications of Search Models

We study a setting in which consumers inelastically demand one unit of a good sold by a finite number of stores. Before searching consumers do not know the realized prices of the available alternatives; in order to have this revealed, consumer i has to pay a constant search cost c_i per alternative, which is assumed to be randomly drawn from a search cost distribution. We study two approaches in modeling consumers' search behavior: fixed sample size search and sequential search.

Consider first the fixed sample size search paradigm in a homogenous good setting. Assuming the consumer believes that each store's price is an i.i.d. draw

from distribution $F(p)$ with density $f(p)$ and support between $\underline{p}$ and $\overline{p}$, a consumer will determine the optimal number of stores k in her sample by minimizing the sum of the expected price and total search cost:

$$k(c) = \arg\min_k \int_{\underline{p}}^{\overline{p}} k \cdot p(1 - F(p))^{k-1} f(p) \, dp + k \cdot c.$$

The integral on the right-hand side is the expected minimum price drawn from the price distribution when sampling k alternatives, so the optimal sample size is determined by finding sample size k that minimizes the sum of the expected minimum price out of searching k times and the total search cost $k \cdot c$.

Next consider sequential search. As shown by McCall (1970) a consumer will continue to search as long as she finds a price higher than some reservation price $r(c)$, where $r(c)$ is given by:

$$(1) \qquad\qquad c = \int_{\underline{p}}^{r(c)} (r(c) - p) f(p) \, dp.$$

As seen in the equation, the reservation price is such that, if the price in hand is $r(c)$, the marginal cost of search c equals the expected benefit from continued search—the integral on the right-hand side is the expected decrease in price from another search, accounting for the option value of discarding higher price draws.

A.   Recall

One important difference between the two search paradigms is in the recall behavior of consumers. By definition, fixed sample size search implies a consumer first samples all alternatives in a subset of bookstores and then decides which alternative to purchase. Therefore, if a consumer searches multiple times and the first option sampled offers the best deal, a consumer will return there to buy. However, for a consumer searching sequentially at a search cost c the reservation price $r(c)$ is constant across searches, which means the consumer will never recall an alternative that she sampled earlier, unless there are a finite number of stores, and the consumer has visited all the stores.[12] Our first test will focus on recall behavior by consumers.

Test 1 (No Recall) *Under the null hypothesis of the standard sequential search model, we should not observe recall of previously sampled alternatives, unless the consumer has sampled* all *of the stores she is aware of.*

The benchmark sequential search model tells us that the only instance that a consumer will recall a store is if she exhausts the search by visiting all stores. If the

---

[12]If all stores have prices exceeding the reservation price $r(c)$ the sequential search literature usually assumes that consumers will buy from the store with the lowest price among all the stores.

consumer does not exhaust the search, the optimal stopping rule is to buy from the last sampled alternative, which means this alternative should have a price that is less than the reservation price.

To test this hypothesis, we have to check whether (i) a consumer recalled a product that was previously sampled, and (ii) if there was recall, whether this was because the consumer searched all stores she is aware of. To do this, we first identify all the stores that a consumer is aware of by looking at previous visits to bookstores by that consumer. For instance, if we observe that the consumer has only visited Amazon and Barnes and Noble in the past, this is a conservative lower bound on the set of stores that the consumer is aware of.

TABLE 3—TEST OF "NO RECALL" HYPOTHESIS

Search window	No. of stores visited	Percentage	If 2 or more stores, bought from:	Percentage	Percentage exhausted search?
7 Days	One	76			
	2 or more	24	Last store sampled	65	
			Recalled	35	55
6 Days	One	77			
	2 or more	23	Last store sampled	64	
			Recalled	36	55
5 Days	One	79			
	2 or more	21	Last store sampled	63	
			Recalled	37	55
4 Days	One	80			
	2 or more	20	Last store sampled	61	
			Recalled	39	55
3 Days	One	82			
	2 or more	18	Last store sampled	61	
			Recalled	39	56
2 Days	One	84			
	2 or more	16	Last store sampled	61	
			Recalled	39	56
1 Day	One	86			
	2 or more	14	Last store sampled	61	
			Recalled	39	56
Same day	One	90			
	2 or more	10	Last store sampled	62	
			Recalled	38	58

For a given transaction the consumer visits one store or the consumer searches more than one store. If the consumer visits more than one store, she either buys from the last store, or she recalls a previously visited store. In the case where

the consumer visits only one store, we cannot distinguish between sequential and fixed sample size strategies.

Table 3 shows the percentage of transactions for each of the three search sequences for differing definitions of the search period. The periods range from one week prior to the same day of the transaction. For example, for the search period defined as the same day of the transaction (bottom row of the table), in 90 percent of the transactions the consumer visited one store in the same day. In 10 percent of transactions, consumers visited more than one bookstore. Among the 10 percent of transactions in which a consumer visited more than one store, 62 percent bought from the last store sampled and 38 percent recalled a previously visited store.

Note that there are a large number of instances where the consumer recalls a product that was previously sampled. This may not immediately be construed as evidence against a sequential model, however, as recall is allowed in a sequential search model in which a consumer has exhausted the search options available to her. The last column presents the percentage of the transactions in which the consumer visited all the stores she is aware of (i.e. the stores she has visited before). If we focus on the bottom row of the table, where we look at search activity only on the day of the transaction, we see that consumers "exhausted" the search possibilities in 58 percent of those transactions where they recalled a previously sampled product. Perhaps more to the point, consumers did not exhaust the search in 42 percent of the recalled instances, which is a violation of the basic sequential search model. Note that our definition of "not exhausting a search" is a conservative one; it may have been the case that the consumer was aware of more bookstores than we were able to capture with our data set.

B. Price Dependence

The previous section examines recall behavior to test the basic sequential search model with a constant reservation price strategy. However, it is possible for more general models of sequential search to rationalize recall behavior. For example, a sequential search setting in which search costs are increasing over the search duration could generate recall. Relatedly, in a model where consumers start with prior beliefs regarding the price distribution and update their beliefs using observed prices (see, for instance, Rothschild, 1974; Rosenfield and Shapiro, 1981), recall is also possible.

A more robust empirical difference between the sequential and fixed sample size paradigms is that the optimal sample size under fixed sample size search is independent of observed prices, while under sequential search the decision to continue searching or not depends on the realization of the previous search. Consumers searching sequentially are therefore more likely to continue searching when a relatively high priced is observed. This also means that consumers who search only once are more likely to have encountered a relatively low price. Our next test focuses on the dependence of search decisions on observed prices.

Test 2 (Price Dependence) *Under the null hypothesis of the standard sequential search model, consumers searching once are more likely to have found a relatively low price, while the first price observation of consumers searching twice is likely to be relatively high.*

To test this hypothesis we use a logit model to regress a variable that represents continued search on the relative position of prices observed so far, assuming expectations are rational. Because the number of books for which we have prices at all four bookstores is very limited, in this test we will only focus on Amazon and Barnes and Noble. This means the dependent variable in our regression is a variable which has value 1 if the consumer searched both Amazon and Barnes and Noble, and 0 if the consumer visited just one of the two before buying. Our explanatory variable of interest is 1 if the price at the first store visited is lower than or equal to the second store's price and 0 if the first store visited has the higher price.

The estimation results are presented in Panel A of Table 4. Specification (A) includes all transactions for which we have prices for both Amazon and Barnes and Noble. The results for this specification show that the coefficient estimate for the first price being lower than or equal to the second price is negative, but not significantly different from zero. This is also shown by its average marginal effect, which is negative and, although insignificant, very close to zero. This goes against predicted behavior when searching sequentially—one would expect relatively low prices at the first store visited to have a significant negative impact on the probability of searching once more. Specification (A) does not control for a consumer's search cost, which makes it difficult to interpret the estimated coefficient: a consumer might simply stop searching because her search cost is so high that any observed prices do not matter. In specifications (B) and (C) we try to capture this by only including consumers with multiple transactions during the sampling period, assuming these consumers have lower search costs and thus are willing to continue searching if the first price observation is relatively high. The estimation results for specification (B) show that the coefficient for the first price being lower than or equal to the second price is now positive but again not significantly different from zero. Even when we add a loyalty dummy in specification (C), which is 1 if the consumer has always bought from the same store and 0 otherwise, the estimates do not change qualitatively: as expected, the loyalty dummy has a negative sign, but the average marginal effect of the first price being lower than or equal to the second price is still indistinguishable from zero.

In specification (D) we narrow the definition of a "searcher" by only including consumers who have searched multiple times shortly before at least one of their transactions, inferring that these consumers have relatively low search costs. As in previous specifications the coefficient for the first price being lower than or equal to the second price is not significantly different from zero. In specification (E) we add consumer fixed effects. For this we need to drop consumers who always

TABLE 4—ESTIMATES OF CONSUMER CONTINUED SEARCH ON FIRST OBSERVED PRICE

Variable	(A)	(B)	(C)	(D)	(E)
Panel A. All transactions					
Coefficients					
Intercept	−1.817	−1.796	−0.818	0.295	
	(0.093)***	(0.123)***	(0.154)***	(0.133)**	
First price lower or equal	−0.071	0.090	0.040	−0.223	−0.073
	(0.119)	(0.152)	(0.157)	(0.165)	(0.371)
Loyal			−1.446		
			(0.151)***		
Average marginal effects					
First price lower or equal	−0.008	0.011	0.005	−0.055	−0.015
	(0.014)	(0.019)	(0.019)	(0.041)	(0.078)
Loyal			−0.171		
			(0.017)***		
Consumer fixed effects	No	No	No	No	Yes
Number of observations	2,593	1,504	1,504	649	235
Panel B. Price difference more than 25 percent					
Coefficients					
Intercept	−1.811	−1.904	−0.812	0.194	
	(0.131)***	(0.174)***	(0.233)***	(0.180)	
First price lower or equal	−0.088	0.207	0.153	−0.054	1.235
	(0.184)	(0.237)	(0.246)	(0.252)	(0.686)*
Loyal			−1.566		
			(0.248)***		
Average marginal effects					
First lower or equal	−0.010	0.025	0.017	−0.014	0.260
	(0.021)	(0.029)	(0.028)	(0.063)	(0.130)**
Loyal			−0.176		
			(0.027)***		
Consumer fixed effects	No	No	No	No	Yes
Number of observations	1,014	590	590	253	62

Note: Standard errors in parentheses. In all specifications the dependent variable has value 1 if searching twice and 0 if searching once. In specification (A) all relevant observations are included, specifications (B) and (C) only include consumers with multiple transactions, specification (D) only includes consumers who have searched more than once in multiple transactions, and specification (E) only includes consumers who have searched both once and twice within the sample period.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

search once or always search twice, leaving us with 235 observations. As shown in the table, the coefficient on the first price being lower than or equal to the second price is again not significantly different from zero.

A potential issue is that for many transactions the price difference between Ama-

zon and Barnes and Noble is relatively small. This might imply the regressions in Panel A of Table 4 have little power. As a robustness check, in Panel B of the table we restrict the transactions to only those in which the price difference between Amazon and Barnes and Noble is more than 25 percent. The regression results do not change that much—all parameters of interest in all specifications are still not significantly different from zero (specifications (A)-(D)), or are significantly different from zero but have the wrong sign (average marginal effects in specification (E)).

TABLE 5—PRICE OF THE FIRST STORE BY NUMBER OF SEARCHES

Price of the first store	Once	Twice	Total
Lower or equal	63.55%	61.89%	63.32%
Higher	36.45%	38.11%	36.68%
Number of observations	2,244	349	2,593

Our finding that there is no strong relation between the decision to continue searching and observed prices is also illustrated in Table 5. This table gives the percentage of transactions for which the consumer has found the lower or equal price at the store visited first. The tabulation is done separately for consumers searching once and consumers searching twice. For both groups of consumers the price at the first store is lower or equal in approximately the same percentage of transactions: in 64 percent of the transactions in which consumers search once the store visited first has the lowest price, while for transactions in which consumers search twice this percentage is 62 percent (a z-test fails to reject the equality of these two proportions, p-value = 0.55). This suggests consumers who searched twice did not find significantly worse prices at the first store visited than consumers who searched only once, which violates the predictions of a sequential search model.

C. Product Differentiation

Both the standard sequential search model and the standard fixed sample size search model assume that prices of the stores are i.i.d. random variables and that stores are the same in every dimension but price. This means these models can explain variation in the number of searches in terms of heterogeneity in consumer search costs, but cannot explain variation across stores in terms of sampling frequency or search order. The i.i.d. assumption implies consumers are indifferent between the stores, so *every* observed search order can be rationalized by these models, even though random (often uniform) sampling is usually assumed.

Nevertheless, any observed unequal sampling in the data is very likely to be related to heterogeneity in consumer preferences. Even though the bookstores in our sample sell homogenous goods, consumers might prefer one bookstore over another because of previous buying experiences, transaction ease, store reputation, etc.

TABLE 6—SEARCH ORDER BY STORE

	One store searched ($n = 26{,}803$)	Two stores searched ($n = 7{,}226$)				
		First store visited	Second store visited			
			Amazon	Barnes & Noble	Book Clubs	Other Bookstores
Amazon	0.69	0.57	-	0.54	0.21	0.25
Barnes and Noble	0.17	0.17	0.80	-	0.08	0.11
Book Clubs	0.11	0.09	0.60	0.27	-	0.13
Other Bookstores	0.03	0.17	0.63	0.24	0.13	-

Table 6 shows that the bookstores in our dataset are indeed sampled unequally. Amazon has by far the highest probability of being searched: 69 percent of consumers searching once visited Amazon and 57 percent of consumers searching twice first visited Amazon. Even consumers starting their search at another bookstore were very likely to visit Amazon second. Moreover, according to our data around 36 percent of consumers do not buy from the lowest priced store in their sample, conditional on sampling more than one store. Both findings suggest that consumers' store preferences matter for the decision regarding where to buy.

To deal with store heterogeneity and unequal sampling probabilities we consider a setting where consumer i's indirect utility of buying a book at store j is given by

$$(2) \qquad u_{ij} = \delta_{ij} + \alpha_i p_j,$$

where δ_{ij} is the consumer's gross utility from each store and α_i is a price coefficient. We assume consumers know δ_{ij} and α_i but as before have to search in order to find price p_j. Furthermore, we assume prices are independently but no longer identically distributed across stores. As shown by Weitzman (1979), in this case the optimal sequential search procedure is to start searching at the alternative with the highest reservation utility and to terminate search whenever the maximum sampled utility exceeds the reservation utilities of all remaining unsampled alternatives.[13]

---

[13]Notice that such a setting might also explain recall behavior. After each search the reservation utilities of the unsampled alternatives will go down, which means that a previously sampled alternative might not pass the threshold initially but does so later on, resulting in recall.

The optimal fixed sample size search procedure in such a setting is described in Chade and Smith (2006). They show a "greedy algorithm" is capable of finding the optimal choice set—this means a consumer should first determine the best singleton, and should keep adding the next best single addition as long as the gains offset additional search costs. After the optimal choice set is determined, the consumer will visit the stores in her choice set and buy the alternative that gives the highest realized utility.

Notice that also with product differentiation the decision to search again in a sequential search framework will depend on the outcome of the previous search, while there is no such dependence in a fixed sample size search framework. Still, Test 2 does not apply in this setting because any observed price differences across stores might not be enough to offset a consumer's store-specific preferences, as captured by δ_{ij} in equation (2). To test for sequential search in a way that is immune to product differentiation, we study the effect of relative price positions on search behavior *within* bookstores instead of across bookstores.[14] For instance, if at some point in time *Freakonomics* is discounted relatively aggressively at Amazon, one would expect consumers to terminate their search early in comparison to other weeks during which *Freakonomics* is priced relatively higher at the same store. Under sequential search, the size of the search set should be smaller because consumers should be more likely to terminate their search early. Our next test is a variant of Test 2 and relies on within-store price dependence to control for store differentiation.

Test 3 (Price Dependence with Product Differentiation) *Under the null hypothesis of the standard sequential search model with product differentiation, consumers are more likely to continue searching if the price of a book is relatively high in a store's price distribution for that book.*

Table 7 presents the effect of the within-store relative price on consumer search. The relative price of each book is calculated as the percentage difference between the price of the first store searched by the consumer and the average price charged for that book at that store over the sample period (using at least two price observations per book at the store). Specifications (A) through (E) correspond to the ones used in Table 4 for transactions at any of the four bookstores. In none of the specifications of Panel A the relative price variable is significantly different from zero, which means that even a relatively high price within the price distribution of the first store sampled does not make a consumer more likely to continue searching. Panel B shows the results of a regression of the number of stores visited by consumers on the within-store relative price. On average consumers should have the same ex ante ranking of stores for a given book, thus under sequential search the number of stores visited should be smaller for lower relative prices because consumers should be more likely to terminate their search early. The price coeffi-

---

[14]We thank Jean-Pierre Dubé for this suggestion.

Table 7—Estimates of Consumer Search on Within-Store Relative Price

Variable	(A)	(B)	(C)	(D)	(E)
Panel A. Dependent variable: Search two or one stores (Logit)					
Coefficients					
Intercept	−2.518	−2.414	−1.434	−0.424	
	$(0.043)^{***}$	$(0.053)^{***}$	$(0.089)^{***}$	$(0.053)^{***}$	
Within store relative	0.090	0.180	0.225	0.305	0.600
price premium	(0.162)	(0.182)	(0.192)	(0.220)	(0.555)
Loyal			−1.320		
			$(0.111)^{***}$		
Average marginal effects					
Within store relative	0.006	0.014	0.016	0.073	0.107
price premium	(0.011)	(0.014)	(0.014)	(0.052)	(0.099)
Loyal			−0.096		
			$(0.009)^{***}$		
Panel B. Dependent variable: Number of stores searched (OLS)					
Coefficients					
Intercept	1.075	1.082	1.193	1.396	0.970
	$(0.003)^{***}$	$(0.004)^{***}$	$(0.010)^{***}$	$(0.013)^{***}$	$(0.276)^{***}$
Within store relative	0.006	0.015	0.016	0.073	0.101
price premium	(0.012)	(0.015)	(0.015)	(0.052)	(0.107)
Loyal			−0.133		
			$(0.010)^{***}$		
Consumer fixed effects	No	No	No	No	Yes
Number of observations	7,796	4,802	4,802	1,472	653

Note: Standard errors in parentheses. In all specifications the dependent variable has value 1 if searching twice and 0 if searching once. In specification (A) all relevant observations are included, specifications (B) and (C) only include consumers with multiple transactions, specification (D) only includes consumers who have searched more than once multiple times shortly before at least one of their transactions, and specification (E) only includes consumers who have searched both once and twice within the sample period.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

cients are not significant across specifications which indicates that consumers are not using a sequential search rule.

III. Implications of the Fixed Sample Size Search Model

In this section we show how we can use consumer level data on browsing and purchases to estimate the distribution of search costs as well as demand elasticities in an environment where consumers search using a fixed sample size search strategy. Based on the patterns we observe in our data, we allow for heterogeneity

in store preferences. Our starting point is the utility specification of equation (2), i.e., we assume consumer i's indirect utility of buying a book at store j is given by $u_{ij} = \delta_{ij} + \alpha_i p_j$, where δ_{ij} is the consumer's gross utility from each store, p_j is store j's price for the book, and α_i is a consumer-specific price coefficient. This gross utility is given by

$$\delta_{ij} = \mu_j + X_i \beta_j + \varepsilon_{ij},$$

where we allow this utility to depend on a store fixed effect μ_j, consumer characteristics X_i, and a idiosyncratic utility draw ε_{ij}. Consumers know the gross utility from each store, but are uncertain about prices, which they learn upon visiting a store.

We assume consumers use a fixed sample search strategy—consumers decide which subset of the J online stores to visit and then make a purchase decision among the visited stores. Sampling stores is costly—we allow the cost of sampling a store c_i to depend on consumer characteristics, i.e., $c_i = c + X_i \beta$. The expected net benefit to consumer i of visiting all online stores in a subset of stores S, denoted by m_{iS}, is the difference between the expected maximum utility of sampling the stores in subset S and the cost of sampling these stores, i.e.,

$$(3) \qquad m_{iS} = E \left[\max_{j \in S} \{u_{ij}\} \right] - k \cdot c_i,$$

where k is the number of stores in subset S.

Consumer i will pick the subset S_i that maximizes the expected net benefits. To smooth the choice set probabilities we add a mean-zero stochastic noise term ς_{iS} to m_{iS}. This stochastic noise term ς_{iS} can be interpreted as reflecting errors in an individual's assessment of the net expected gain of visiting all stores in subset S.[15] Assuming ς_{iS} is i.i.d. type I extreme value distributed with scale parameter σ_ς, the probability that consumer i finds it optimal to sample the set of stores S_i is then[16]

$$(4) \qquad P_{iS} = \frac{\exp\left[m_{iS}/\sigma_\varsigma\right]}{\sum_{S' \in \mathcal{S}} \exp\left[m_{iS'}/\sigma_\varsigma\right]}.$$

In the second stage the uncertainty about prices for the selected stores is resolved and consumer i purchases from the store j that provides the highest utility in her sample S_i, i.e., $\arg\max_{j \in S_i} u_{ij}$. This happens with probability $P_{ij|S}$, where $P_{ij|S}$

---

[15]Alternatively, ς_{iS} can be interpreted as the idiosyncratic part of the total cost of sampling subset S, i.e., $c_{iS} = k \cdot c_i - \varsigma_{iS}$, where c_{iS} is the total cost of sampling subset S (see Moraga-González, Sándor and Wildenbeest, 2011).

[16]In the denominator of the choice set probability in equation (4) we have to sum over all possible choice sets. In our application we have only four stores, but if N is large this might be problematic. Honka (2010) offers a solution to this dimensionality problem by assuming first order stochastic dominance among price distributions. As shown by Chade and Smith (2005) this makes it optimal to rank stores according to expected utility and search only the top N firms. Alternatively, Moraga-González, Sándor and Wildenbeest (2011) use importance sampling to deal with the dimensionality problem.

is given by

(5) $$P_{ij|S} = \Pr\left(u_{ij} > u_{ik} \quad \forall \ \ k \neq j \in S_i\right).$$

Since consumers condition their search behavior on ε_{ij}, a store with a relatively large idiosyncratic utility draw is more likely to be selected. Therefore ε_{ij} will not be i.i.d. in the conditional buying stage, which means equation (5) does not have a closed-form solution.

To get the probability of observing a consumer i selecting a choice set S and buying product j we take the product of the probabilities in equations (4) and (5), i.e.,

$$P_{ijS} = P_{iS}P_{ij|S}.$$

A.   Estimation

We assume consumers know δ_{ij} as well as the distribution of prices for each store, but have to sample stores to find actual draws from the price distribution. Calculation of the expected maximum utility of visiting all stores in subset S therefore depends on price expectations. To obtain a closed-form expression for $E\left[\max_{j\in S}\{u_{ij}\}\right]$ we follow Mehta, Rajiv and Srinivasan (2003) and Honka (2010) in assuming that prices follow a type I extreme value distribution with (known) store-specific location parameter γ_j and common scale parameter σ, i.e.,[17]

$$
\begin{aligned}
E\left[\max_{j\in S}\{u_{ij}\}\right] &= \alpha_i\sigma\log\left(\sum_{j\in S}\exp\left[\frac{\delta_{ij} + \alpha_i\gamma_j}{\alpha_i\sigma}\right]\right); \\
&= \alpha_i\sigma\log\left(\sum_{j\in S}\exp\left[\frac{\mu_j + X_i\beta_j + \varepsilon_{ij} + \alpha_i\gamma_j}{\alpha_i\sigma}\right]\right).
\end{aligned}
$$
(6)

We estimate the store specific price distributions by fitting a type I extreme value distribution (with common scale but store-specific location parameters) to the observed prices using maximum likelihood, after accounting for unobserved differences in the characteristics of the books. We then treat the parameters of these distributions as known by the consumers during the actual estimation procedure.

We estimate the model by maximum likelihood. The log-likelihood function is

$$LL = \sum_i \log \check{P}_{ijS} = \sum_i \log \check{P}_{iS}\check{P}_{ij|S},$$

---

[17]We have omitted the Euler constant from equation (6) because it does not affect choices. Note that assuming (for instance) a normal distribution requires numerical integration, which will slow down the estimation substantially.

where $\check{P}_{ijS}$ is the probability that individual i bought at store j from the observed choice set S. The probability of observing choice set $\check{P}_{iS}$ follows from equation (4), while we use a logit-smoothed AR simulator to smooth the conditional buying probability $\check{P}_{ij|S}$.

As in standard discrete choice models of demand the coefficients of the observed variables in our search model are only identified relative to the variance of the unobserved factors. The unobserved factors in our model are the random utility term ε and the choice set specific stochastic term ς. We normalize the variance of the random utility term, which allows us to estimate the variance of the stochastic optimization error term.

B.   Results

We estimate the model using books for which we have at least 20 transactions in total. Most of these books appeared on the *New York Times* Bestseller list for at least part of the sampling period. Table 8 gives descriptive statistics for the books we use to estimate the model. Mean prices are relatively similar across books, with *The Last Juror* (by John Grisham) having the highest average price, while *Angels and Demons* (by Dan Brown) has the lowest mean price. The dispersion of book prices varies: the coefficient of variation ranges from 0.04 to 0.94. The reported shares of consumers sampling k stores shows little variation across books. In line with findings for the complete sample, consumer search activity is very modest: between 52 percent and 95 percent of consumers visits no more than one bookstore before buying and consumers search more than twice for fewer than half of the books in our sample.

The parameters of the type I extreme value price distributions are estimated using prices for only those transactions for which we have prices at all four stores. We control for unobserved differences in book characteristics by first de-meaning prices for each observation and then fitting a type I extreme value distribution with store-specific location parameters and a common scale parameter to the de-meaned prices.[18] The parameters of the rest of the model are estimated by a maximum simulated likelihood procedure, using 100 simulated consumers per observation and assuming the random utility term follows a type I extreme value distribution.

Table 9 gives the parameter estimates for the model, obtained using our maximum likelihood procedure. The results in column (1) are for the sample consisting of transactions corresponding to the 24 books in Table 8. Note that to estimate the model we only need to observe prices for the bookstores actually sampled by the consumer. A book price for at least one bookstore within the consumer's choice set is unavailable in 41 of the transactions, so we exclude these from the sample leaving us with 602 transactions.

---

[18]The fitted (by maximum likelihood) parameters are 1.747 for Amazon, 2.482 for Barnes and Noble, -0.097 for Book Clubs, and 0.061 for Other Bookstores; the fitted (common) scale parameter is 2.262.

Table 8—Descriptive Statistics

Product name	No. Obs.	Price			Consumers by Sample Size (share)		
		Mean	Std. Dev.	CV	μ_1	μ_2	μ_3
Best sellers 2002							
Answered Prayers	26	16.66	2.39	0.14	0.92	0.08	0.00
Dr Atkins New Diet Revolution	25	9.97	3.67	0.37	0.84	0.16	0.00
Four Blind Mice	35	17.09	2.54	0.15	0.80	0.17	0.03
From a Buick 8	37	17.49	2.16	0.12	0.73	0.24	0.03
Haley's Cleaning Hints	22	17.62	4.97	0.28	0.95	0.05	0.00
Harry Potter Paperback Boxed Set 1-4	25	21.46	3.79	0.18	0.60	0.36	0.04
Leadership	21	18.26	17.25	0.94	0.76	0.19	0.05
Let Freedom Ring	23	17.59	4.30	0.24	0.87	0.13	0.00
Q is for Quarry	27	16.97	3.24	0.19	0.78	0.22	0.00
Red Rabbit	22	15.98	4.30	0.27	0.82	0.18	0.00
The Lovely Bones	23	15.50	11.59	0.75	0.78	0.17	0.04
Best sellers 2004							
Angels and Demons	24	7.51	2.55	0.34	0.92	0.08	0.00
Harry Potter and the Half-Blood Prince	35	20.66	8.25	0.40	0.77	0.17	0.06
He's Just Not That into You	26	12.93	1.56	0.12	0.77	0.23	0.00
London Bridges	23	16.35	4.19	0.26	0.70	0.30	0.00
My Life	28	21.08	0.94	0.04	0.75	0.18	0.07
R is for Richochet	21	16.93	4.63	0.27	0.67	0.29	0.05
The Automatic Millionaire	22	13.20	2.02	0.15	0.73	0.27	0.00
The Da Vinci Code	52	14.29	3.28	0.23	0.71	0.25	0.04
The Five People You Meet in Heaven	25	11.47	2.38	0.21	0.72	0.28	0.00
The Last Juror	24	23.16	11.54	0.50	0.79	0.21	0.00
The South Beach Diet	35	14.46	1.78	0.12	0.83	0.17	0.00
Trace	21	16.18	3.73	0.23	0.52	0.43	0.05
Unfit for Command	21	16.49	3.30	0.20	0.90	0.05	0.05

Note: Coefficient of variation (CV) is calculated as the standard deviation over the mean. Prices are in US dollars.

TABLE 9—ESTIMATION RESULTS

	Search: ε observed (1)		Search: ε observed (2)		Search: ε unobserved (3)		Full information (4)	
Variable	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Price								
Income less than $35,000	-0.243	(0.024)***	-0.464	(0.080)***	-0.398	(0.205)*	-0.109	(0.084)
Income $35,000–$75,000	-0.205	(0.009)***	-0.347	(0.055)***	-0.316	(0.173)*	-0.168	(0.074)**
Income more than $75,000	-0.236	(0.015)***	-0.484	(0.027)***	-0.325	(0.186)*	-0.102	(0.082)
Bought before at store	3.217	(0.066)***	4.710	(0.310)***	2.704	(0.999)***	5.503	(1.103)***
Search cost								
Constant	0.309	(0.042)***	1.187	(0.358)***	1.647	(0.409)***	—	
Household size	-0.005	(0.008)	-0.095	(0.054)*	-0.127	(0.068)*	—	
Broadband connection	-0.003	(0.020)	-0.020	(0.119)	-0.026	(0.150)	—	
Household income < $50,000	0.025	(0.024)	0.136	(0.142)	0.193	(0.178)	—	
Age 60 and over	-0.012	(0.022)	0.048	(0.148)	0.059	(0.187)	—	
Store fixed effect								
Amazon	1.997	(0.375)***	2.706	(1.191)**	2.775	(1.202)**	3.890	(1.813)**
Barnes and Noble	1.211	(0.378)***	1.368	(1.093)	1.933	(1.130)*	2.015	(1.848)
Book Clubs	0.736	(0.370)**	1.277	(1.168)	1.755	(1.011)*	3.126	(1.801)*
Household size								
Amazon	-0.201	(0.066)***	-0.264	(0.248)	-0.089	(0.249)	-0.246	(0.499)
Barnes and Noble	-0.204	(0.067)***	-0.115	(0.219)	0.029	(0.248)	0.107	(0.497)
Book Clubs	-0.305	(0.066)***	-0.113	(0.247)	-0.058	(0.244)	-0.065	(0.492)
Broadband connection								
Amazon	-0.409	(0.157)***	-1.019	(0.539)*	-0.649	(0.570)	-0.754	(1.079)
Barnes and Noble	-0.009	(0.166)	0.142	(0.525)	-0.091	(0.573)	0.329	(1.114)
Book Clubs	-0.807	(0.170)***	-1.085	(0.550)*	-0.664	(0.577)	-1.057	(1.075)
Household income < $50,000								
Amazon	0.299	(0.191)	0.391	(0.513)	0.417	(0.603)	0.747	(1.206)
Barnes and Noble	-0.066	(0.196)	-0.074	(0.526)	0.069	(0.625)	0.012	(1.254)
Book Clubs	0.441	(0.196)**	1.045	(0.519)**	0.622	(0.581)	1.366	(1.189)
Age 60 and over								
Amazon	-0.465	(0.183)**	-1.088	(0.580)*	-1.005	(0.570)*	-1.582	(1.121)
Barnes and Noble	-0.324	(0.175)*	-0.762	(0.350)**	-0.803	(0.585)	-1.275	(1.149)
Book Clubs	0.141	(0.170)	-1.241	(0.535)**	-0.995	(0.588)*	-1.846	(1.135)
Scale ς (σ_ς)	0.082	(0.012)***	0.344	(0.106)***	0.432	(0.127)***	—	
Number of observations	602		173		173		173	
Log-likelihood	831.85		275.71		268.45		124.36	

Note: Number of simulated consumers in columns (1) and (2) is 100 per observation. Number of simulated prices drawn from the empirical price CDF in column (3) is 1,000.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

We estimate a separate price coefficient for three different income groups (less than \$35,000, between \$35,000 and \$75,000, and more than \$75,000). The estimates show that the magnitude of the price coefficients is largest for the low-income group. The estimated search cost constant is highly significant—normalizing the estimates of the search cost parameters by the estimated price coefficients indicates that search costs are on average around \$1.35. Having a broadband connection decreases search costs, and each additional household member also has a negative effect on search costs. Figure 2(a) gives a kernel density plot of the estimated search costs for the consumers in our sample, which contains the estimated effects of the demographics included in column (1) of Table 9.
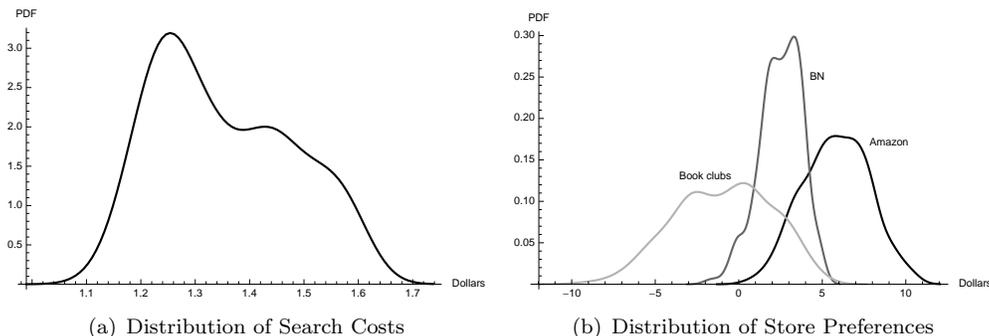

(a) Distribution of Search Costs   (b) Distribution of Store Preferences

Figure 2. Estimated Search Costs and Store Preferences

Our estimate of the scale parameter of the choice-set-specific stochastic term ς is found to be relatively small in magnitude—the scale of ε_{ij} in the utility function is normalized to one, which means that the scale parameter of the utility-specific stochastic term is estimated to be about 12 times higher than the scale parameter of the choice-set-specific stochastic term. The small scale of ς suggests the impact of optimization error on the expected net benefit of the choice sets is relatively small.

The estimated store fixed effects are quite different across stores. Amazon has the highest store fixed effect, which does not come as a surprise given its high sampling probability. Also Barnes and Noble and the Book Clubs have higher store fixed effects than the Other Bookstores. In addition the estimates indicate that store preferences depend on some of the consumer demographics. For instance, compared to buying at Other Bookstores, an additional household member has a negative effect on the marginal utility of buying at all three bookstores listed in the table. To get a better picture of the extent to which estimated store preferences are heterogeneous in our sample, we combine the store fixed effects and the effects of the several consumer characteristics on the marginal utilities of buying at each store. Figure 2(b) gives kernel density plots of how this mean utility (without accounting for price and the preference shock), measured in dollar terms,

is distributed across consumers. For the interpretation of the density plots it is important to keep in mind that we normalize the gross utility of the Other Bookstores to zero, so store preferences are relative to the Other Bookstores category. The plots show that consumers have the highest store preferences for Amazon, with Amazon generating $3.89 more value on average than Barnes and Noble, and $7.53 more on average than Book Clubs. Note that the graph does not include the effect of having bought before from a store—as can be seen in Table 9, consumers put substantial value on this, which, if a consumer has indeed bought before from a store, can compensate for any of the calculated differences in store preferences, even if price expectations would be similar across stores.

TABLE 10—OWN-PRICE ELASTICITIES

Book title	Amazon	Barnes & Noble	Book Clubs	Other Bookstores
Dr Atkins New Diet Revolution				
Expected price	-0.636	-1.156	-1.907	-1.379
Price	-0.112	-0.267	-0.091	-0.268
Combined	-0.748	-1.424	-1.998	-1.647
Four Blind Mice				
Expected price	-1.656	-2.090	-0.766	-5.293
Price	-0.682	-0.949	-0.243	-3.951
Combined	-2.339	-3.038	-1.009	-9.244
From a Buick 8				
Expected price	-1.595	-2.261	-1.488	-5.925
Price	-0.492	-0.937	-0.487	-2.277
Combined	-2.088	-3.197	-1.976	-8.202
Q is for Quarry				
Expected price	-1.384	-2.387	-1.029	-2.869
Price	-0.313	-0.740	-0.249	-1.353
Combined	-1.697	-3.127	-1.279	-4.222
The Da Vinci Code				
Expected price	-0.711	-1.748	-1.251	-3.342
Price	-0.300	-0.918	-0.257	-1.777
Combined	-1.011	-2.666	-1.508	-5.119

Table 10 gives demand elasticities for the books in our data set for which we have all four prices. Prices enter the model in two ways: through consumers' store-specific price expectations that are used to determine choice set probabilities, as well as directly in the conditional buying probabilities. A price change by a store may therefore not be "known" by a consumer unless she visits the store. This means demand elasticities can be disentangled into two separate effects: the demand elasticity with respect to a change in the store-specific mean of the expected price distribution, as well the effect of a price change on the conditional buying probabilities. In Table 10 both effects are shown separately as well as combined,

assuming a similar marginal increase in both prices and the mean of the expected price distribution.[19] The estimated combined own-price elasticities in Table 10 indicate that demand for almost all books is elastic at all four bookstores. On average demand at Barnes and Noble is more elastic than demand at Amazon. For three out of five books demand at the Book Clubs is the most inelastic, while for the other two books Amazon has the most inelastic demand. Most of the combined change in demand is driven by changes in price expectations—if consumers' price expectations do not change, demand is inelastic at all bookstores, with the exception of three out of four books sold at the Other Bookstores. This is not surprising since most consumers have very few stores in their choice sets.

In Table 11 we look in more detail at substitution patterns for the The Da Vinci Code, which is the book with the highest number of transactions. The estimated (combined) cross-price elasticities indicate that price changes at Other Bookstores do not have a substantial impact on market shares of competitors, whereas price changes at Amazon have a much much larger effect on competitors' market shares.

TABLE 11—DEMAND ELASTICITIES (COMBINED) FOR THE DA VINCI CODE

| | Market share | | | |
Price	Amazon	Barnes & Noble	Book Clubs	Other Bookstores
Amazon	-1.011	1.735	1.130	2.366
Barnes and Noble	0.543	-2.666	0.503	2.081
Book Clubs	0.301	0.465	-1.508	0.594
Other	0.129	0.348	0.122	-5.119

C. Alternative Models

Most traditional discrete choice demand models assume consumers observe all prices. This is clearly not the case in our data. To investigate what happens to the parameter estimates if we would incorrectly assume consumers have full information we estimate a similar model imposing that consumers have sampled all stores. To estimate a multinomial logit demand model without search frictions we can only use observations for books for which we have prices at all four stores. This leaves us with 173 transactions. For comparison purposes we re-estimate our search specification using this smaller sample as well—the results for the search model are presented in column (2) of Table 9 while column (4) gives the parameter estimates for the full information model. A first observation is that even though

---

[19]We calculate the elasticities by simulating the average change in buying probability of a marginal increase in either the price or the mean of the expected price distributions (the location parameter of store j's fitted price distribution), which is then multiplied by the ratio of the store's average price and market share of the book. The combined own-price elasticity is the sum of the two individual effects.

most parameters increase in magnitude in comparison to the search model, the estimated price coefficients decrease in magnitude, resulting in less price sensitive consumers on average. This result is also apparent in Table 12, which gives the estimated own-price elasticities for The Da Vinci Code for the search model in the first column as well as the full information model in the last column, both estimated using the smaller sample. For all stores the estimated (combined) own-price elasticities are smaller in absolute value for the full information model, which suggests that if we incorrectly assume consumers have full information we underestimate how sensitive consumers are to price changes (see also Draganska and Klapper, 2011, for a similar finding in an advertising model). This is intuitive: the price changes we—as econometricians—observe in the data are not observed by consumers who sample only a subset of the stores. A full information logit model assumes that all prices are observed, thus ascribing unresponsiveness to price changes to low price elasticity.

TABLE 12—OWN-PRICE ELASTICITIES THE DA VINCE CODE SEARCH VERSUS FULL INFORMATION MODEL

	Search		Full
	ε observed	ε unobserved	Information
Amazon	-1.870	-2.944	-0.522
Barnes and Noble	-4.350	-6.250	-1.160
Book Clubs	-2.315	-3.726	-0.656
Other	-7.100	-6.093	-1.267

So far we have assumed that the random utility term is observed by consumers before searching. To see how robust our estimates are to this assumption, and to capture uncertainty about book availability and other information that is unobserved before searching, we assume consumers do not observe prices and do not observe the random utility term ε_{ij} before visiting a store. Consumers do know the distribution of ε_{ij}, so calculation of $E\left[\max_{j \in S}\{u_{ij}\}\right]$ depends on expectations about prices as well as the random utility term.

As shown by Moraga-González, Sándor and Wildenbeest (2011), if we assume ε_{ij} follows a type I extreme value distribution we get the following closed-form expression for a consumer's expected maximum utility of searching all stores in subset S, conditional on prices:

$$(7) \qquad E\left[\max_{j \in S}\{u_{ij}\}\right] = \log\left(\sum_{j \in S}\exp\left[\mu_{ij} + X_i\beta_j + \alpha_i p_j\right]\right).$$

Since consumers' choice sets no longer depend on any realized values of the random utility term, ε_{ij} will be i.i.d. type I extreme value in the the conditional buying stage. As shown above, this means that also the conditional buying prob-

ability equation has a closed form solution, which facilitates the estimation of the model. Nevertheless, once we integrate out ε_{ij} the price distribution cannot be integrated out analytically from equation (4). We proceed by assuming consumers know each store's empirical price CDF and believe prices are random draws from these distributions. To integrate out these price distributions we randomly draw a price for each store from the corresponding empirical price CDF (after accounting for unobserved heterogeneity in the characteristics of the books by de-meaning the price observations), and calculate $E[\max_{j \in S}\{u_{ij}\}]$ using equation (7) for each choice set. For each observation we repeat this a number of times and use the mean as an estimate of $E[\max_{j \in S}\{u_{ij}\}]$

Column (3) of Table 9 gives the parameter estimates for this specification. The estimated price coefficients are very similar to those estimated for the main model in column (2). Search costs are higher—normalizing the estimated search costs by the price coefficients indicates that search costs are on average $4.14. As in the model with the random utility term unobserved ex ante most consumers prefer Amazon. Own-price elasticity estimates for the Da Vinci Code are shown in the second column of Table 12. Assuming the random utility term is unobserved before searching results in higher combined price elasticity estimates for all but the Other Bookstores.

D. Discussion

Our price elasticities provide an interesting comparison with the results of Chevalier and Goolsbee (2003), who found an own-price elasticity of -3.5 for Barnes and Noble and -0.45 for Amazon, using the very different methodology of investigating the effect of price changes on sales ranks of books. If we only look at a change in prices, assuming consumers' price expectations are unaffected, our estimated own-price elasticities are on average not very different for Amazon (between -0.1 and -0.7), but substantially lower for Barnes and Noble (between -0.3 and -0.9). If we take the effect of a change in price expectations into account as well, our estimated own-price elasticities for Amazon are mostly higher (between -0.7 and -2.3 across books), but still lower for Barnes and Noble (between -1.4 and -3.2). The difference between our findings may be due to several factors: first, Chevalier and Goolsbee's estimates are based on a much larger sample of books; our sample is restricted to several bestsellers. It is plausible that consumers are more price elastic when purchasing bestsellers (which could be used as "loss leaders" by bookstores to attract new customers), explaining the higher price elasticities for Amazon. Second, Chevalier and Goolsbee's results are based on 2001 data; whereas ours is based on a mix of 2002 and 2004 data. It is possible that online book shoppers have gotten somewhat savvier at searching for deals than they were in 2001. Third, our methodologies are quite different: whereas Chevalier and Goolsbee have the advantage of using exogenous price shocks, but are limited by lack of sales data (and have to extrapolate using a Pareto distribution), our method relies crucially on the specification of our demand model. We

hope that further research can identify data sets that can overcome the limitations of these two approaches.

IV. Conclusion

In this paper we have investigated to what extent consumers are indeed using sequential and fixed sample size search strategies put forth by the large theoretical literature on search behavior. By using detailed data on the browsing and purchasing behavior of a large panel of consumers, we have tested various restrictions that classical search models put on search behavior. We have shown that the benchmark model of sequential search, which assumes consumers know the distribution of prices, can be rejected based on the recall patterns observed in the data, even if there is a finite number of stores. In addition we do not find support for any within-store and across-store price dependence of search decisions—if consumers were searching sequentially, even in a setting with store differentiation, they would be more likely to continue searching when a relatively high price is observed.

Our finding that the fixed sample size search strategy outperforms the sequential search model in terms of explaining observed search behavior for the subjects in our sample is to some extent surprising given that fixed sample size search strategy is often thought of as a constrained version of sequential search. However, as shown by Morgan and Manning (1985) the optimal search model allows consumers to choose both the size of the sample and how many samples to take, and as such encompasses both the sequential and fixed sample size search models. When there is a large time lag between making the search decision and obtaining the actual quotation fixed sample size search is typically optimal because it allows the searcher to gather information more quickly than would have been possible with sequential search.

Although a typical online shopper will not face large time lags when searching, a fixed sample size search strategy might still be a good approximation of the optimal strategy if there exist economies of scale to sampling or if the searcher discounts the future. As argued by Manning and Morgan (1982), sufficiently large economies of scale from sampling will make it optimal to sample more stores at once and stop afterwards, even if the consumer can continue sampling. Indeed, after one has gone through the hassle of finding the right book and obtaining a price quote at one online bookstore, simply copying and pasting the ISBN number of the website of another bookstore is enough to obtain an additional price quotation.

Finally, we have explored the quantitative implications of our favored model by estimating the price elasticities implied by the fixed sample size search demand model. According to our estimates consumers are more price sensitive in this search model than in a model that assumes consumers have full information. Moreover, depending on the exact specification of the search model and whether price expectations remain constant, our findings indicate mostly higher price elasticities for Amazon but lower price elasticities for Barnes and Noble than found

in Chevalier and Goolsbee (2003). In Section III.D we discuss several factors that may explain the differences in results.

Our application has two important limitations: price differences in the online book market are small and the online book market is dominated by two booksellers. This means our conclusions may not generalize to other settings. We nevertheless hope that this exercise demonstrates the usefulness of the consumer search model as a "demand-side" model that could be applied in environments where consumer search is deemed an important factor.

REFERENCES

Axell, Bo. 1977. "Search Market Equilibrium." *Scandinavian Journal of Economics*, 79: 20–40.

Brown, Meta, Christopher J. Flinn, and Andrew Schotter. 2011. "Real-Time Search in the Laboratory and the Market." *American Economic Review*, 101: 948–74.

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

Camerer, Colin. 1995. "Individual Decision Making." In *Handbook of Experimental Economics.* , ed. John H. Kagel and Alvin E. Roth, 587–703. Princeton: Princeton University Press.

Carlson, John A., and R. Preston McAfee. 1983. "Discrete Equilibrium Price Dispersion." *Journal of Political Economy*, 91: 480–493.

Chade, Hector, and Lones Smith. 2005. "Simultaneous Search." http://www.public.asu.edu/ hchade/papers/ectrawp.pdf.

Chade, Hector, and Lones Smith. 2006. "Simultaneous Search." *Econometrica*, 75: 1293–1307.

Chen, Xiaohong, Han Hong, and Matthew Shum. 2007. "Nonparametric Likelihood Ratio Model Section Tests Between Parametric Likelihood and Moment Condition Models." *Journal of Econometrics*, 141: 109–140.

Chevalier, Judith, and Austan Goolsbee. 2003. "Measuring Prices and Price Competition Online: Amazon and BarnesandNoble.com." *Quantitative Marketing and Economics*, 1: 203–222.

De los Santos, Babur I. 2008. "Consumer Search on the Internet." NET Institute Working Paper #08-15.

Draganska, Michaela, and Daniel Klapper. 2011. "Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data." *Journal of Marketing Research*, 48.

Harrison, Glenn W., and Peter Morgan. 1990. "Search Intensity in Experiments." *Economic Journal*, 100: 478–486.

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

Honka, Elisabeth. 2010. "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry." http://home.uchicago.edu/~ehonka/Paper_EHonka_100310.pdf.

Janssen, Maarten C. W., and José Luis Moraga-González. 2004. "Strategic Pricing, Consumer Search and the Number of Firms." *Review of Economic Studies*, 71: 1089–1118.

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

Kogut, Carl A. 1990. "Consumer Search Behavior and Sunk Costs." *Journal of Economic Behavior and Organization*, 14: 381–392.

Koulayev, Sergei. 2009. "Estimating Demand in Search Markets: The Case of Online Hotel Bookings." Federal Reserve Bank of Boston Working Paper No. 09-16.

Manning, Richard, and Peter Morgan. 1982. "Search and Consumer Theory." *Review of Economic Studies*, 49: 203–216.

McCall, John J. 1970. "Economics of Information and Job Search." *Quarterly Journal of Economics*, 84: 113–126.

Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan. 2003. "Price Uncertainty and Consumer Search: a Structural Model of Consideration Set Formation." *Marketing Science*, 22: 58–84.

Moraga-González, José Luis, Zsolt Sándor, and Matthijs R. Wildenbeest. 2011. "Consumer Search and Prices in the Automobile Market." http://www.kelley.iu.edu/mwildenb/searchautomobiles.pdf.

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

Mortensen, Dale T. 1970. "Job search, the Duration of Unemployment and the Phillips Curve." *American Economic Review*, 60: 847–862.

Reinganum, Jennifer F. 1979. "A Simple Model of Equilibrium Price Dispersion." *Journal of Political Economy*, 87: 851–858.

Roberts, John H., and James M. Lattin. 1991. "Development and Testing of a Model of Consideration Set Composition." *Journal of Marketing Research*, 82: 429–440.

Rob, Rafael. 1985. "Equilibrium Price Distributions." *Review of Economic Studies*, 52: 487–504.

Rosenfield, Donald B., and Roy D. Shapiro. 1981. "Optimal Adaptive Price Search." *Journal of Economic Theory*, 25: 1–20.

Rothschild, Michael. 1974. "Searching for the Lowest Price when the Distribution of Prices is Unknown." *Journal of Political Economy*, 82: 689–711.

Schotter, Andrew, and Yale M. Braunstein. 1981. "Economic Search: an Experimental Study." *Economic Inquiry*, 19: 1–25.

Sonnemans, Joep. 1998. "Strategies of Search." *Journal of Economic Behavior and Organizaton*, 35: 309–332.

Stahl, Dale O. 1989. "Oligopolistic Pricing with Sequential Consumer Search." *American Economic Review*, 79: 700–712.

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

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