Search and Product Differentiation at an Internet Shopbot

Erik Brynjolfsson,* Astrid A. Dick† and Michael D. Smith‡

July 2004

Abstract

Price dispersion among commodity goods is typically attributed to consumer search costs. We explore the magnitude of consumer search benefits and costs using a data set obtained from a major Internet shopbot. For the median consumer, the benefits to searching lower screens are \$6.55 while the cost of an exhaustive search of the offers is a maximum of \$6.45. We are also able to estimate price elasticities and find that they are relatively high compared to offline markets, with a decrease in demand of 7 to 10 percent for each percentage increase in price, in our base model. Interestingly, in our setting, consumers who search more intensively are *less* price sensitive than other consumers, reflecting their increased weight on retailer differentiation in delivery time and reliability. Our results demonstrate that even in this nearly-perfect market of the shopbot, substantial price dispersion can exist in equilibrium from consumers preferences over both price and non-price attributes.

*Sloan School of Management, Massachusetts Institute of Technology, erikb@mit.edu.
†Federal Reserve Bank of New York, astrid.dick@ny.frb.org.
‡H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, mds@cmu.edu.
Any views expressed do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System. We especially thank Susan Athey and Nancy Rose for their useful and careful comments. We are also grateful to Patrick Bajari, Gregory Crawford, Ali Hortaçsu, Andrew Sweeting and participants at the 2003 Quantitative Marketing and Economics Conference for their comments and suggestions. Any errors are our own.

1 Introduction

Price dispersion in a market is typically attributed to imperfect information and consumer search costs. Low consumer search costs are one of the most frequently discussed aspects of Internet markets both in the academic literature and the popular press. On the Internet, consumers can discover prices and product offerings from competing retailers much more easily than they could in a comparably conventional environment.

Nonetheless, empirical research has often found a high level of price dispersion across Internet retailers. For instance, Clay et al. (2002) find a price dispersion of 27 percent for a random selection of hardcover books and 73 percent for paperback bestsellers. Similarly, Brynjolfsson and Smith (2000) find that Internet retailer prices differ by an average of 33 percent for books and 25 percent for CD's. These findings contrast with the classic "Law of One Price" in spite of the fact that the underlying products being compared are homogeneous and the marginal costs of the books are essentially identical across retailers (Brynjolfsson, Hu and Smith 2003).

Building on Stigler's (1961) claim that "Price dispersion is a manifestation — and, indeed it is a measure — of ignorance of the market," consumer search theory provides a class of models which can explain equilibrium price dispersion in the presence of consumer search costs. According to the standard economic theory, price dispersion arises when individuals are not perfectly informed about the prices or qualities of what is being sold (see, for example, Butters (1977), Varian (1980), Burdett and Judd (1983), Rob (1985), Stahl (1989)). As information is usually costly to gather, a buyer will stop searching for better deals as soon as the anticipated price reduction falls short of her cost of search. Applying the predictions of these models to markets with low search costs, Sorensen (2001) notes:

> "An intuitive result that arises from such models is that exogenous increases in consumers' propensities to search (for instance, due to a decrease in search costs) will constrain prices to be lower and less dispersed."

The economic prediction of these models regarding consumers weighing the costs and benefits of search has usually been difficult to test. The empirical work on search in the economics literature includes Sorensen (2000) and Sorensen (2001). Sorensen (2000) finds that patterns in price dispersion across prescription drugs are consistent with the predictions of a search model, as repeatedly purchased prescriptions show lower dispersion and price-cost margins. In subsequent work, Sorensen (2001) uses data on retail pharmacy transactions to make inferences about search for prescription drugs, based on a structural model, and finds that search intensities are generally low and are higher for maintenance medications.

However, search costs are not the only possible explanation for significant price dispersion. In the studies of Internet price dispersion mentioned above, while the products themselves are identical across retailers, they differ in regard to retailer-level attributes such as shipping service, product availability, return policies, and retailer reputation. Even if search costs were zero, consumers might be willing to pay different prices for different sets of these characteristics.

Our research uses a flexible demand model to estimate consumer search benefits and costs among users of a major Internet shopbot. Shopbots are Internet services that allow consumers to easily compare prices and product offerings among competing retailers. At a shopbot's site, a consumer places a product search for a unique product and obtains a list of the retailers' offers with price as well as other attributes such as shipping time and product availability displayed in a tabular format. The consumer evaluates these offers and makes a selection by "clicking" on a particular offer.

These shopbots may represent a particularly important service for Internet markets. The increasing use of shopbots, or online comparison-shopping services, should dramatically reduce consumer search costs in markets where they are available (Brown and Goolsbee 2000). Stigler (1961) highlighted the important role of organizations that specialize in the collection and dissemination of product and price information. Likewise, Bakos (1997) observed that electronic marketplaces, such as those facilitated by shopbots, are likely to become increasingly pervasive, with significant effects on buyer and seller welfare.

Our data contain 10,627 actual consumer searches for books offers over a 12 month period resulting in 460,814 separate retailer offers. In the data, we observe what offers the consumer was shown, the position of the offers on the consumer's screen, and how the consumer responds through their observed selection of offers. By focusing on books, a homogeneous physical product, we are able to eliminate product heterogeneity and only focus on heterogeneity across retailer service characteristics such as reputation, return policies, and shipping services. We are also able to obtain data on what consumers are observing and some of the actions they take, data that would be difficult to obtain in a conventional environment: Do they click on lower screens by scrolling down? Do they re-sort the data by shipping time, availability, or other characteristic? Do they sequentially click on multiple alternative offers before choosing one?

Taking advantage of the format of our data, we are able to estimate consumer benefits to search as well as an upper bound for search costs. The former is based on the comparison of the welfare generated by the first set of offers shown to the consumer in the default screen, and that generated by the entire set of offers, which the consumer could inspect by scrolling to lower screens and clicking on offers in these screens. For those consumers that do scroll down in our data, this represents an upper bound to their search costs. We use a compensating variations approach to calculate the welfare, based on the estimates of consumers' marginal utilities. While Sorensen (2001) offers a structural methodology for estimating search costs as an economic primitive, our methodology, while not allowing for the simulation of a consumer's search behavior, is straightforward to implement by comparison.

Our random coefficients model estimates imply that the benefits to searching lower screens are \$6.55 for the median consumer, while the cost of carrying a more exhaustive search of the offers is a maximum of \$6.45 for the median consumer that we observe chooses to search lower screens. Using a very different methodology, Bajari and Hortaçsu (2003) quantified the implied cost of entering an eBay auction to be \$3.20. Given a price dispersion of approximately \$11 in our data, as measured by the average standard deviation of the total price within a session, search costs represent about a maximum of sixty percent of this price dispersion for the median consumer. Since we can only identify consumer behavior based on actual click-throughs, and therefore do not observe their choice sets perfectly, we introduce a measure of "distance" to control for how far down the list an offer appears in a given session. Thus, we can make the choice set as large as possible, by including all listed offers, even if the consumer only saw a fraction of these offers, since the distance measure should allow each offer to be given the appropriate weight in terms of the consumer's decision.

Interestingly, in contrast to most search models, we find that increased search is associated with *reduced* price sensitivity in this setting. This search is driven by non-price factors, given that the first offer on the screen is the one with the lowest price. Across the various consumer types, we find, on the one hand, that consumers who spend the least are the most price sensitive. Consumers that scroll down to search lower screens, on the other hand, have lower price sensitivity. Instead, brand appears to play a relatively important role for them. Presumably they choose to inspect lower screens because they care relatively more about this and other attributes besides price. Similarly, consumers that take the time to re-sort offers as well as those that inspect several offers before choosing one are particularly sensitive to brand and less sensitive to price. Our results suggest that even when seemingly homogenous products are considered, non-price factors can be very important.

Our work is related to several earlier papers on consumer search costs. We use a similar dataset to Smith and Brynjolfsson (2001) (hereafter S&B) but depart from their work in several ways. First, while S&B study the importance of brand in the Internet, we model and estimate search costs and benefits. Second, we explicitly explore consumer heterogeneity and its implications for consumer behavior based on observable behavior of consumer across offers and screens. Third, this paper uses the random coefficients model, which allows for complex demand patterns. Fourth, while S&B use a sample for a period of roughly two months, our data covers a period of over twelve months, providing us with a richer set of options for our empirical work. Lastly, we introduce an easily implemented methodology for inferring search benefits and costs.

Our work is most closely related to recent studies analyzing how information environments impact consumer choice. Lynch and Ariely (2000) use an experiment to show that lowering search costs for quality information (as opposed to price information) at a simulated online store reduces consumer price sensitivity. Diehl, Kornish, and Lynch (2003) conduct a similar experiment where heterogeneous options are ordered according to the consumer's quality preferences. Thus, products listed earlier in the set of offers have a better fit to the consumer's preferences but by searching to lower offers consumer's can gain more price information. The authors find that in this choice environment consumers select lower priced offers than they would in an environment where offers are presented without regard to fit or price.

Our research is also related to recent studies analyzing customer behavior in online markets. Johnson, Bellman, and Lohse (2002) use MediaMetrix data to show that the time consumers spend on web sites declines with experience and that the sites with the fastest declines also have the highest customer loyalty. Baye, Morgan and Scholten (2002) find that various identifiable sources of firm heterogeneity can account for some, but not all, of the observed price dispersion in their sample of 36 online markets. Finally, Ellison and Ellison (2001) use shopbot data to analyze consumer price elasticity and retailer obfuscation strategies. They find evidence both of extraordinarily strong price competition and strategies on the part of retailers to increase consumer search costs. Our approach differs from these papers in that we use observed consumer choice behavior to place explicit bounds on consumer search costs.

Our research is also related to recent empirical studies that analyze Internet market behavior. Brynjolfsson, Hu and Smith (2003) use sales rank data to infer sales volume for specific products and find that the ease of accessing the greater product selection at Internet book retailers has generated approximately one billion dollars of consumer surplus each year. Chevalier and Goolsbee (2002) apply a similar method to compare specific products at Amazon.com and Borders and from this estimate own and cross price elasticity for the two retailers. They show that Barnes and Noble faces much stronger competition from Amazon than Amazon does from Barnes and Noble. Brown and Goolsbee (2000) use survey data to show that the introduction of shopbots for life insurance products placed significant price pressure on products listed by these shopbots. Brynjolfsson and Smith (2000a) collect data on a matched set of products across Internet and conventional channels to show that Internet markets appear to be more efficient with regard to price and menu costs, but that significant levels of price dispersion persist in spite of presumed low search costs. Indeed, they find that by several measures, price dispersion is at least as high across Internet retailers as across conventional retailers. Other authors have studied retailer differentiation strategies (Clay et al 1999) and price discrimination strategies (Clemons et al 2002). However, none of these studies estimates search costs from direct observations of consumer choice behavior.

The rest of the paper is organized as follows. Section 2 presents the data and the empirical framework, including the discussion of consumer heterogeneity in our sample, the model of consumer behavior which is taken to the data, and how we identify search costs. Section 3 presents a brief review of the literature and introduces a simple theoretical framework of analysis. Section 4 presents results for both the logit and random coefficients model. Section 5 presents the implied price elasticities, and finally the estimated search benefits and costs. Lastly, we provide some concluding remarks in Section 6.

2 Data

2.1 Description of the data

The data used in our analysis come from DealTime.com,[1] a prominent online comparison-shopping service.[2] See Figure 1 for a sample screen shot. As noted above, this dataset is similar to the one used by S&B. However, while S&B use data for a period of 69 days during August 25 to November 1, 1999, this paper uses a sample covering a period of over 12 months, roughly from September 1999 to September 2000. The sample is restricted to the top 100 bestselling books, as opposed to

[1]Formerly EvenBetter.com, acquired by DealTime.com on May 19, 2000. In July 2000, DealTime was the 13th most popular site among online retail shopping sites (with Amazon.com at the top), with 3.5 million of unique monthly visitors (4% of total web users). During this time, DealTime was also the most popular price-comparison site (with more unique visitors than MySimon.com and PriceScan.com, for instance).

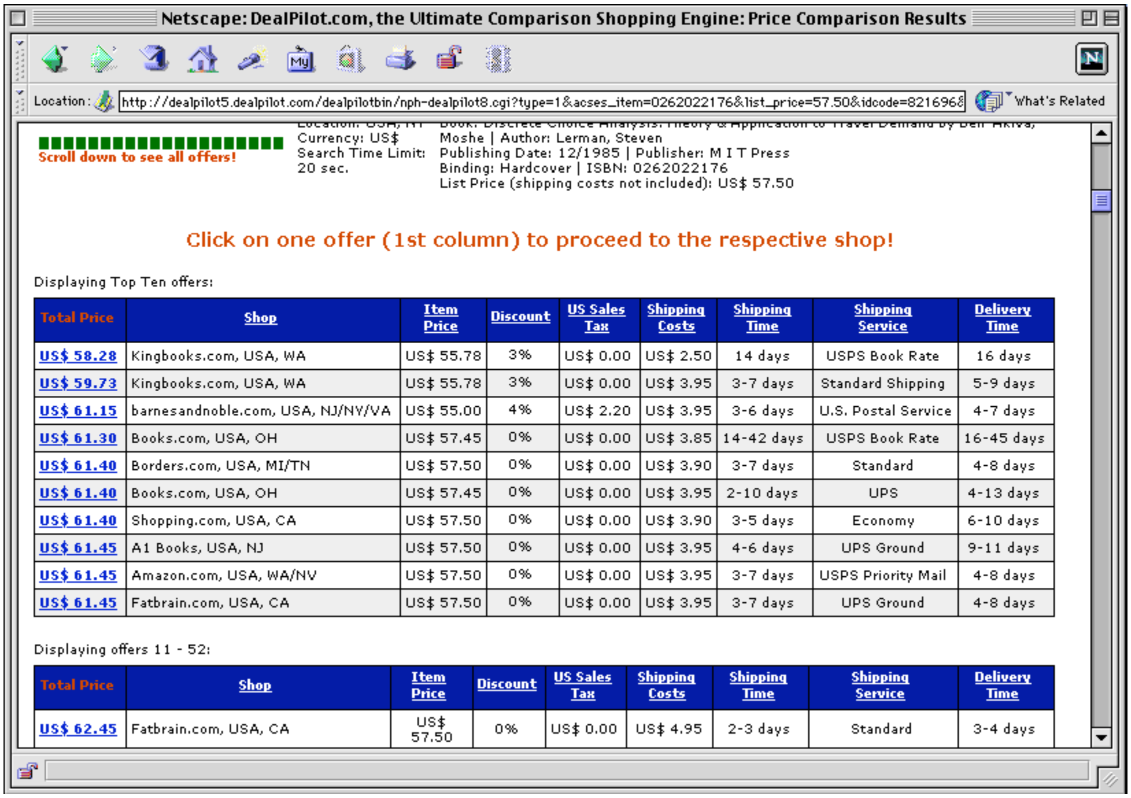
[2]Shopbots are free Internet-based services that provide a comparison-shopping search tool that presents prices of an item, as well as other product attributes, from various competing retailers. Shopbots have been changing through time, however, as they have gone from more objective presentation of price data to listing only products from companies that pay to be included in the search, or favoring retailers that advertise on their sites or pay a premium for a logo. Smith (2001) reviews the academic literature relating to shopbots.

all searches carried on the shopbot (see Table 1 for a list of the top ten books in our dataset). To facilitate the use of our choice model we focus here on the sub-sample of U.S.-based customers, sessions that lead to at least one click-through by the consumer, and searches that return more than one retailer. Even with these restrictions, we are able to observe 10,627 book searches or sessions, with roughly 460,000 total offers by retailers. The maximum number of offers any consumer is presented in a single search session is 67, including multiple offers by some retailers, for instance if they offer multiple shipping options. Table 2 shows the retailers and their shares of last clicks. There are 46 distinct retailers present in the dataset. During the sample period, we observe the behavior of 7,042 consumers.

In order to place a search, consumers must first choose the specific book they are interested in buying, which reduces their selection to a unique and physically homogeneous product, leaving item variation solely in terms of the retailers' conditions for price, shipping and product availability. When a consumer initiates a search, DealTime looks for offers for this selection in real time from a large set of retailers which account for the vast majority of books sold online. It displays the resulting price and product information to the consumer in a tabular format (Figure 1). The information displayed to the consumer through DealTime is the same that the consumer would obtain were she to go directly to the retailer's web site. Once a specific book offer is chosen by the consumer, she enters her country and state location in order for applicable taxes to be calculated. The attributes of the product offers include item price for the underlying book, sales tax, shipping costs, shipping time and service, delivery time, and total price. Up to ten offers fit on a single screen, and the offers are ranked by total price (item price plus shipping and taxes), from lowest to highest. By clicking on an offer, the consumer is taken directly to the retailer's web site to finalize the purchase. Our data include all the above information, as well as all consumer clicks, and whether the consumer sorts on a column other than total price (the default ordering).

It is important to note that in our data we only observe click-throughs as opposed to actual purchases. In related research Brynjolfsson and Smith (2000b) find that the factors that drive traffic to a site are also good predictors of sales at the retailer level.[3] However, a conservative interpretation of our approach is as a model of click-throughs and not of sales per se. If the consumer clicks on multiple offers, we use the offer she clicks on last as an indicator of her final choice.

[3]Based on information provided by DealTime.com, the sale/click ratio was not only very similar across retailers, but also the conversion rate between actual sales and clicks was approximately 50 % within our sample period.


Netscape: DealPilot.com, the Ultimate Comparison Shopping Engine: Price Comparison Results

Location: http://dealpilot5.dealpilot.com/dealpilotbin/nph-dealpilot8.cgi?type=1&acses_item=0262022176&list_price=57.50&idcode=821696

What's Related

Scroll down to see all offers!

Currency: US$
Search Time Limit: 20 sec.

Moshe | Author: Lerman, Steven
Publishing Date: 12/1985 | Publisher: M I T Press
Binding: Hardcover | ISBN: 0262022176
List Price (shipping costs not included): US$ 57.50

Click on one offer (1st column) to proceed to the respective shop!

Displaying Top Ten offers:

Total Price	Shop	Item Price	Discount	US Sales Tax	Shipping Costs	Shipping Time	Shipping Service	Delivery Time
US$ 58.28	Kingbooks.com, USA, WA	US$ 55.78	3%	US$ 0.00	US$ 2.50	14 days	USPS Book Rate	16 days
US$ 59.73	Kingbooks.com, USA, WA	US$ 55.78	3%	US$ 0.00	US$ 3.95	3-7 days	Standard Shipping	5-9 days
US$ 61.15	barnesandnoble.com, USA, NJ/NY/VA	US$ 55.00	4%	US$ 2.20	US$ 3.95	3-6 days	U.S. Postal Service	4-7 days
US$ 61.30	Books.com, USA, OH	US$ 57.45	0%	US$ 0.00	US$ 3.85	14-42 days	USPS Book Rate	16-45 days
US$ 61.40	Borders.com, USA, MI/TN	US$ 57.50	0%	US$ 0.00	US$ 3.90	3-7 days	Standard	4-8 days
US$ 61.40	Books.com, USA, OH	US$ 57.45	0%	US$ 0.00	US$ 3.95	2-10 days	UPS	4-13 days
US$ 61.40	Shopping.com, USA, CA	US$ 57.50	0%	US$ 0.00	US$ 3.90	3-5 days	Economy	6-10 days
US$ 61.45	A1 Books, USA, NJ	US$ 57.50	0%	US$ 0.00	US$ 3.95	4-6 days	UPS Ground	9-11 days
US$ 61.45	Amazon.com, USA, WA/NV	US$ 57.50	0%	US$ 0.00	US$ 3.95	3-7 days	USPS Priority Mail	4-8 days
US$ 61.45	Fatbrain.com, USA, CA	US$ 57.50	0%	US$ 0.00	US$ 3.95	3-7 days	UPS Ground	4-8 days

Displaying offers 11 - 52:

Total Price	Shop	Item Price	Discount	US Sales Tax	Shipping Costs	Shipping Time	Shipping Service	Delivery Time
US$ 62.45	Fatbrain.com, USA, CA	US$ 57.50	0%	US$ 0.00	US$ 4.95	2-3 days	Standard	3-4 days

Figure 1: Dealtime Sample Comparison Screen

Table 3 shows summary statistics for the variables used in our analysis based on the total number of offers. Total price is defined as the sum of the item price, the shipping charge, and applicable taxes. Delivery time is provided by the retailer, and reflects both shipping time and what we call acquisition time, the time it takes the retailer to get the item out of the warehouse before it can be shipped, which includes product availability.[4] Given that the retailer sometimes provides a time range for delivery, we construct the variable "average delivery time" which transforms this range into a single number, by taking the maximum and the minimum in the range and averaging them. "Delivery time not available" is an indicator variable for whether the retailer specifies a delivery time, which takes the value of one if the retailer did not provide the information. This variable basically reflects the item's availability, since the shipping time, the other component of delivery time, is always known.[5] The variable "Big three retailers" refers to an indicator variable we construct for whether the offer is for one of the following large, well-known retailers: Amazon,

[4]Due to the item's availability component, the average delivery time variable sometimes may take on very large values.

[5]Note that if delivery time is not provided by the retailer, we set the delivery time equal to the shipping time, since this is all the information the consumer has. The indicator variable for whether delivery time is not available should capture the additional effect that the lack of information has on the consumer's decision to choose the offer.

Barnes & Noble and Borders. Screen number indicates on what screen the offer is listed, taking on the value of 1 if the offer is listed within the top ten offers (the default screen), 2 if the offer is listed within the 11th to the 20th offer on the screen following the default screen, and so on. See the Appendix for further description of the variables.

Given that our sample is restricted to consumers who choose to use DealTime in their book search, our analysis might be interpreted as being only applicable to consumers conditional on them choosing this shopbot first. The reason is that customers who elect to go through a shopbot, and in particular DealTime, might be systematically different from other consumers.

2.2 Consumer heterogeneity across shopbot consumers

Consumer behavior is analyzed following the discrete choice literature. In our setup, a consumer places a search for a specific book on the shopbot's site and is then presented with several offers from various retailers, with several characteristics, such that as researchers we know a lot about what the consumer sees before making choices.

On the shopbot screen, offers are sorted according to price, from lowest to highest. Yet what is remarkable in our data is that about half of the consumers do *not* click on the offer with the lowest price. While the underlying good is homogeneous, the final product, including the bundled retailer services, is perceived as a differentiated product by the consumer. Clearly, especially in the case of consumers that choose to go to a shopbot as opposed to going directly to the retailer's site, total price should be an important component of the purchase decision. However, the consumer is expected to care about the overall utility derived from the final product, or the sum of the weighted product attributes, given the valuable diversity in terms of non-price attributes such as reliability, availability, and delivery time (Smith and Brynjolfsson 2001).

While shopbot consumers share some common characteristics, there are differences among them observable to the researcher. Based on what we observe consumers do in the shopbot, we can identify the following potentially distinct consumer groups: (i) those that click only on offers on the default screen, that is, the screen the consumer is shown after a search, which we call *first screen* consumers; (ii) those that scroll to lower screens by clicking on offers situated past the default screen, which we refer to as *low screen* consumers; (iii) those that sort by a column other than the total price column (the default sorting), which we call *sorting* consumers;[6] and (iv) those that

[6] Shopbot consumers have the option to sort by any of the columns shown, including item price, shipping price, ship time, availability, type of shipping service, tax and retailer.

choose to inspect more than one offer, by clicking on them, which we call *multiple click* consumers. Presumably, the preferences of these consumer types follow different distributional properties. By dividing consumers into the four groups listed above, we control for these differences and carry a more nuanced analysis of demand for books bought online.

In order to make inferences about search benefits and costs, it is appropriate to take into account consumer heterogeneity as much as possible. If a consumer chooses to scroll down the screen, this might not only be reflective of the search cost differences across consumers, but also of differences in preferences as well. For instance, heavily branded retailers (i.e., Amazon, Barnes and Noble, and Borders), often appear only on lower screens: 29 percent of sessions have no branded retailer in the top ten offers (the first screen), and an additional 9 percent have only 1, with half of the sessions having no more than two branded retailers in the first screen. If brand, for instance, is important for a given consumer she might choose to search offers on lower screens if the option she is looking for is not available on the first screen. As a result, *ceteris paribus* she might not only be less sensitive to price, but also be more sensitive to branding. These facts highlight the possibility that more intensive search may be motivated by a desire to locate products with attributes other than merely low price, a scenario often ignored, if not actively ruled out, in much of the theoretical work on search costs.

Table 4 reports how consumers are distributed based on their behavior at the shopbot. First screen consumers represent the majority, with almost 91 percent of the sessions falling in this category, while low screen consumers represent the remaining 9 percent.[7] However, as mentioned earlier, in spite of the products being identical within a session as books with the same ISBN, most consumers do not choose the offer which has the lowest total price and is listed at the top of the Dealtime session screen. Sixteen percent of the sessions have consumers click in more than one offer and less than one percent of consumers choose to sort by a column other than the default column of total price.[8]

The main specification in our analysis is the random coefficients model. This flexible model

[7]Note that 13 percent of first screen consumers click on multiple offers (within the first screen), while 61 percent of low screen consumers click on multiple offers. Within the group of low screen consumers, last clicks are concentrated within the second screen (46 percent), with the rest mostly divided up among lower screens.

[8]Within the group of consumers that sort, the majority (55 percent) sort by item price. Around 20 percent sort by product availability, 11 percent by retailer, 7 percent by number of shipping days, and the remaining 7 percent sort by tax, shipping price or shipping service. Given the scant number of sessions where consumers sort, we pool these observations together as opposed to dividing the group further by what consumers chose to sort on. It is interesting that sorting customers choose to sort mostly by item price. This might suggest that these consumers care not only about the total price they will need to pay, but also about how that price is apportioned to components like item price, shipping cost, and taxes. Smith and Brynjolfsson (2001) reported precisely such an effect, and retailers themselves often tout "free shipping" and other partitions of pricing, in addition to the total price charged itself.

allows for tastes to vary across consumers through interactions of an idiosyncratic utility component with product characteristics. As a result, we not only control for consumer heterogeneity *a priori*, by dividing consumers according to some observed characteristics, but also allow for consumer tastes to vary within a given consumer group.

3 Analytical framework

3.1 Literature on search under product differentiation

If books with a given ISBN are completely identical, and the first shopbot offer always has the lowest price, then why do most people search beyond the first offer? Presumably, consumers care about the non-price retailer attributes bundled with the purchase of a given book.[9] In other words, they are looking for the best product fit possible, which is based on a multidimensional set of product characteristics.

Thus, the appropriate framework of analysis is one where consumers search under product differentiation, which in our setting comes primarily from retailer attributes such as service quality, reputation, and shipping policies. However, while the economics literature is well developed for both search and product differentiation, respectively, the former has focused on search under homogeneous products while the latter has mostly left search aside. Three important exceptions are Bakos (1998) and Anderson and Renault (1999) and Chen and Hitt (2003). Bakos (1998) introduces search costs to a version of Salop's (1979) unit circle model of spatial differentiation. In the absence of product differentiation, the model follows the standard predictions in the literature: lower search costs lead to more competition and lower prices. However, in the presence of product differentiation, these predictions reverse: the model predicts that as sellers reduce the cost of obtaining information related to the non-price features of the product, prices increase as consumers care less about price and search more to find the right fit. Chen and Hitt develop a model that includes both retailer differentiation and search costs (through shopbot use). They find that Bertrand competition only occurs when both retailer differentiation is eliminated and when all consumers use shopbots. They also show that price dispersion and price premiums charged by heavily branded retailers can increase with increasing shopbot use. Anderson and Renault (1999) draw on insight in Wolinsky (1986) to construct a model of price competition in the presence of search costs and

[9]In some settings, the act of searching (a.k.a., shopping) might have intrinsic utility, but, we find that implausible in this setting.

product differentiation, modeled through the discrete choice approach. In their model, prices are first high when consumers have a very low value for product diversity, as in Diamond (1971) where consumers do not search at all, and subsequently fall and then rise again as the taste for diversity becomes large enough and consumers engage in more search. The latter case coincides with Bakos' (1998) result.[10]

3.2 A taxonomy of search and product diversity

The literature provides predictions about consumer behavior for situations of imperfect information under homogeneous products as well as product diversity without search. Using these insights, we can develop a simple framework to analyze the behavior that we would expect from shopbot consumers who both search and care about product diversity.

Consumers are likely to be heterogeneous in both their taste for product diversity as well as their search costs. Product differentiation is at the heart of our analysis, with products defined over a multidimensional space. While some models assume homogeneous consumer search, differences in the costs of search among consumers is likely to be prevalent in reality, as people have different costs of time and different tastes for search. In our data, we see a range of consumer behavior in the shopbot that suggests heterogeneity in preferences — from consumers clicking on the first offer to those that click on multiple ones before deciding on an offer.

As far as search costs under product homogeneity, economic intuition and theory suggest that lower search costs lead to consumers increased search, as well as to lower prices, as the source of market power when products are homogeneous is derived only from the existence of imperfect information (see Stiglitz (1989) for a review of this literature). In terms of taste heterogeneity without search, the theory suggests that equilibrium prices increase as consumers value for diversity increases (Anderson, de Palma, and Thisse (1992), for instance). Here the source of market power is the intensity of preference for diversity.

Putting together these two dimensions of consumer heterogeneity can give us insight into consumer behavior when products are differentiated and information is imperfect. Thus, allowing consumers to be different in these two dimensions, we can expect the taxonomy shown in Figure 2. We can infer that for any given level of search costs, consumers should become less price sensitive

[10]Chen and Sudhir (2002) have an Internet model where they interact consumer search costs and targeted pricing predicting that competition may be reduced and prices may rise as consumer search and targeting becomes easier. Finally, in a recent paper, Kuksov (2003) develops a model in which search costs affect not only prices but product design, with firms increasing their differentiation in response to lower search costs as a way to avoid lower prices (and therefore sustaining price dispersion).

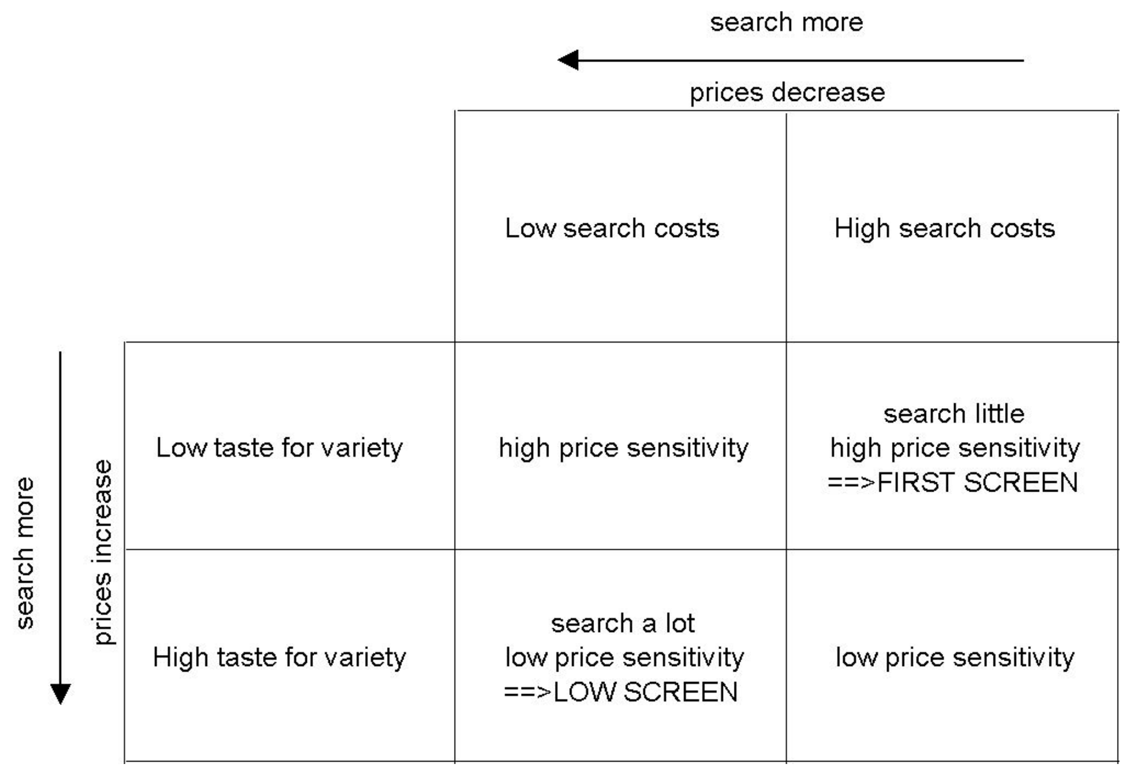


Figure 2: A Taxonomy of the Effects of Search Costs and Differentiation on Price Sensitivity

and search more as their taste for variety increases. For any given level of variety, consumers should search more as their search costs decrease.

Consumers with a low taste for non-price attributes should be the most price sensitive, and will search less the higher their search costs. First screen consumers are likely to fall within the category of consumers with low value for differentiation, especially those consumers that click on the first offer. Consumers with a high taste for non-price attributes should be the least price sensitive, given that they care a lot about other product characteristics. These consumers are willing to search more to find the right "fit," especially those that have low search costs. Low screen consumers are likely to fall within this category, though whether consumers with both low (high) search costs and taste for variety search a lot or a little will depend on the relative magnitudes of each.

Consumer types and their presumed behavior

The above taxonomy is useful in terms of allowing us to make certain predictions about how we expect consumers on the shopbot to behave. Low screen consumers, who search the most intensively and by revealed preference care about non-price attributes, are expected to have a lower price elasticity and a larger weight on branding. First screen consumers, on the contrary, should have a higher price sensitivity and a low preference for branding. Within this group, those consumers that choose the first offer should be the most price sensitive. Multiple click and sorting consumers, clearly, search for attributes beyond total price, so that retailer attributes should be important.

4 Empirical framework

4.1 Model

Base model

We start by introducing the base demand model for our analysis. Under the discrete choice framework, consumers are assumed to maximize an indirect utility function of the form

$$u_{ij} \equiv \delta_j + \epsilon_{ij} \equiv z_j\theta + \epsilon_{ij}, \tag{1}$$

where i stands for the consumer and j for the offer, $z_j = (p_j \; x_j)$ is a K+1-dimensional row vector of observed product characteristics, including the product price p_j and the observed product characteristics x_j, and ϵ_{ij} is a mean zero random disturbance. The K+1 dimensional vector $\theta = (\alpha, \beta)$ represents the taste parameters, where α is the coefficient on price. In particular, $z_j\theta = x_j\beta_j - \alpha p_j$.

Assuming an extreme value distribution for ϵ implies that the conditional choice probability is given by the logit formula:

$$P_j(\delta) = \frac{\exp(\delta_j)}{\sum_{r=1}^{J}\exp(\delta_r)} \qquad\qquad r = 1, ..., J. \tag{2}$$

Note that given the above assumption on the distribution of ϵ, the choice probabilities do not depend on individual characteristics.

Random coefficients choice model

The utility model above assumes additive separability between the two terms in (1), such that the δ term depends on product characteristics only, and the disturbance term solely on consumer characteristics. One implication from this utility model is that substitution patterns only depend on the δ_j's. The random coefficients model overcomes this limitation by allowing for the interaction of consumer heterogeneity and observed product characteristics. If the coefficient on observed product characteristic k is allowed to vary by consumer, θ_{ik}, where $\theta_{ik} = \theta_k + \sigma_k \nu_{ik}$, one obtains the random coefficients model where each individual may assign a different utility level to each observable product characteristic (or some of them). The indirect utility function takes on the following form:

$$u_{ij} \equiv z_j \theta + \sum_k \sigma_k z_{jk} \nu_{ik} + \epsilon_{ij} \tag{3}$$

where ν is i.i.d. across individuals and characteristics.

The choice probability of consumer i choosing offer j now becomes:

$$P_{ij} = \frac{\exp(\delta_j + \sum_k \sigma_k z_{jk} \nu_{ik})}{\sum_{r=1}^{J} \exp(\delta_r + \sum_r \sigma_r z_{jr} \nu_{ir})} \qquad r = 1, ..., J. \tag{4}$$

Therefore, unlike the basic multinomial logit, the choice probabilities depend on individual characteristics, which in terms of substitution patterns implies that consumers will substitute towards similar products (McFadden, 1984).

Price exogeneity assumption

Implicit in the above analysis is the assumption of the exogeneity of price. Given that we observe individual consumers making choices in what constitutes a micro-level or disaggregate dataset, price can be assumed to be exogenous given that a single household should have no impact on a retailer's price, or any other attribute for that matter.[11] However, this advantage of disaggregate demand models comes at the cost of being unable to carry out the type of counterfactual exercises that aggregate demand models allow for, since the assumption of price exogeneity is inadequate in a forecasting context where prices are determined by market forces (Goldberg, 1995).

Choice set and the "distance" measure

The choice set, or the set of offers that the consumer actually observes, is expected to vary across consumer types. However, our data do not allow us to identify exactly which offers the consumer sees before making a click-through. As a result we must infer what the most appropriate

[11] This could be violated if the price variation for a given book is mostly driven by demand factors, as opposed to supply factors, with the demand shock being common across individuals. In our setting, we expect the price variation to be driven mostly by cost and brand factors, as the total price used in our analysis includes shipping price (which should mostly reflect actual firm costs), there is large variation in product availability and retailers are highly differentiated, as our estimation results confirm. Moreover, when we include fixed effects for the retailers, the price coefficient does not appear to be affected in any significant way (results not shown).

choice set is for each consumer type. The key question is how far down on the list of offers can an offer be and still be part of the consumer's choice set. Here, we include in the choice set the full set of offers in the session, by adjusting for the relative unattractiveness of lower offers. Even if the consumer does not see all offers listed, including offers that are at the bottom of the list should be appropriate as long as one includes a measure of "distance" for how far away from the default screen that offer is. This should take into account the fact that the consumer finds a lower offer less appealing compared to an offer on the default screen, because of the cost of effort involved in scrolling down. In particular, we include as an additional retailer attribute the screen number where the offer appears listed. Including a distance measure is also useful because we avoid imposing assumptions on the consumer's priors about lower screens, and it diminishes the problem that we do not see consumer behavior unless she clicks on an offer. For the case of first screen consumers, this will actually turn out to be equivalent to estimating the model based on the first ten offers on the default screen, as opposed to the full choice set, since none of the lower screen offers are chosen by definition in the case of this consumer group.

4.2 Identification of search benefits and costs

When a consumer initiates a book search on the shopbot, she receives a list of the retailers' offers ranked by total price. Ten offers are shown to the consumer in the first screen, and the consumer must scroll down to see additional offers. We might expect consumers to find it somewhat costly to scroll down the screen in order to observe all offers, since this involves waiting time and cognitive effort for evaluating these offers (Shugan, 1980). Once the consumer inspects the first screen, she has knowledge about the attributes of the first ten offers, but is uncertain about the characteristics of the offers in lower screens unless she exerts additional effort. If shipping time or brand is important, a consumer might choose to look at lower screens if the options she is searching for are not available on the first screen and she expects to obtain an increase in utility that is large enough to at least cover the cost involved in scrolling down to lower screens.

Since we can only identify search behavior based on actual click-throughs, our key assumption is that individuals who do not click-through on lower screens — and therefore remain within the default screen in terms of our data — do not search lower screens either. While this might be violated for some consumers in the default screen group, we expect it to be true for the majority of consumers in this group, where over 80 percent of the consumers click-through on one of the first three offers listed. The latter makes it unlikely that these consumers scroll down to lower screens, do not click on any offers in these lower screens, and then return and click on a top offer. Given that offers are ranked by price, from lowest to highest, the consumer has knowledge about the prices of offers on lower screens, and the fact that they tend to click on the cheapest offer suggests that price is most important to them. Moreover, as will be seen later, our estimates in terms of these consumers' responsiveness to price and other attributes provide further support for our assumption.[12]

In this paper we exploit the nature of our data to estimate consumer benefits to search as well as an upper bound for search costs. The former is based on the comparison of the welfare generated by the first set of offers and that generated by the entire choice set of offers, which the consumer can inspect if she chooses to scroll down to lower screens. We use a compensating variations approach to calculate the welfare, based on the estimates of consumers' marginal utilities. Presumably, the reason why the consumer chooses to look at lower offers is that the expected utility gain is higher than the idiosyncratic search costs incurred in scrolling to lower screens, as she might get an offer that gives her greater utility.

Equivalent variation: Benefits to search

One economic prediction is that consumers weigh the costs and benefits of search when making search decisions. In other words, a buyer will stop searching for better deals as soon as the anticipated price reduction falls short of her cost of search (Stigler, 1961). In our setup, if the consumer chooses to scroll down, she believes that the expected gain in utility will be at least as large as the cost incurred in scrolling down, that is

$$Expected\ Utility\ Gain \geq Cost\ of\ Scrolling\ Down \tag{5}$$

One way to measure the utility gain is by computing the consumer welfare change from adding the full set of offers. Just as consumer welfare is enhanced when consumers can select from 2 million books at Amazon instead of only 40,000 books at a typical conventional store (Brynjolfsson, Hu and Smith, 2003), so is welfare enhanced when additional retailers' offers, with varying prices, shipping times and branding, are made available for any given book title. Following Small and Rosen (1981) and Trajtenberg (1989), in the context of the discrete choice model, the change in welfare from expanding the set to all offers is similar to measuring the changes in welfare from

[12] Note that if the assumption were violated, the interpretation of our estimates would still remain valid, except that we could no longer associate increased search with lower price sensitivity.

changes in the choice set between periods s and $s-1$ in some market t as the expected equivalent variation (EV) of the changes. The latter is defined as the amount of money that would make consumers indifferent, in expectation, between facing the two choice sets. This computation simply generalizes the methods of welfare economics to handle cases in which discrete choices are involved, representing a measure of compensating variation. Then, letting θ represent demand parameters, x product attributes, and p price, one has

$$EV = S_s(p_t, x_t; \theta) - S_{s-1}(p_t, x_t; \theta) \tag{6}$$

where

$$S(p, x; \theta) = \frac{1}{\alpha_j} ln[\sum_{j}^{J} exp(\delta_j(p_j, x_j; \theta))]. \tag{7}$$

In our setup, S_s represents the surplus generated by the entire set of offers, while S_{s-1} the surplus generated by the first screen offers, with the θ parameters being identified from the model based on the full choice set of offers. The coefficient on price, α, can also be interpreted as the marginal utility of income of the consumer, and it is used here to convert utility units into dollars. This equivalent variation represents the benefits the consumer would obtain if she chose to search.

Specifically, we measure the surplus generated to the consumer from the first screen offers, as well as the surplus generated from the full set of offers (all screens). The way the surplus is calculated will depend on consumer type, since first screen consumers should value price relatively more than brand, say, while consumers that scroll multiple screens should care relatively more about brand.

What is identified: Upper bound to search costs for consumers that scroll down

In the case of a consumer that we observe click in lower screens, the above measure of benefits represents an upper bound to search costs.[13] In particular, we are able to measure her *realized* gains from scrolling. Presumably, the reason why the consumer chooses to look at lower offers is that the expected utility gain is higher than the idiosyncratic search costs incurred in scrolling to lower screens. This gain, however, is higher than the average gain from search for all consumers who search low screens. The reason is that there are some consumers that scroll down to lower screens but do not click-through in lower screens, and as a result we have no way to identify the fact that they surfed lower screens. These consumers get zero benefit from searching. In other

[13]Note that search costs are inferred from the utility model we impose; they are not a free parameter.

words, our estimate is a conservative upper bound on search costs for consumers that we observe click on lower screens, since some zeros are omitted from the computation. Every consumer that clicks-through on a lower screen gets some benefit from searching, though this might be greater or less than what she expected. For multiple-click and sorting customers, it is possible to estimate this upper bound, based on the sub-sample of consumers that click on lower offers within each consumer category. Note that, by definition, there are no first screen consumers that scroll down.

The nature of this search is different from most prior analyses of search, where consumers care about finding a lower-priced product, all else equal. In our setup, consumers presumably perceive some degree of product differentiation and value other product attributes such as shipping and delivery time. This is confirmed by our results, presented later in the paper, and is consistent with the results in S&B.

5 Estimation results

5.1 Logit model results

We first explore consumer behavior using the multinomial logit model. Column (i) in Table 5 reproduces one of the main logit results of S&B, while column (ii) presents the results we obtain here with our sample under the same specification. As can be appreciated from the table, while our sample contains almost four times the number of observations of S&B and covers a different time period, the results are very similar.

The specification includes price broken up into the item price, shipping charge and tax. Product attributes include the average delivery time, whether the delivery time was provided by the retailer, and an indicator variable for whether the retailer belongs to the "big three" branded retailers: Amazon, Barnes and Noble, or Borders. These three retailers are well-known to consumers throughout the sample, and including this fixed effect should capture the intangible, non-contractible or unobserved retailer characteristics that play a role in the consumer decision. Note that the dependent variable takes on the value of one if the consumer picked the offer, and zero otherwise.

The point about brand deserves some explanation. Products sold over the Internet contain both contractible and non-contractible characteristics. They represent, from the perspective of the consumer, a kind of product bundle, including both an underlying product, as well as a service component provided by the Internet retailer, such as delivery and web site characteristics. Contractible aspects of the product bundle include attributes for which the consumer has clear avenues of recourse in the case the retailer defaults on any of them. In contrast, other characteristics, such as delivery time, are non-contractible. In the presence of non-contractible product characteristics, consumers may use a retailer's brand name as a proxy for their credibility in fulfilling their promises on non-contractible aspects of the product bundle (Wernerfelt, 1988).

As mentioned earlier, in order to examine consumer heterogeneity, we divide consumers into four groups based on some observable characteristics in our data. Column (i) of Table 6 reports results for the logit model for the entire sample. Column (ii) presents results for *first screen* consumers that only inspect offers in the first screen; column (iii) for *low screen* consumers that clicked at least once in offers in lower screens; column (iv) for *sorting* consumers that resorted the offers, and column (v) for *multiple click* consumers. Note that in these specifications we focus on total price which includes item price, shipping costs and tax, in order to keep our random coefficients analysis parsimonious in light of its greater computational demands. Also, note that the number of sessions for the sorting consumer group is very small relative to the other groups. Based on our earlier discussion, the choice set on which the model is estimated is the entire set of offers in the session.

As we would expect, ceteris peribus consumers in all groups are less likely to choose an offer with a higher total price. We find that first screen consumers have the highest coefficient on price, while low screen consumers who click on lower screens have the lowest. These results coincide with our expectations that search in this setting is generally motivated by non-price factors.

Consumers also value shorter delivery times. Sorting consumers present the largest responsiveness to delivery time (though only significant at the 10 percent level of confidence), while multiple click consumers have the lowest. Both low screen and multiple click consumers appear to have a strong taste for brand, as they depict significantly higher, positive demand effects on the big-three retailers indicator.[14] Again, this suggests the importance of non-price factors for search intensive consumers in our setting.[15] The small magnitude of the brand coefficient for first screen consumers contrasts with results for the rest of the consumer groups.

[14]Note that only three sessions of multiple click customers who surfed lower screens have consumers last click on an offer in the first screen (that is, where consumers click on at least one lower screen offer but then decide to go back up to the first screen to last click-through on an offer there). In the case of low screen consumers, 123 sessions have consumers check on offers on lower screens to then end up last clicking on an offer on the first screen. For these consumers, the realized gain of searching lower screens must have been lower than the expected gain. However, another possibility is that these consumers could be using the information provided on the product by lower screen offers (which could be of better quality) to then last click on a low price offer in the first screen. Nevertheless, the fraction of these sessions is rather small.

[15]Note that while most of multiple click consumers click on offers on the first screen (80 percent), most of these offers (54 percent) belong to the big three retailers, suggesting that multiple click consumers are sensitive to brand.

The coefficient on the indicator variable for whether delivery time is not available is negative, suggesting that not listing the delivery time has an adverse effect on the demand for a retailer. Using the coefficient on the average delivery time (measured in days) to interpret this negative effect on demand suggests that, on average, the value consumers put on delivery information not being provided is equivalent to about 4 additional delivery days. However, for low screen and sorting consumers this variable presents no statistical significance at reasonable levels of confidence.

It is worth noting that even first screen consumers — who we expect to be the most price sensitive, and that, indeed, present the highest coefficient on price — put significant weight on other attributes of the product-retailer bundle, highlighting the importance of retailer differentiation in this context.

Table 7 estimates the logit model adding the distance measure "screen number" that captures how far down the list each offer appears, thus taking into account the relative unattractiveness of lower offers given the effort involved in scrolling down. This measure also allows us to safely define the largest possible choice set even if a consumer did not see some of the offers in this choice set, since the screen number should give proper weighing to each offer. Note that in the case of first screen consumers, the model based on the entire set of offers and a distance measure is identical to the model without a distance measure estimated based on the restricted choice set including the top ten offers only. This is due to the fact that first screen consumers, by definition, never click on a lower screen offer.[16]

The coefficient on screen number is statistically significant for all the consumer groups shown on the table, and it is usually negative, such that offers that appear on lower screens are worse to the consumer than those that are on higher screens. The exception is the case of low screen consumers, where the coefficient is positive, since these consumers rarely go back to click on a first screen offer after having clicked on lower screens. These higher priced offers yield them more utility, as clearly they are looking for attributes that go beyond price. Overall, the estimations appears to be more precise and the coefficients are sometimes affected by the introduction of the distance measure, suggesting the importance of controlling for this variable when all offers are included as part of the choice set — something that we must do as we do not exactly know what screens the consumers actually looked at, unless they clicked on an offer on that screen.

[16]We present the latter model since it presents the appropriate R-squared and underlying set of offers, and does not include the distance measure which is insignificant (the rest of the coefficients are identical under both models).

5.2 Random coefficients model results

Tables 8 through 12 show results for the random coefficients specification, for the entire sample as well as for each consumer type. Note that each consumer is allowed to have an individual-specific marginal utility, as described in section 4. The results are robust to various optimization routines and are based on Halton draw sampling 125 individuals from a standard normal distribution.[17]

We present two specifications in each case. Both Model I and Model II allow all marginal utilities to differ across individuals, thus making the model flexible.[18] Model II, however, includes the distance measure "screen number," also allowed to vary across consumers.[19]

Across all specifications, the estimation results are consistent with the way we expect the coefficients to enter the indirect utility function. In Table 8, based on the entire sample, we find that consumers respond negatively to total price, as well as to delivery time and product availability. We also find a significant positive brand effect on demand for the big three retailers (Amazon, Barnes & Noble, and Borders), as before. All the random coefficients have a standard deviation that is significant at the one percent level of confidence. This suggests that it is appropriate to allow the marginal utilities to vary across consumers.

In terms of price sensitivity, it is low screen consumers who have the lowest absolute value coefficient, suggesting once again that they care relatively more about non-price attributes — evidenced by the high coefficient on branded retailers. Sorting and multiple click consumers, who also spend more time searching through offers, present similar coefficients. First screen consumers, who do not search, have the highest coefficient on price and put the lowest weight on brand.

These results are similar to our previous logit results. However, the random coefficients model allows for more reasonable substitution patterns. For instance, if there were a zero standard deviation on the distribution of marginal utilities of delivery time, then when a low delivery time retailer increases its price, consumers who substitute away from this retailer will do so proportionately toward all other retailers, regardless of their delivery time, as substituting consumers have the same

[17]Based on Train (1999), we use Halton draws instead of random draws, a type of what is known in the literature as "intelligent" draw, to save computation time. Train finds that the simulation variance in the estimation of random coefficients is lower with 100 Halton draws than with 1000 random draws, confirming earlier results in the literature. In our computations, we have benefited greatly by the estimation algorithm developed by Kenneth Train, David Revelt and Paul Ruud.

[18]The fully flexible model is our preferred specification — relative to a more parsimonious model where, say, only the price coefficient is allowed to vary across individuals — as it provides a better fit of the data, given that the estimates of the standard deviation of the distribution of tastes are usually significantly different from zero for all attributes.

[19]Once again, in the case of first screen consumers, model II is estimated with no distance measure but on the basis of the restricted choice set of the top ten offers, which, as measured earlier, is similar to including a distance measure and using the unrestricted choice set.

marginal utility as any other consumer. On the contrary, if the standard deviation on taste for delivery time were nonzero, as we find is the case here, when a low delivery time retailer increases its price, consumers who substitute away will do so towards other low delivery time retailers, as they originally showed a strong taste for low delivery time. The latter has to do with the way consumers decide on purchases by choosing the one which provides the highest utility: if a consumer found a low delivery time retailer to provide her with the greatest utility, on average this consumer will have a relatively large marginal utility for low delivery time.

Note that model II in all instances provides a better fit of the data, with the distance measure, represented by screen number, always significantly different from zero. Therefore, we choose model II as our preferred specification in the analysis that follows.

5.3 Price elasticities

Based upon the above estimates, one can obtain price elasticities, which will allow for the interpretation of the coefficient magnitudes.

Logit elasticities

Recalling that $z_j\theta = x_j\beta_j - \alpha p_j$, under the logit model, as defined in section 4, the price elasticity for offer j is

$$\eta_{jk} = \frac{\partial P_j}{\partial p_k}\frac{p_k}{P_j} = \begin{cases} -\alpha p_j(1-P_j) & \text{if} j = k \\ \alpha p_k P_k & \text{otherwise.} \end{cases} \tag{8}$$

Results for own-price elasticities are shown on table 13. We present various percentiles for the distribution of price elasticities across offers for the entire sample as well as for each consumer group and based on our two logit models. The median of the distribution of elasticities is -9.77 in the base model, and -6.75 when the distance measure is introduced, such that for a one percent increase in the retailer's total price, there is a reduction of nearly 10 and 7 percent in the retailer's demand, respectively. This is quite high compared to most offline markets, as might be expected given the ease with which consumers can compare prices at a shopbot like Dealtime. Ignorance and geography are virtually eliminated as barriers to price search. At the same time, the median price elasticity is significantly *less* than the elasticities found by Ellison and Ellison (2001) in their analysis of a shopbot for computer memory chips, where retailer differentiation is less evident. In results below, we further explore this finding.

Another important result is the large variation across consumer types. In particular, low screen consumers have a remarkably low price elasticity with a median of less than one. These results are directionally consistent with our expectations, as discussed in section 3.2, although the small magnitude of the price elasticity is notable. First screen consumers, as we would expect, present the highest price elasticities, with a median of -14.46 and -6.00, respectively. Thus, it appears that when we correct for the position of the offers on the screen, and therefore adjust for the fact that some low offers might not be seen by the consumer at all, the marginal utilities are affected significantly, suggesting that not making the adjusting can introduce distortions.

Random coefficients elasticities

As mentioned earlier, the flexibility of the random coefficients model has several advantages over the multinomial logit model. While the logit model is attractive due to its tractability, it imposes restrictions on the own- and cross-price elasticities (see McFadden, 1981, 1984; Berry, Levinsohn and Pakes, 1995). As we saw earlier, the price elasticities of the logit model are driven only by market shares. In the case of cross-price elasticities, for instance, this implies that if two retailers have similar market shares, whenever the price of a third retailer increases, consumers will substitute towards both retailers similarly, regardless of how far apart in the characteristics space the two retailers are located from each other. The random coefficients model allows for flexible price elasticities. Own-price elasticities in this model are driven by the different price sensitivities of diverse consumers, as opposed to the functional form assumptions of how price enters the indirect utility (additive separability). Cross-price substitution is driven by product characteristics, as the error term includes interaction between individual idiosyncrasies and characteristics.

In particular, the price elasticities derived from the random coefficients model introduced in section 4 are as follows:

$$\eta_j = \begin{cases} \frac{-p_j}{P_j} \int \alpha_i P_{ij}(1 - P_{ij}) dP(\nu) & \text{if} j = k \\ \frac{p_j}{P_j} \int \alpha_i P_{ij} P_{ik} dP(\nu) & \text{otherwise.} \end{cases} \tag{9}$$

where P_{ij} is the choice probability for consumer i for retailer j, as depicted in equation 4, $\alpha_i = \alpha + \sigma v_i$, $P(\nu)$ is the distribution of ν (which can be empirically estimated or imposed a priori as we do here), and

$$P_j = \int P_{ij} dp(\nu) \tag{10}$$

The elasticities implied by the random coefficients specification are reported in Table 14. This specification usually leads to substantially smaller elasticity estimates than the logit model, with the elasticity distributions being shifted to the right. Note that the low screen consumer elasticities are modified only slightly. This is not surprising, given that only the standard deviation on delivery time is significantly different from zero (at the ten percent level) in the random coefficients model. That is, the implicit assumption of the multinomial logit about the standard deviation of the taste distribution being zero, actually holds true under the random coefficients for price and for the indicator for whether delivery time is not available.

In the case of the entire sample estimates, while over a quarter of the elasticities are positive under the base model, only the tenth percentile is positive when the distance measure is included, suggesting once again the importance of introducing this measure. Some positive elasticities result from the fact that we let the individual characteristics be normally distributed, so that elasticities can take on any value. An alternative is to impose another distribution such as restricting the individual's marginal utility of price to take on negative values only. However, when we try imposing a log-normal distribution on the marginal utility of price, no convergence is obtained. Even if we did, however, we would be forcing the elasticities to be negative by imposing such distribution, and the validity of such an approach might be questionable.[20]

Note that the price elasticities tend to be more similar between the logit and the random coefficients model when the distance is included. The ordering of the magnitudes of the elasticities (based on the median) remains approximately the same under both the logit and the random coefficients. First screen consumers, as expected, have the highest price elasticity, with a median of around -5, while multiple and low screen consumers have the lowest price elasticity with a median of no more than -1. Also note that while some of the elasticities might appear to be low (especially those below unity), Chevalier and Goolsbee (2003) also find a low own-price elasticity of demand for Amazon.com — the most popular online retailer at the time of both theirs and our sample — of -0.45 during 2001.

[20]The fact that some percentage of the elasticities are positive is understandable given that the random draws are taken from a normal distribution so that, in principle, the tail of the distribution can take on positive values. This is also common in the results in the literature. Nevo (1997), for instance, finds as many as 13 percent of the price coefficients to be positive. It is through the flexible interactions with demographics, which we do not have as part of our data set here, that Nevo (2001) obtains, in subsequent, related work, a dramatic reduction in the positive price coefficients, to only 0.7 percent. As demographic data are included, the distribution of demographics, which is not normal, modifies the final coefficient distribution away from the normal.

5.4 Search benefits and costs

Using the random coefficients model estimates based on the entire sample and the equations in Section 4.2 above, we find that there is a median gain of \$6.55 from scrolling down to a lower screen. Table 15 shows the distribution of consumer welfare improvements from the full set of offers across various percentiles. As discussed earlier, a consumer derives this consumer welfare gain when choosing one alternative from the full set of offers at the shopbot. In the case of low screen consumers, who click on offers beyond the first screen, this welfare gain actually represents an upper bound for search costs. For these consumers, search benefits are rather high, given that their marginal utility of income, used here to adjust utils into dollars, is very low, as evidenced by the coefficient on price and their price elasticity.

Note that the search benefits of first screen consumers are rather low. This makes sense since these consumers appear to care mostly about price, such that having access to more offers, all of which are on lower screens and therefore more expensive, does not usually increase their utility. It is important to note that while search benefits of low screen consumers represent an upper bound to their search costs, the search benefits of first screen consumers represent a lower bound to their search costs — at least as long as we are willing to accept the assumption that first screen consumers never looked at lower screens (a fact that we cannot corroborate since we can only identify consumer behavior through actual click-throughs).

Table 16 presents various percentiles for the upper bound for various consumer groups. The estimates are based on the subset of consumers, within each consumer type (entire sample, low screen, sorting and multiple click consumers),[21] that scroll down to lower screens.

The estimates imply that the benefits to searching lower screens are \$6.55 for the median consumer, while the cost of carrying an exhaustive search of the offers is a maximum of \$6.45 for the median consumer that we observe chooses to search lower screens. Using a very different methodology, Bajari and Hortaçsu (2003) estimate the implied cost of entering an eBay auction — the effort spent on estimating the object's value and the opportunity cost of time spent bidding — to be \$3.20. This is an interesting point of comparison as we expect the consumer groups in both analyses to be "Internet savvy" and therefore somewhat similar.

Given a price dispersion of approximately \$11 for the median consumer in our data (measured by the average standard deviation of total price within a session), the median consumer search costs are equal to sixty percent of this price dispersion. Thus, search costs can only explain part of

[21] Note that, by definition, there are no first screen consumers that scroll down.

the observed dispersion of prices in our shopbot data. However, this level of search costs, combined with product differentiation, may be sufficient to sustain persistent price dispersion. For example, using aggregate data for the mutual fund industry, Hortaçsu and Syverson (2004) find that small search search costs and a large degree of product differentiation can sustain the large fee dispersion between S&P 500 index funds.

It is worth noting that consumer search costs online and offline might actually mean different things. We might expect consumers to be more willing to search when all retailers are just a click away than when they are a car ride away. Yet, we do not observe consumers searching as much in spite of the apparent ease (see also Johnson et al. 2004). One possible explanation for this is that while in a brick and mortar environment most search costs are costs of time — on which people might assign a low value — on the Internet most search costs are of a cognitive nature. At the shopbot, a consumer will not incur a large amount of time scrolling down, but rather she is going to have to think more — which presumably consumers dislike.

5.5 Discussion

The Internet has facilitated a substantial amount of consumer search by concentrating large amounts of data, particularly through shopbots. The data available at a shopbot is not only detailed in terms of price, but also on product features. This should have invariably decreased the cost of search for both prices and other product attributes. When products are homogeneous, economic theory suggests that as search costs decrease, prices should go down and become less dispersed. The fact that price dispersion has been documented to be high in the Internet (Brynjolfsson and Smith, 2000a) suggests the importance of product diversity sold on the Internet (Smith, Bailey, and Brynjolfsson 2000). In terms of the current findings, if a significant portion of retailers' traffic is driven by shopbots, therefore affecting their pricing decisions, our results are generalizable beyond the shopbot. In June 2000, for instance, a quarter of Amazon.com's traffic was driven by DealTime.com, suggesting that shopbot activity might be important in determining the market's equilibrium price dispersion.

As searching on product features is made easier and product diversity is commonplace, the theory suggests that consumers will search more intensively on characteristics other than price while becoming less sensitive to price (Bakos, 1997; Anderson and Renault, 1999), leading to an increase in price dispersion. Our results support this prediction.

Consistent with our expectations, we find that consumers that click in lower screens have low price sensitivity and high taste for diversity. Multiple click and sorting customers exhibit similar behavior. First screen consumers are, as expected, the most sensitive to price and assign a lower value to brand. It is important to note that all of these customers — in spite of being shopbot customers and thus likely to be more sensitive to price than the overall consumer base — appear to value branding, as well as other product attributes, and therefore product diversity.

6 Concluding remarks

Shopbots are free Internet-based services that offer a comparison-shopping search tool that presents prices and other product characteristics from various competing retailers. To researchers interested in learning about consumer behavior patterns, they provide a unique opportunity to observe consumers revealing their preferences by making actual choices. This opens up a rich new set of possibilities for testing not only theories about search, but about the decision-making of consumers, retailers, and intermediaries more generally.

In this paper, we quantify consumer benefits to search and place an upper bound on consumer search costs. We find that search costs are significant in our setting, amounting to sixty percent of the price dispersion in the market. We also analyze the consumer heterogeneity present in the data and its implications in terms of consumer behavior and search benefits and costs. Across the various consumer types we find that first screen consumers are the most price sensitive. Consumers that scroll down multiple screens, on the other hand, have low price sensitivity but brand appears to play a relatively important role for them as presumably they choose to inspect lower screen because they care relatively more about other attributes besides price. Similarly, multiple click and sorting consumers appear to assign a high value on brand and are less price sensitive. Thus, in our setting increased search intensity is not correlated with greater price sensitivity, contrary to common assumption in most search cost theory.

Collectively, the results highlight two important factors regarding Internet commerce. First, the presence of search costs in this setting of nearly-perfect price and product information provides one possible explanation for the continuing presence of high levels of price dispersion in Internet markets. Second, the importance of non-price factors, even for homogeneous physical products, highlights the importance of retailer differentiation on the Internet through service characteristics and reputation. This is consistent with recent observations that in a modern economy, ancillary services take on an increased importance relative to the physical product. This also suggests that the future development of analytic models in the context of the Internet should take into account both consumer search costs and retailer differentiation. Analytic models that ignore retailer differentiation may not be that relevant for homogeneous physical products such as books — not to mention more complex and differentiated products such as electronics, cars, and computers.

References

[1] Anderson, S.P. and R. Renault (1999). "Pricing, product diversity, and search costs: a Bertrand-Chamberlin-Diamond model," *RAND Journal of Economics*, 30:719-735.

[2] Bajari, P. and A. Hortaçsu (2003). "The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions." *RAND Journal of Economics, 34:329-355.*

[3] Bakos, Y. (1997). "Reducing Buyer Search Cost: Implications for Electronic Markets." *Management Science*, 42:1613-1630.

[4] Baye, M.R., J. Morgan and P. Scholten (2002) "Persistent Price Dispersion in Online Markets", Indiana University, mimeo, April.

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

[6] Brynjolfsson, E., Hu, and M.D. Smith (2003). " Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Bookseller." *Management Science*, forthcoming.

[7] Brynjolfsson, E. and M.D. Smith (2000a). "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science*, 46:563-585.

[8] Brynjolfsson, E. and M.D. Smith (2000b). "The Great Equalizer? Consumer Behavior at Internet Shopbots." MIT, mimeo.

[9] Brown, J.R. and A. Goolsbee (2000). "Does the Internet Make Markets More Competitive? Evidence From the Life Insurance Industy." *Journal of Political Economy,* 110:481-507.

[10] Burdett, K. and K.L. Judd (1983). "Equilibrium Price Dispersion." *Econometrica*, 51:955-970.

[11] Butters (1977). "Equilibrium Distributions of Sales and Advertising Prices." *Review of Economic Studies*, 44:465-491.

[12] Chen, P. and L. Hitt (2003). "Understanding Price Dispersion in Internet-Enabled Markets." Working Paper, Carnegie Mellon University, Pittsburgh, PA.

[13] Chen, Y. and K. Sudhir (2002). "The Interacting Role of Consumer Search and Targeted Pricing: Implications for Price Competition on the Internet." Working paper.

[14] Chevalier, J. and A. Goolsbee (2003). "Measuring Prices and Price Competition Online: Amazon and Barnes and Noble." *Quantitative Marketing and Economics,* 1:203-222.

[15] Clay, K., R. Krishnan, E. Wolff, D. Fernandes (2002). "Retail Strategies on the Web: Price and Non-price Competition in the Online Book Industry." *Journal of Industrial Economics*, 50:351-367.

[16] Clemons, E.K., I. Hann and L.M. Hitt (2002). "Price Dispersion and Differentiation in On-Live Travel: An Empirical Investigation." *Management Science* 48:534-549.

[17] Diamond, P. (1971). "A Model of Price Adjustment." *Journal of Economic Theory*, 3:156-168.

[18] Diehl, K., L.J. Kornish, J.G. Lynch, Jr. (2003). "Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity."*Journal of Consumer Research* 30:56-71.

[19] Ellison, G. and S.F. Ellison (2001). "Search, Obsfuscation and Price Elasticities on the Internet." MIT, mimeo.

[20] Goldberg, P.K. (1995). "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry." *Econometrica*, 63:891-951.

[21] Hortaçsu, A. and C. Syverson (2004). "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: ACase Study of S&P 500 Index Funds." *Quarterly Journal of Economics,* 119:403-456.

[22] Johnson, E.J. and E. Russo (1984). "Product Familiarity and Learning New Information." *Journal of Consumer Research,* 11:542:550.

[23] Johnson, E.J., S. Bellman and G.L. Lohse (2002). "Cognitive Lock In and the Power Law of Practice." Columbia University, mimeo.

[24] Johnson, E. J., W. W. Moe, P. S. Fader, S. Bellman, J. Lohse (2004). "On the Depth and Dynamics of World Wide Web Shopping Behavior." *Management Science,* 50:299-308.

[25] Kuksov, D. (2003). "Buyer Search Costs and Endogenous Product Design." University of California, Berkeley, mimeo.

[26] Lynch, J.G., Jr. and D. Ariely (2000). "Wine online: Search cost and competition on price, quality, and distribution." *Marketing Science*, 19:83-103.

[27] McFadden, D.L.(1981). "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data*, ed. by C. Manski and D. McFadden. Cambridge: MIT Press.

[28] McFadden, D.L. (1984). "Econometric Analysis of Qualitative Response Models." *Handbook of Econometrics*, Volume II, ed. by Z. Griliches and M.D. Intriligator, Amsterdam: North Holland.

[29] Montgomery, A.L., K. Hosanager, R. Krishnan, K. Clay (2002). "Designing a Better Shopbot." CMU, mimeo.

[30] Morwitz, V., E.A. Greenleaf and E.J. Johnson (1998). "Divide and Prosper: Consumers' Reactions to Partitioned Prices." *Journal of Marketing Research,* 35:453-463.

[31] Nevo, A. (1997). "Demand for Ready-to-Eat Cereal and Its Implications for Price Competition, Merger Analysis, and Valuation of New Goods." Ph.D. Dissertation, Harvard University.

[32] Nevo, A. (2001). "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69:307-342.

[33] Rob, R. (1985). "Equilibrium Price Distributions." *Review of Economic Studies*, 52:487-504.

[34] Rossi, P.E., R.E. McCulloch and G.M. Allenby (1996). "The Value of Purchase History Data in Target Marketing." *Marketing Science*, 15:321:340.

[35] Small, K.A. and H.S. Rosen (1981). "Applied Welfare Economics with Discrete Choice Models." *Econometrica*, 49:105:130.

[36] Smith, M.D., J. Bailey, E. Brynjolfsson (2000). "Understanding Digital Markets." In *Understanding the Digital Economy* ed. by E. Brynjolfsson and B. Kahin. MIT Press, Cambridge, MA, 99-136.

[37] Smith, M.D. and E. Brynjolfsson (2001). "Consumer Decision-Making at an Internet Shopbot: Brand Still Matters." *Journal of Industrial Economics*, 49:541-557.

[38] Smith, M.D. 2002. "The Impact of Shopbots on Electronic Markets." *Journal of the Academy of Marketing Science* 30:442-450.

[39] Sorensen, A.T. (2000). "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs." *Journal of Political Economy*, 108:833-850.

[40] Sorensen, A.T. (2001). "Price Dispersion and Heterogeneous Consumer Search for Retail Prescription Drugs." University of California, San Diego, mimeo.

[41] Stahl, D.O. (1989). "Oligopolistic Pricing with Sequential Consumer Search." *American Economic Review*, 79:700-712.

[42] Stigler, G.J. (1961). "The Economics of Information." *Journal of Political Economy,* 72:44-61.

[43] Shugan, S. (1980) "The Cost of Thinking," *Journal of Consumer Research,* 7:99-111.

[44] Train, K. (1999). "Halton Sequences for Mixed Logit." University of California, Berkeley, working paper.

[45] Trajtenberg, M. 1989. "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners." *Journal of Political Economy*, 97:444-479.

[46] Varian, H. (1980)."A Model of Sales." *American Economic Review*, 70:651-659.

[47] Wernerfelt, Birger (1988). "Umbrella branding as a signal of new product quality: an example of signalling by posting a bond." *RAND Journal of Economics*, 19:458-466.

[48] Wilkie, W.L. and P.R. Dickson (1985). "Shopping for Appliances — Consumers' Strategies and Patterns of Information Search," Marketing Science Institute, working paper.

Table 1: TOP TEN BESTSELLERS IN DATASET

Book Title	Author(s)	Number of last clicks
Harry Potter and the Goblet of Fire	J.K.Rowling	1303
Harry Potter and the Chamber of Secrets	J.K.Rowling	408
Harry Potter and the Prisoner of Azkaban	J.K.Rowling	349
Java How to Program	P.J.Deitel and H.M.Deitel	318
C++ How to Program	H.M.Deitel and P.J.Deitel	233
The Carbohydrate Addict's Lifespan Program	R.F.Heller	214
A Tale of Two Cities	C.Dickens	203
Computer Networks	A.Tanenbaum	191
Harry Potter and the Sorcerer's Stone	J.K.Rowling	186
Who Moved My Cheese?	S.Johnson and K.H.Blanchard	180

Table 2: RETAILERS AND THEIR CLICK SHARES

Retailer	Share rank	No. last clicks	Click share
Borders	1	1173	0.1103792
1Bookstreet	2	1090	0.1025689
ecampus.com	3	923	0.0868542
Amazon	4	876	0.0824315
the BigStore.com	5	707	0.0665287
Fat Brain	6	603	0.0567423
Amazon.co.uk	7	495	0.0465795
Alphabetstreet	8	464	0.0436624
BN.com	9	428	0.0402748
Half.com	10	408	0.0383928
AlphaCraze	11	348	0.0327468
elgrande.com	12	319	0.0300179
A1Books	13	292	0.0274772
Countrybookstore	14	280	0.0263480
Shopping.com	15	251	0.0236191
Classbook.com	16	240	0.0225840
Indigo.ca	17	237	0.0223017
buy.com	18	202	0.0190082
uk.bol.com	19	187	0.0175967
ChaptersGLOBE.com	20	169	0.0159029
Davista	21	79	0.0074339
bol.de	22	75	0.0070575
Bookbuyer's Outlet	22	75	0.0070575
Internet Book Shop	24	74	0.0069634
Wordsworth	24	74	0.0069634
Hamilton Books	26	69	0.0064929
AllBooks4Less.com	27	67	0.0063047
Blackwells	27	67	0.0063047
Angus and Robertson	29	52	0.0048932
Dymocks	30	43	0.0040463
seekbooks	31	40	0.0037640
Page1Book	32	36	0.0033876
StudentBookWorld.com	33	33	0.0031053
lion.cc	34	30	0.0028230
Powells	35	23	0.0021643
Textbook.com	36	20	0.0018820
BCYbookloft.com	37	19	0.0017879
Amazon.de	38	15	0.0014115
Brians	39	11	0.0010351
1000's of Discount Books	40	10	0.0009410
WHSmith Online	41	9	0.0008469
BookCloseOuts	42	5	0.0004705
Lesezone	42	5	0.0004705
Magusbooks	44	3	0.0002823
WATERSTONES Online	45	1	0.0000941
ChristianBooks.com	46	0	0
Cherryvalley	46	0	0
Books For Cooks	46	0	0

Table 3: SUMMARY STATISTICS

Variable	Mean	St. Dev.	Min	Max
Click (1=yes)	0.0300	0.1705	0	1
Last click (1=yes)	0.0231	0.1501	0	1
Total price	52.11	32.78	1.25	212.91
Item price	42.14	31.42	0.50	180.40
Shipping price	9.70	6.66	0	59.92
Tax	0.26	1.04	0	13.08
Minimum delivery time	6.23	8.74	0	63
Maximum delivery time	9.47	12.69	0	85
Average delivery time	7.85	10.53	0	73.5
Delivery time not available†	0.4298	0.4951	0	1
Big three retailers (1=yes)	0.1813	0.3853	0	1
Screen number	2.85	1.45	1	7
Number of observations (offers)	460814			
Number of sessions	10627			

Source: Information constructed on the basis of Dealtime.com data. †the variable equals 1 if the retailer *did not* provide the delivery time.

Table 4: SUMMARY STATISTICS: SEARCH

	Percentage of sessions
First screen consumers (clicked only within default screen)	90.76%
Low screen consumers (scrolled down)	9.24%
Multiple click consumers (clicks>1)	16.48%
Sorting customers	0.79%
Last-clicked in offer number one	49.68%
Last-clicked one of the first three offers	75.49%
Last-clicked offer in second screen	4.23%
Last-clicked offer in third screen	1.55%
Scrolled to lower screens but chose first screen offer	1.16%
Number of sessions	10627

Table 5: RESULTS: S&B vs. CURRENT RESULTS

Explanatory Variable	Dependent Variable: 0/1 (i)	(ii)
Item price	−0.193	−0.190
	(0.001)**	(0.002)**
Shipping price	−0.367	−0.386
	(0.002)**	(0.004)**
Tax§	−0.361	−0.265
	(0.012)**	(0.024)**
Average delivery time	−0.018	−0.038
	(0.001)**	(0.002)**
Delivery time not available	−0.361	−0.235
	(0.015)**	(0.025)**
Big three retailers	0.332	0.356
	(0.014)**	(0.027)**
Sessions	39,635	10,627
Pseudo R-squared	0.285	0.365

NOTE.— Standard errors are in parentheses. ** significant at 1%. See text and Appendix for description of variables. Columns (i) and (ii) present results for S& B and our current results, respectively. §S&B use weighted tax, which tries to take into account locality taxes, unobserved to the researcher, in addition to state sales tax. S&B sample covers all book searches during Aug.25-Nov.1, 1999. Our sample is restricted to the top 100 bestseller book searches, and covers the period Sep. 1999-Sept. 2000.

Table 6: RESULTS: LOGIT MODEL

Explanatory Variable	Dependent Variable: 0/1				
	Entire sample (i)	First screen (ii)	Low screen (iii)	Sorting (iv)	Multiple Clicks (v)
Total price	−0.239	−0.353	−0.016	−0.126	−0.103
	(0.002)∗∗	(0.003)∗∗	(0.002)∗∗	(0.017)∗∗	(0.003)∗∗
Avg. delivery time	−0.026	−0.032	−0.020	−0.040	−0.014
	(0.002)∗∗	(0.002)∗∗	(0.005)∗∗	(0.022)†	(0.003)∗∗
Delivery time N/A	−0.096	−0.290	−0.042	0.074	−0.138
	(0.025)∗∗	(0.028)∗∗	(0.073)	(0.265)	(0.056)∗
Big 3 retailers	0.258	0.174	0.867	0.580	0.759
	(0.027)∗∗	(0.031)∗∗	(0.069)∗∗	(0.278)∗	(0.060)∗∗
Observations	460814	416373	44441	3209	75620
Sessions	10627	9645	982	84	1751
Pseudo R-squared	0.336	0.452	0.031	0.204	0.178

NOTE.— Standard errors are in parentheses. †significant at 10%; *significant at 5%; ** significant at 1%. See text and Appendix for description of variables. The models are estimated on the entire set of offers in a session.

Table 7: RESULTS: LOGIT MODEL WITH DISTANCE MEASURE

Explanatory Variable	Dependent Variable: 0/1				
	Entire sample (i)	First screen‡ (ii)	Low screen (iii)	Sorting (iv)	Multiple Clicks (v)
Total price	−0.165	−0.237	−0.023	−0.120	−0.058
	(0.003)∗∗	(0.004)∗∗	(0.003)∗∗	(0.018)∗∗	(0.003)∗∗
Avg. delivery time	−0.028	−0.034	−0.019	−0.031	−0.017
	(0.002)∗∗	(0.002)∗∗	(0.005)∗∗	(0.020)	(0.003)∗∗
Delivery time N/A	−0.072	−0.223	−0.033	−0.129	−0.153
	(0.024)∗∗	(0.027)∗∗	(0.073)∗∗	(0.287)	(0.056)∗∗
Big 3 retailers	0.172	0.058	0.901	0.191	0.689
	(0.027)∗∗	(0.030)∗∗	(0.070)∗∗	(0.309)	(0.061)∗∗
Screen number	−0.999		0.110	−1.298	−0.813
	(0.026)∗∗		(0.036)∗∗	(0.205)∗∗	(0.042)∗∗
Observations	460814	95200	44441	3209	75620
Sessions	10627	9645	982	84	1751
Pseudo R-squared	0.360	0.205	0.032	0.311	0.214

NOTE.— Standard errors are in parentheses. †significant at 10%; *significant at 5%; ** significant at 1%. See text and Appendix for description of variables. The models are estimated on the entire set of offers in a session, except where noted. ‡The variable screen number is not shown in the case of first screen consumers because, by definition, no first screen consumer clicks on a lower screen, and as a result the model with a distance measure estimated on the entire set of offers is exactly identical to the model shown here which is based on the top ten offers.

Table 8: RESULTS: RANDOM COEFFICIENTS MODEL (ENTIRE SAMPLE)

Explanatory Variable	MODEL I Mean	MODEL I Standard deviation	MODEL II Mean	MODEL II Standard deviation
Total price	−0.6461	0.3903	−0.5941	0.4255
	(0.0108)**	(0.0085)**	(0.0119)**	(0.0095)**
Average delivery time	−0.0591	0.0555	−0.0549	0.0488
	(0.0033)**	(0.0041)**	(0.0033)**	(0.0045)**
Delivery time not available	−0.6119	1.3280	−0.5664	0.8045
	(0.0389)**	(0.1239)**	(0.0358)**	(0.1477)**
Big three retailers	0.3001	1.2178	0.3013	1.2625
	(0.0477)**	(0.1392)**	(0.0462)**	(0.1362)**
Screen number			−2.3589	2.1596
			(0.1200)**	(0.1016)**
Sessions	10627		10627	

NOTE.— Robust standard errors are in parentheses. ** significant at 1%. See text and Appendix for description of variables.

Table 9: RESULTS: RANDOM COEFFICIENTS MODEL FOR FIRST SCREEN CONSUMERS‡

Explanatory Variable	MODEL I Mean	MODEL I Standard deviation	MODEL II Mean	MODEL II Standard deviation
Total price	−0.8178	0.4313	−0.7830	0.5404
	(0.0152)**	(0.0103)**	(0.0190)**	(0.0156)**
Average delivery time	−0.0658	0.0687	−0.0634	0.0647
	(0.0038)**	(0.0047)**	(0.0040)**	(0.0060)**
Delivery time not available	−0.7808	2.0785	−0.7232	1.4343
	(0.0506)**	(0.1296)**	(0.0462)**	(0.1329)**
Big three retailers	0.2572	1.1896	0.2581	1.0748
	(0.0521)**	(0.1591)**	(0.0490)**	(0.1639)**
Sessions	9645		9645	

NOTE.— Robust standard errors are in parentheses. ** significant at 1%. See text and Appendix for description of variables. ‡The first model is estimated on the entire set of offers, while the second model is estimated on the basis of the first ten offers only.

Table 10: RESULTS: RANDOM COEFFICIENTS MODEL FOR LOW SCREEN CONSUMERS

Explanatory Variable	MODEL I Mean	MODEL I Standard deviation	MODEL II Mean	MODEL II Standard deviation
Total price	−0.0164	0.0010	−0.0235	0.0003
	(0.0025)**	(0.0075)	(0.0035)**	(0.0075)
Average delivery time	−0.0279	0.0236	−0.0250	0.0202
	(0.0082)**	(0.0106)*	(0.0081)**	(0.0117)†
Delivery time not available	−0.0743	0.0064	−0.0588	0.0039
	(0.0773)	(0.2359)	(0.0773)	(0.2343)
Big three retailers	0.7104	1.1454	0.7427	1.1717
	(0.1823)**	(0.7277)	(0.1856)**	(0.7339)
Screen number			0.1095	0.0168
			(0.0374)**	(0.1738)
Sessions	982		982	

NOTE.— Robust standard errors are in parentheses. *significant at 5%; ** significant at 1%. See text and Appendix for description of variables.

Table 11: RESULTS: RANDOM COEFFICIENTS MODEL FOR CONSUMERS THAT SORT

Explanatory Variable	MODEL I Mean	MODEL I Standard deviation	MODEL II Mean	MODEL II Standard deviation
Total price	−0.3458	0.2403	−0.3641	0.2043
	(0.0690)**	(0.0701)**	(0.0842)**	(0.0592)**
Average delivery time	−0.1526	0.1045	−0.1464	0.1224
	(0.0530)**	(0.0325)**	(0.0636)*	(0.0548)*
Delivery time not available	−0.3619	1.9671	0.5253	7.9340
	(0.4835)	(1.4885)	(1.4539)	(3.6632)*
Big three retailers	0.3485	0.4921	0.0374	1.3162
	(0.3606)	(1.0509)	(0.5502)	(1.0974)
Screen number			−4.4250	3.1047
			(1.3537)**	(0.9115)**
Sessions	84		84	

NOTE.— Robust standard errors are in parentheses. ** significant at 1%. See text and Appendix for description of variables.

Table 12: RESULTS: RANDOM COEFFICIENTS MODEL FOR CONSUMERS WITH MULTIPLE CLICKS

Explanatory Variable	MODEL I Mean	MODEL I Standard deviation	MODEL II Mean	MODEL II Standard deviation
Total price	−0.1938	0.1288	−0.1050	0.0981
	(0.0084)**	(0.0079)**	(0.0077)**	(0.0084)**
Average delivery time	−0.0279	0.0224	−0.0270	0.0179
	(0.0057)**	(0.0099)*	(0.0053)**	(0.0105)†
Delivery time not available	−0.4204	1.6959	−0.3370	0.5870
	(0.0853)**	0.3651)**	(0.0655)**	(0.3417)†
Big three retailers	0.8718	0.0569	0.8134	0.3727
	(0.0675)**	(0.3924)	(0.0732)**	(0.4399))
Screen number			−1.6327	1.3868
			(0.1248)**	(0.1203)**
Sessions	1751		1751	

NOTE.— Robust standard errors are in parentheses. ** significant at 1%. See text and Appendix for description of variables.

Table 13: LOGIT PRICE ELASTICITY PERCENTILES

Price	10%	25%	Median	75%	90%
Entire sample	−23.89	−18.15	−9.77	−5.56	−4.11
	−16.50	−12.53	−6.75	−3.87	−2.83
First screen	−35.14	−26.80	−14.46	−8.20	−5.96
	−17.01	−12.50	−6.00	−3.70	−2.74
Low screen	−1.62	−1.18	−0.68	−0.38	−0.30
	−2.35	−1.71	−0.98	−0.55	−0.43
Sorting	−13.08	−9.75	−4.36	−2.76	−2.14
	−12.49	−9.27	−4.18	−2.63	−2.02
Multiple clicks	−10.74	−8.43	−5.35	−2.66	−1.92
	−6.11	−4.78	−3.03	−1.53	−1.09

NOTE.— Based on estimates from Table 6 and Table 7, respectively. The elasticity distribution is over offers in the corresponding consumer sample. Figure indicates percentage change in the choice probability for a given retailer given a 1 percent increase in the retailer's price.

Table 14: RANDOM COEFFICIENTS PRICE ELASTICITY PERCENTILES

	10%	25%	Median	75%	90%
Entire sample	−12.68	−7.91	−4.19	0.29	6.06
	−10.58	−6.33	−2.95	−0.27	3.78
First screen	−15.11	−9.42	−5.01	−0.12	5.86
	−20.18	−10.66	−5.41	−1.07	4.91
Low screen	−1.66	−1.23	−0.70	−0.39	−0.31
	−2.40	−1.75	−1.00	−0.56	−0.44
Sorting	−10.06	−6.62	−3.93	−0.65	4.46
	−16.28	−8.26	−4.50	−1.67	0.66
Multiple click	−9.60	−6.34	−3.77	−1.79	0.94
	−4.78	−2.27	−0.61	1.10	3.21

NOTE.— Based on estimates from model II and II, respectively. The elasticity distribution is over offers in the corresponding consumer sample.

Table 15: BENEFITS TO SEARCH ($ units)

	10%	25%	Median	75%	90%
Entire sample	2.27	3.58	5.74	9.54	18.76
	2.82	4.20	6.55	11.15	20.51
First screen	1.37	2.59	4.72	8.41	16.88
	0.42	0.76	1.17	2.02	4.13
Low screen	74.42	83.72	93.17	103.18	112.76
	53.54	58.62	64.07	70.97	76.29
Sorting	4.96	6.64	9.14	13.47	22.63
	3.46	6.22	9.68	13.94	21.37
Multiple click	8.73	11.82	16.08	23.05	36.30
	16.74	22.14	29.75	41.50	65.04

NOTE.— Based on estimates from model I and model II, respectively. The distribution is over customers in the corresponding consumer sample. Figures represent U.S. dollars.

Table 16: UPPER BOUND TO SEARCH COSTS ($ units)

	10%	25%	Median	75%	90%
Entire sample (n = 982)	2.52	3.78	5.47	9.01	17.70
	3.15	4.38	6.45	10.88	19.89
Low screen (n = 982)	74.42	83.72	93.17	103.18	112.76
	53.54	58.62	64.07	70.97	76.29
Sorting (n = 14)	5.03	7.19	9.16	14.90	19.99
	4.97	7.10	10.30	18.04	30.63
Multiple click (n = 471)	9.71	12.35	16.18	22.64	36.97
	17.47	22.77	30.52	44.02	71.41

NOTE.— Based on estimates from model I and model II, respectively. The distribution is over customers in the corresponding consumer sample. Note that for low screen consumers search benefits are identical to search costs, since by definition these consumers scrolled down. Also, in the case of first screen consumers, their search benefits represent a lower bound to search costs (shown on previous table). Figures represent U.S. dollars.

APPENDIX: DESCRIPTION OF VARIABLES

Variable	Description
Click	=1 if the offer was one on which the customer clicked (which may not be the last click).
Last click	=1 if the offer was the last one on which the customer clicked on.
Total price	Total price as listed in the shopbot's screen. Total price = item price + shipping cost + sales tax.
Item price	Item price as listed in the shopbot's screen.
Shipping price	Shipping price as listed in the shopbot's screen.
Tax	Sales tax as listed in the shopbot's screen
Minimum delivery time	The smallest number in the range specified by the retailer for delivery time, whenever a range as opposed to a single number of days is provided.
Maximum delivery time	The largest number in the range specified by the retailer for delivery time, whenever a range as opposed to a single number of days is provided.
Average delivery time	Delivery time = Acquisition time + Shipping time. "Average" delivery time is the average between maximum delivery time and minimum delivery time offered by the retailer, whenever a time range is provided by the retailer. Otherwise it is just the specific time indicated.
Delivery not available	=1 if the retailer *did not* provide a delivery time.
First screen consumer	=1 if the consumer only clicked on offers in first screens.
Low screen consumer	=1 if the consumer clicked on offers in lower screens.
Sorting consumer	=1 if consumer sorted by column other than total price, which is how the screen is ordered when first shown to the consumer.
Multiple click consumer	=1 if consumer clicked on multiple offers.
Big three retailers	=1 if the retailer is one of the well-known retailers throughout the sample. Namely, Amazon.com, Barnes & Noble, Borders.
Screen number	=1 if the offer is listed within the default screen; =2 if the offers is listed on the second screen; and so on.