DRAFT
COMMENTS WELCOME

Consumer Decision-making at an Internet Shopbot

Michael D. Smith* and Erik Brynjolfsson**

Available at http://ebusiness.mit.edu/papers/cdis

Acknowledgements: We thank Austan Goolsbee, Il-Horn Hann, John D.C. Little, Thomas Malone, Allen Montgomery, Nancy Rose, Robert Zeithammer, and seminar participants at Boston University, Carnegie-Mellon University, The University of Chicago, Indiana University, the University of Maryland, MIT, the University of Michigan, the University of Texas at Austin, Stanford University, The Wharton School, the Brookings conference on "Measuring E-Commerce" (Brookings Institution, September 24, 1999), the Workshop on Information Systems and Economics (December 11, 1999), and the NBER E-Commerce Project for valuable comments on this research. We thank Christoph Janz and Christopher Muenchhoff of DealTime.com for generously collecting the data necessary to conduct this research and for providing valuable insight into the shopbot market. Thomas Cheng and Frederic Mahoue provided excellent research assistance. The Center for eBusiness@MIT provided generous financial support under a grant from Fleet Bank.

* H. John Heinz III School of Public Policy and Management, Carnegie Mellon University; email: mds@cmu.edu; web: http://www.heinz.cmu.edu/~mds

** Sloan School of Management, Massachusetts Institute of Technology; email: erikb@mit.edu; web: http://ebusiness.mit.edu/erik

Consumer Decision-making at an Internet Shopbot

Abstract

Internet shopbots allow consumers to almost effortlessly compare prices and service levels of dozens of competing retailers. This creates a valuable laboratory for assessing the determinants of consumer choice and accords well with the assumptions underlying commonly-used logit models. We analyze the choices of 20,227 shopbot consumers who choose among 33 competing retailer offers for books over a sample period of 69 days. We find that consumers are remarkably sensitive to how the total price for good is allocated among the item price, the shipping cost, and tax, and are also quite sensitive to the ordinal ranking of retailer offerings with respect to price. We also find that even in this setting, brand is important and that in particular, consumers appear to use brand as a proxy for a retailer's credibility with regard to non-contractible aspects of the product bundle such as shipping time. The shopbot "laboratory" appears to provide a useful new tool for testing a variety of economic theories.

1. Introduction

Shopbots are Internet-based services that provide one-click access to price and product information from numerous competing retailers. In so doing, they reduce buyer search costs for product and price information by at least 30-fold compared to telephone-based shopping and even more compared to physically visiting the retailers (Brynjolfsson and Smith, 2000a). Shopbots list summary information for both well- and lesser-known retailers, and typically rank the retailers based on a characteristic of interest to the shopper such as price or shipping time. These comparison tables reveal a great deal of variation across retailers in relative price levels, delivery times, and product availability. Further, shopbots provide researchers with an opportunity to observe choice behavior as consumers evaluate these comparison tables and click on a particular product offers. Consumer choice behavior can then be analyzed using econometric models to reveal how consumers respond to different aspects of the product bundle.

The use of econometric models may be particularly applicable to this setting since available choices provided by a typical Internet shopbot are conducive to rational, objective decision making by shoppers. Indeed, it closely approximates the idealized setting implicit in logit models of consumer choice: consumers have most of the relevant information about each alternative and can directly compare them side-by-side. The computer automatically calculates the net effect of various components of not only the pricing, (e.g. shipping cost, tax, item price) relevant to their particular circumstances but also the net effects of non-price characteristics like shipping and delivery time. More sophisticated shopbots will even attempt to calculate and rank the utility-maximizing options based on the consumers own weighting function, provided explicitly or implicitly by the consumer. In even the simplest shopbots, consumers can re-rank the choices based on characteristics of interest to them with minimal effort. For the shopbot data we study, advertising budgets and word of mouth do not affect the way the information is presented on the screen. Similarly, switching costs are minimized. Information about the offerings of new retailers is presented in the same way as information about retailers that have been visited many times before.

Even the best shopbots do not provide full information about the alternatives, however. For instance, not all potential retailers are included in shopbot databases. Furthermore, the data may not always be accurate: retailers can promise to deliver by a certain date, but will they deliver on their promise? Worse yet, some retailers may be outstanding at posting low prices and at efficiently processing credit cards, but may be sloppy or worse when it comes to actually providing the goods and services purchased.

Shopbots provide a valuable tool for understanding the decision-making processes of consumers. As with supermarket scanner data, the data collected by shopbots can provide insight into the trade-offs consumers make when confronted with a choice of alternatives. How important is price and does it matter how the total cost of a purchased is apportioned among various components? What is the role of branding? Is it more important as a signal for some types of consumers and for some types of decisions than for others?

We address these questions through panel data gathered from an Internet shopbots in the market for books. We use these data to study two major aspects of Internet shopbot markets. First, we analyze consumer response to partitioned pricing strategies (separating total price into item price, shipping cost, and sales tax), and the ordinal rankings of prices. Second, we use the responses of observable groups of consumers to analyze how consumers respond differently to contractible aspects of the product bundle versus non-contractible aspects such as promised delivery times.

Our approach to analyzing electronic markets complements recent empirical studies that examine the pricing behavior from the perspective of efficiency (Bailey 1998; Brynjolfsson and Smith 2000a), retailer differentiation (Clay, Krishnan, Wolff, Fernandes 1999), and price discrimination (Clemons, Hann, and Hitt 1998). While these studies are able to analyze competitive strategies across retailers and markets, they provide only second-order evidence of consumer behavior in electronic markets. In contrast, the current paper and a companion paper (Brynjolfsson and Smith, 2000b), directly analyze customer behavior by using the shopbot as a laboratory of sorts where consumers respond to heterogeneous offers from variety of retailers.

Our first set of results contrasts with what might be expected from a straight-forward application of utility theory and rational consumer behavior: we find that shopbot consumers do not treat all components of prices equivalently. They are significantly more sensitive to changes in shipping cost than they are to changes in item price, even when the total price they must pay is unaffected. They are also quite sensitive to the ordinal rankings of prices. When it comes to affecting purchasing behavior, some pennies are more equal than others. Our second set of results suggests that branding is more important for consumers who care about non-contractible aspects of the product bundle. In particular, consumers who care about shipping times are especially likely to prefer well-known brands, potentially because promised shipping times are difficult to enforce.

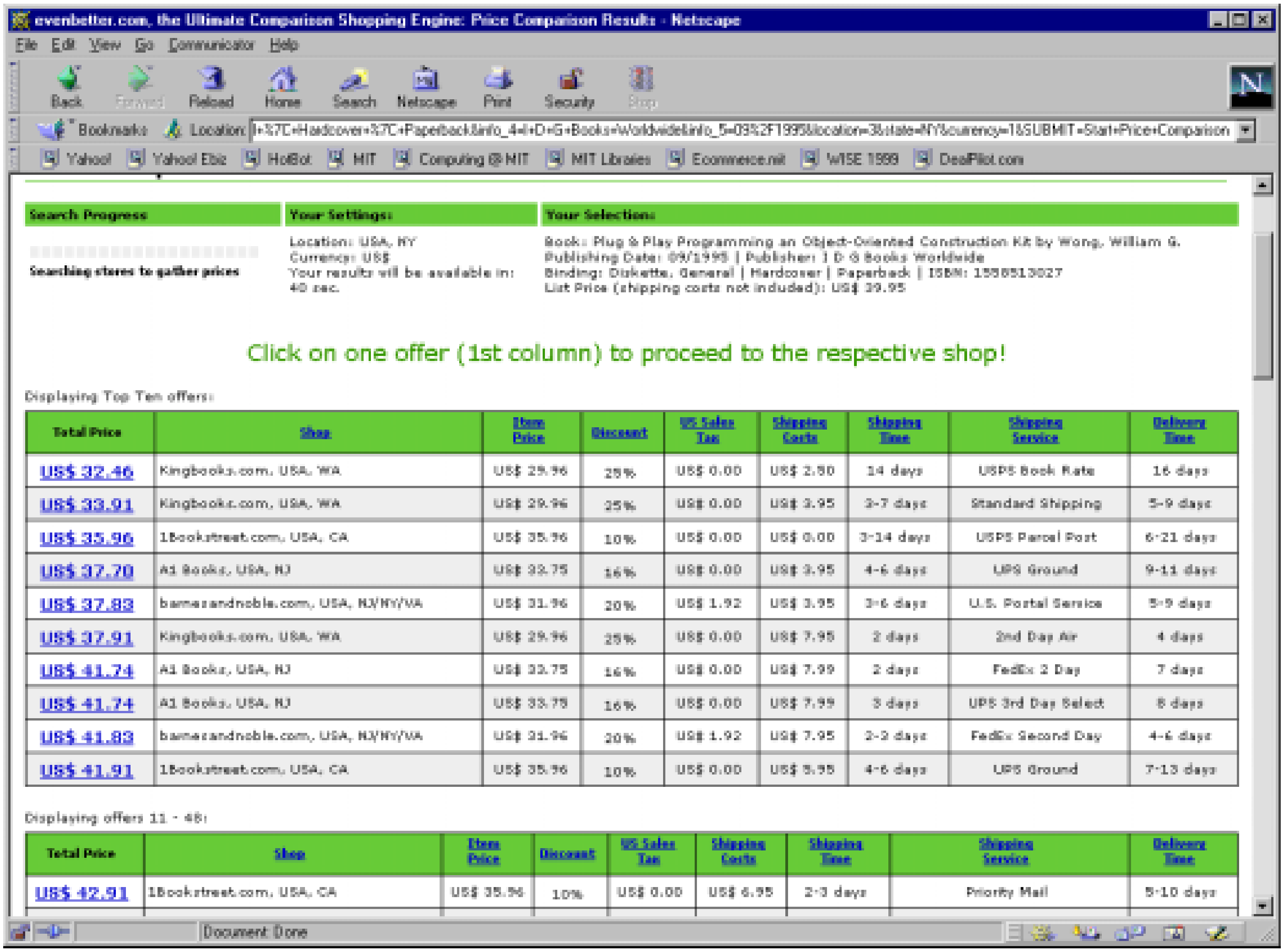
The remainder of this paper is organized in three parts. Section 2 summarizes the data we collect and the empirical models we use to analyze our data. Section 3 presents our main results and we summarize our conclusions in Section 4.

2. Data and Methods

2.1. Data Source and Characteristics

Our analysis uses panel data collected from book consumers at DealTime, a prominent Internet shopbot. Consumers using this service first identify the book they are interested in purchasing, narrowing their selection to a unique, and physically homogeneous, product selection. Once the book is chosen, the consumer provides their country and state so that local currency and local taxes can be calculated correctly. After the consumer initiates their search, DealTime queries prices for this selection in real time from 33 different book retailers, which collectively account for the vast majority of books sold online. Because this information is queried in real time directly from the retailers, the information displayed by DealTime’s comparison tables are the same as those obtained by visiting retailers’ sites directly.

Figure 1: Sample Screen from DealTime.com



Displaying Top Ten offers:

Total Price	Shop	Item Price	Discount	US Sales Tax	Shipping Costs	Shipping Time	Shipping Service	Delivery Time
US$ 32.46	Kingbooks.com, USA, WA	US$ 29.96	25%	US$ 0.00	US$ 2.50	14 days	USPS Book Rate	16 days
US$ 33.91	Kingbooks.com, USA, WA	US$ 29.96	25%	US$ 0.00	US$ 3.95	3-7 days	Standard Shipping	5-9 days
US$ 35.96	1Bookstreet.com, USA, CA	US$ 35.96	10%	US$ 0.00	US$ 0.00	3-14 days	USPS Parcel Post	6-21 days
US$ 37.70	A1 Books, USA, NJ	US$ 33.75	16%	US$ 0.00	US$ 3.95	4-6 days	UPS Ground	9-11 days
US$ 37.83	barnesandnoble.com, USA, NJ/NY/VA	US$ 31.96	20%	US$ 1.92	US$ 3.95	3-6 days	U.S. Postal Service	5-9 days
US$ 37.91	Kingbooks.com, USA, WA	US$ 29.96	25%	US$ 0.00	US$ 7.95	2 days	2nd Day Air	4 days
US$ 41.74	A1 Books, USA, NJ	US$ 33.75	16%	US$ 0.00	US$ 7.99	2 days	FedEx 2 Day	7 days
US$ 41.74	A1 Books, USA, NJ	US$ 33.75	16%	US$ 0.00	US$ 7.99	3 days	UPS 3rd Day Select	8 days
US$ 41.83	barnesandnoble.com, USA, NJ/NY/VA	US$ 31.96	20%	US$ 1.92	US$ 7.95	2-3 days	FedEx Second Day	4-6 days
US$ 41.91	1Bookstreet.com, USA, CA	US$ 35.96	10%	US$ 0.00	US$ 5.95	4-6 days	UPS Ground	7-13 days

Displaying offers 11 - 48:

Total Price	Shop	Item Price	Discount	US Sales Tax	Shipping Costs	Shipping Time	Shipping Service	Delivery Time
US$ 42.91	1Bookstreet.com, USA, CA	US$ 35.96	10%	US$ 0.00	US$ 6.95	2-3 days	Priority Mail	5-10 days

Based on the information returned, DealTime provides consumers with a list of product attributes including separate fields for the total price, item price, shipping cost, sales tax, delivery time, shipping time, and shipping service (e.g., Figure 1). By default, the table is sorted in ascending order on total price but the consumer can sort the table based on any of the other eight columns if they desire. Consumers view these tables and make an observable choice by clicking on an offer. By clicking through on an offer, the consumer is taken directly to the retailers’ web site where they can finalize their purchase. Where consumers click on multiple offers in a search, we use the offer they click on last as an indication of their final choice.

In our data we gather the information shown to the consumer in the offer comparison table and the consumer's behavior (whether they sort on a column other than total price and their last click). The data we gather is shown in Table 1 and described in more detail in Brynjolfsson and Smith (2000b).

We obtained data for the period from August 25 to November 1, 1999. In this paper, we focus on the subset of the sample that 1) includes U.S.-based consumers (75.4% of sessions), 2) lead to at least one click-through by the consumers (26.3% of remaining sessions) and 3) return more than one retailer (99.9% of remaining sessions). The resulting data set reflects searches by 20,227 distinct consumers, including 7,478 repeat visitors. These consumers conducted at total of 39,654 searches and the shopbots returned a total of 1,513,439 distinct retailer offerings, an average of just over 38 offers per search.

Table 1: Shopbot Data Collected

Offer Data	
Item Price	The price for the item
Shipping Cost	The price for shipping
State Sales Tax	Sales tax (if applicable)
Weighted Sales Tax	State sales tax plus city/county taxes weighted by Internet population (1998)
Retailer	Retailer Name (used to create dummy variables for each retailer)
Delivery Time	Average of the minimum and maximum delivery range quoted by the retailer
Shipping Method	Priority (1-day or 2-day), Standard (3-7 day), Book Rate (>7 day)
Delivery NA	=1 if retailer can't quote the time to get the book from the distributor
Rank	The position of the offer in the comparison table
Session Data	
Date/Time	Date and time search occurred
ISBN	ISBN number of book searched for (used to calculate book type)
Sort Column	Identifies which column the consumer sorted on (default is total price)
Choice Data	
Last Click-Through	=1 if the consumer's last click through was on this offer

Two important factors are observed in this data. First, there is a high degree of price dispersion across homogeneous books in the offers: the difference between the lowest priced offer and the tenth lowest priced offer averages $10.77 (or 32.3% of the price of the product) in our data. While such a large difference in prices among homogeneous goods may seem surprising, this level of dispersion is comparable to Brynjolfsson and

Smith (2000a, p. 575) who analyzed the Internet book retailing market in a comparable setting between 1998-1999.

Second, less than half of the consumers in our sample choose the lowest priced offer. Among consumers who do not choose the lowest priced offer, the average selected offer is $6.79 (20.4%) higher than the lowest priced offer in the session. Consumers presumably perceive some differences among retailers or products that make at least some of them willing to pay a premium in at least some cases. For instance, different retailers have different policies on shipping and a few have invested significantly in creating branded identities. This opens the door for us to use econometric techniques to analyze the non-price factors that may influence consumer choice in this setting.

2.2. Methodology

As noted above, shopbot data are well suited to analysis via multinomial logit and nested logit models because the presentation of the data to the consumers appears to lend itself to rational, attribute-based decision processes. Our analysis these models, using the techniques described in Brynjolfsson and Smith (2000b). We briefly describe these models below, referring the interested reader to more detailed presentations of the models (e.g., McFadden (1974), Ben-Akiva and Lerman (1985)) and prior empirical settings using these models (e.g., Guadagni and Little (1983), Fader and Hardie (1996)).

Two points are important to note with regard to applying these model to our data. First, our analysis is, by necessity, restricted to consumers who actually choose to use a particular shopbot. Thus, our logit model predictions are conditioned on a consumer choosing to use this shopbot. While conditioning in this way does not bias the logit results, they should be interpreted as applying to a self-selected set of consumers who are likely to differ systematically from consumers in general and even from other Internet shoppers.

Second, because we observe click-throughs and not purchases in our data, our models reflect those factors that drive traffic to a site, not necessarily those that drive sales. However, in related research (Brynjolfsson and Smith 2000b) we find that the factors that drive traffic are also relatively unbiased predictors of sales, increasing the validity of inferences in this regard. Nonetheless, a conservative interpretation our approach is as a model of click-throughs, not of sales per se.

2.2.1. Multinomial Logit Model

The Internet shopbot we study presents a set of offers to each consumer. As discussed above, each offer is for the identical book but varies in price, taxes, delivery time, shipping method and brand of retailer. In the multinomial logit model,[1] consumer utility (U_{it}) for each offer (t) in each session (i) is the sum of a systematic component (V_{it}) and a stochastic component (ε_{it}):

$$U_{it} = V_{it} + \varepsilon_{it} \quad (1)$$

In the model the stochastic disturbance reflects both unobserved taste variation across consumers and measurement error in evaluating offers. The systematic component is the linear combination of the product's attributes ($\mathbf{x}'_{it}$) and the consumer's preferences for those attributes (β), yielding:

$$U_{it} = \mathbf{x}'_{it}\beta + \varepsilon_{it} \quad (2)$$

Given a utility maximizing consumer facing a choice between multiple offers, we can assume that the probability that offer t is chosen in session i as:

$$P_t(\mathbf{x}_{it}, \beta) = \Pr\{U_{it} = \arg\max(U_{i1}, U_{i2,} \ldots U_{iT_i})\} \quad (3)$$

[1] The multinomial logit model is also occasionally referred to as the conditional logit model (McFadden 1974). The model presented here refers to choice specific (as opposed to individual specific) attributes (Greene 1997, pp. 912-920). We adopt the multinomial logit terminology for consistency with the majority of the choice literature.

If one assumes that the errors are distributed according to the type I extreme value distribution, (2) and (3) simplify to the following form (McFadden 1974):

$$P_t(\mathbf{x}_i, \beta) = \frac{e^{\mu\beta'\mathbf{x}_{it}}}{\sum_{\tau=1}^{\tau_i} e^{\mu\beta'\mathbf{x}_{i\tau}}} \tag{4}$$

This equation can then be used to estimate the weights consumers place on product attributes (β) using standard maximum likelihood techniques.

2.2.2. Nested Logit Model

For some our analyses, we also estimate a Nested Logit model, with the choice between branded (Amazon, Barnes and Noble, or Borders) and unbranded retailers as the top nest. This accommodates the potential violation of the Independence of Irrelevant Alternatives assumption implicit in the standard logit model. These models provide a useful robustness check on the validity of the basic multinomial logit approach – does the IIA assumption significantly affect the results obtained using standard multinomial logit models for our data?

The nested logit model assumes that consumers face a two-step choice process with an initial choice set *S* (e.g., *S={branded retailers, unbranded retailers}*) and a restricted choice set *R* (e.g., *R={{amazon, barnesandnoble, borders},{albooks, kingbooks, 1bookstreet}}*). In this setting, if one assumes (1) $var(\varepsilon_r)=0$, (2) ε_s and ε_{sr} are independent for brand and retailer selections in the consumer's choice set, (3) ε_{sr} independent and identically Gumbel distributed with a scale parameter μ_r, and (4) ε_s distributed such that $\max_{r \in R_{ns}} U_{rs}$ is Gumbel distributed with a scale parameter of μ_s, the choice probabilities reduce to the following:

$$p(r \mid s) = \frac{e^{\mu_r (V_{sr} + V_r)}}{\sum_{j \in R_{ns}} e^{\mu_r (V_{sj} + V_j)}} \tag{5}$$

$$p(s) = \frac{e^{\mu_s (V_s + V_s')}}{\sum_{i \in S_n} e^{\mu_s (V_i + V_i')}} \tag{6}$$

where

$$V_s' = \frac{1}{\mu_r} \ln \sum_{r \in R_{ns}} e^{(V_r + V_{sr})} \tag{7}$$

Thus, at the lower level nest, choice probabilities are given by the standard multinomial logit formula (conditional on the upper level choices). At the upper level, choice probabilities are given by the standard multinomial logit formula including a term (V_s') which represents the expected value of the consumer's maximum utility from choosing *s* (Ben-Akiva and Lerman 1985).

3. Empirical Results

Our analysis addresses two empirical questions: consumer response to partitioned pricing strategies, and consumer response to contractible and non-contractible aspects of the product bundle. We address these questions in turn below using multinomial logit and nested logit models.

3.1. Consumer Response to Partitioned Pricing

We consider consumer response to the elements of total price: item price, shipping cost, and sales taxes. Prices that are comprised of a base cost and various surcharges are referred to as "partitioned prices" in the marketing literature. One might expect that, as long as the total price remains the same, a consumer should be indifferent among different ways of portioning the price to various components.

However, Morwitz, Greenleaf, and Johnson (1998) analyze partitioned prices in environments where it is difficult for the consumer to calculate the total price from the presentation of the base price and surcharge and found this was not always the case.[2] They find that the use of partitioned prices increases demand and decreases the consumer's recalled total costs relative to when partitioned prices are not used. They show that surcharges can therefore be an effective pricing strategy for retailers. They also find that these effects are stronger when more effort is required to associate the base price and the surcharge. These results may explain why Internet retailers commonly use partitioned prices for their web-site direct consumers. If the cost of surcharges such as shipping cost are not known until the final step of a purchase, the effect may be to decrease the Internet consumer's perception of total price during their initial evaluation of the product.

However, shopbots present consumers with a very different environment with regard to partitioned prices. Here consumers are no only made aware of the total price, but can easily evaluate the elements of total price — item price, shipping cost, and sales tax — when making their choice. Indeed, for our data the default sort order is based on the total price, including all these elements.

Table 2 presents results regarding customer response to partitioned prices in a shopbot environment. The first column of Table 3 includes only variables relating to the partitioned prices themselves. The tax variable is a combination of the state tax and any relevant city and locality taxes weighted by the number of Internet users in each locality at the end of 1998.[3] While DealTime consumers are shown the state taxes alone, they might also factor in the possibility of having to pay locality taxes.[4] Columns 2 and 3 add variables controlling for delivery time and branded retailer dummies respectively. In column 4 we add variables controlling for the position of the offer in the price comparison table. Column 5 adds separate variables for the 12 most popular retailers at DealTime.com. Finally, column 6 adds separate variables for price changes within the three standard shipping bands: express shipping (1-2 day), standard shipping (3-7 day),

---

[2] I.e., because they are computationally difficult to calculate (base cost plus a percentage) or involve search costs (shipping costs not quoted with base costs).
[3] We thank Austan Goolsbee for providing these data.
[4] This inclusion of these taxes represents a conservative assumption. Excluding locality taxes increases the sensitivity to taxes in our models.

and book rate (>7 day). This controls for the possibility that the increased sensitivity in shipping cost is due to sensitivity across shipping bands as opposed to within shipping bands.

Table 2: Models of Partitioned Price Sensitivity

	1	*2*	*3*	*4*	*5*	*6*
Item Price	-.183 (.001)	-.192 (.001)	-.194 (.001)	-.039 (.001)	-.043 (.001)	-.039 (.001)
Ship Price	-.343 (.002)	-.362 (.002)	-.368 (.002)	-.089 (.002)	-.104 (.002)	
Ship Pr. (1-2 days)						-.120 (.003)
Ship Pr. (3-7 days)						-.070 (.004)
Ship Pr. (>7 days)						-.041 (.003)
Weighted Tax	-.350 (.011)	-.374 (.011)	-.350 (.011)	-.067 (.010)	-.059 (.011)	-.053 (.011)
Avg. Del Time		-.018 (.001)	-.019 (.001)	-.035 (.001)	-.027 (.001)	-.029 (.001)
Delivery "N/A"		-.449 (.014)	-.364 (.015)	-.394 (.016)	-.397 (.022)	-.356 (.022)
First Price				2.25 (.014)	2.22 (.014)	2.19 (.014)
First 10 Prices				2.31 (.022)	2.26 (.023)	2.17 (.023)
Amazon			.481 (.020)	1.05 (.022)	.790 (.029)	1.02 (.030)
BarnesandNoble			.193 (.023)	.590 (.025)	.370 (.031)	.604 (.032)
Borders			.269 (.020)	.402 (.022)	.109 (.029)	.363 (.031)
A1Books					*-.043 (.030)*	.188 (.032)
Kingbooks					-.495 (.028)	-.283 (.029)
1Bookstreet					-.469 (.034)	-.164 (.037)
Alphacraze					*.036 (.031)*	.224 (.032)
Alphabetstreet					-.789 (.038)	-.749 (.037)
Shopping.com					-.446 (.034)	-.235 (.035)
Fat Brain					-.438 (.038)	-.151 (.039)
Classbook					.205 (.043)	.238 (.044)
Books.com					-.759 (.034)	-.618 (.034)
Other Retailers	0	0	0	0	0	0
Log Likelihood	-99,151	-98,349	-98,030	-98,090	-78,833	-78,471
Adjusted U^2	.2806	.2864	.2888	.2883	.4280	.4307

* Standard Errors listed in parenthesis. Italicized results are insignificant at p<.05. Adjusted U^2 = 1-(LL(*)-# of variables)/LL(0) (Ben-Akiva Lerman 1985, p. 167). N=39,654 sessions.

As noted above, these coefficients can be interpreted as factor weights in a latent utility index. Thus, using equation (2) one can calculate the increase in shipping price that would be required to offset an increase in item price as:

$$\Delta P_{SHIP} = \frac{\beta_{ITEM}}{\beta_{SHIP}} \Delta P_{ITEM} \tag{8}$$

From (8) we see that, in contrast to Morwitz, Greenleaf, and Johnson, Table 4 indicates that consumers are nearly twice as sensitive to changes in shipping price as they are to changes in item price. This suggests that merchants would increase their sales by posting zero shipping costs and including the full price of these services in their item prices.

We also note that within shipping bands (specification 6) customers are significantly more sensitive to changes in shipping price for express and priority shipping than they are to changes in book rate shipping where price sensitivity is statistically the same as it is for item price. We discuss this result in more detail below.

While there is obvious collinearity between price and position in table, our basic results regarding sensitivity to shipping price hold both before and after introducing these variables into the model. In evaluating the reliability of these models standard errors are generally stable across specifications suggesting that collinearity is not a significant problem for our analysis. This inference is confirmed in other standard tests of data collinearity.[5]

With regard to position in table, as might be expected in this setting we find that relative prices matter more than absolute prices, as reflected in the significance of the order in table coefficients. In particular being listed first or in the first page in a table has a strong effect on consumer evaluation of an offer. We also note that while it is not significantly different from item price in all specifications, sales tax also plays a strong role in consumer evaluations of offers. This result is consistent with Goolsbee (2000) who finds increased shopping over the Internet in areas with high state and local sales tax rates.

In confirming these results it is possible to evaluate the effect of the IIA assumption on our analysis. To address this concern, we analyze a nested logit model where consumers choose whether to purchase from a branded or unbranded retailer at the top level and then choose a particular retailer within these categories at the second level nest. We model choice in this way following Brynjolfsson and Smith (2000b) who find empirical support for this decision process and who find significantly better predictive performance for this

[5] For example, R^2 tests on linear regressions involving key explanatory variables do not reveal significant increases as would be expected if collinearity was present.

decision model against a model where shipping bands are used as the top level decision process. The results from our nested logit analysis are given in Tables 3 and 4 below.

Table 3: Nested Logit Model: Top Nests

Variable	*Coefficient*
Price Difference if Brand Lowest Price	.033 (.006)
Price Difference if Unbranded Lowest Price	.046 (.002)
Prior Last Click Brand	.347 (.035)
Lowest Priced Category	1.030 (.023)
Branded Retailer	.486 (.034)
Unbranded Retailer	0

* Standard Errors are listed in parenthesis. Italicized results are insignificant at p<0.10. n=39,654 sessions

As with Brynjolfsson and Smith (2000b), the top level choice between branded and unbranded retailers is modeled as arising from (1) the difference between the lowest priced branded offer and the lowest priced unbranded offer when branded retailers have the lowest price and (2) the analogous variable if unbranded retailers have the lowest price, (3) the choice with regard to branded or unbranded retailers observed in any prior visits by the consumer, (4) a dummy variable taking on the value of 1 for all offers coming from the lowest priced category (branded or unbranded), and (5) a dummy variable for branded retailers. The variables in the lower level nests are the same as those in table 2 except that we add a variable for the offer with the lowest price in each nest ("Best Price in Nest"). We estimate our nested logit model sequentially as described in Ben-Akiva and Lerman (1985, pp. 297-298) and Guadagni and Little (1998).[6]

The results in Table 4 are consistent with the results presented above for the multinomial logit model: consumers are significantly more sensitive to changes in shipping price than they are to changes in item price with this result being more observed more strongly among branded customers.[7] We also note that the similarity in the multinomial and nested

[6] Sequential estimation produces consistent but asymptotically inefficient estimates, causing the standard errors to be too small (Amemiya 1978). However, it has been shown that in many applications the resulting standard errors are not significantly different from those resulting from Full-Information Maximum Likelihood estimation (Bucklin and Gupta 1992, p. 205). Given the strong significance of nearly all our coefficient estimates it is highly unlikely that Full Information Maximum Likelihood estimation would change our results.

[7] Because the specifications in Table 12 control for different retailers (by construction) it is infeasible to use the same techniques presented in section 4.4 to compare coefficients between nests.

logit results with regard to coefficients and predictions also provides confirmation that the IIA problem does not significantly impact our previous results.

Table 4: Nested Logit Model: Bottom Nests

	Branded Retailers	*Unbranded Retailers*
Price		
Item Price	-.053 (.004)	-.041 (.001)
Shipping Price	-.370 (.008)	-.060 (.002)
Weighted Sales Tax	*-.035 (.022)*	-.080 (.013)
Position in Table		
First Price Listed	1.220 (.048)	1.552 (.061)
In First 10 Prices	1.009 (.046)	2.553 (.030)
Best Price In Nest	.367 (.041)	.804 (.060)
Delivery Time		
Delivery Average.	-.021 (.002)	-.028 (.002)
Delivery "N/A"	-.310 (.068)	-.436 (.026)
Retailer Brand		
Amazon.com	.727 (.039)	
BarnesandNoble	.317 (.041)	
Borders	0	
A1Books		*.005 (.034)*
Kingbooks		-.407 (.031)
1Bookstreet		-.333 (.039)
AlphaCraze		.180 (.035)
AlphabetStreet		-.806 (.040)
Shopping.com		-.356 (.038)
Fat Brain		-.288 (.042)
Classbook.com		.312 (.049)
Books.com		-.649 (.037)
Other Retailers		0

* Standard Errors are listed in parenthesis. Italicized results are insignificant at p<.05. (Branded Retailer n=4,023, Unbranded Retailers n=11,480)

The source of the difference between our results and those of Morwitz, Greenleaf, and Johnson is likely due to the difference in consumer cognitive processing costs when associating the base price and surcharge at a retailer's web site and at a shopbot. As noted above, partitioned prices are typically used in a situation where it is computationally difficult for the consumer to compute the total price from the separate base price and surcharge information. In contrast, at most shopbots shipping cost and tax are included in the total price and identified separately in the offer comparison table, making the effect of shipping cost and tax on the offer price fully observable to the consumer.

Still, finding a higher sensitivity to shipping costs than item price is surprising insofar as it conflicts with the most straightforward application of utility theory and rational consumer behavior. We would expect that if there were no cost to calculate the total price, the effect of a \$0.01 increase in price would be the same whether it enters total price through item price or through shipping cost or sales tax. Apparently this is not the case for at least some of DealTime's consumers, presenting us with an economic puzzle. There are several possible explanations for these findings.

First, consumers may be considering the fact that shipping and handling charges are non-refundable in the event that they return the book. In this case, the expected cost of a book would be

$$E(P) = SHIPPING + (1-a)(ITEM + TAX) \tag{9}$$

where a is the probability of returning the book. However, for this to explain all of the observed difference in response to item price and shipping costs, consumers would have to estimate that the probability of making a return is 48% (i.e., 1- $\beta_{item} / \beta_{shipping}$). This is much higher than the 3-5% return rate observed in the monthly sales reports from DealTime.com's associate program relationships with its retailers.

A second possible explanation is that customers could be planning to purchase multiple books and are factoring savings in shipping prices across all products in their order. Two factors are important to consider in this regard. First, DealTime has a "shopping cart" service which allows customers planning to purchase multiple books to generate a single price comparison for all books in their shopping cart. We would expect that consumers interested in purchasing multiple books would use this service. Since our data does not include sessions from the multiple book service we expect only to see single item shoppers in our data. However, if customers were not aware of the shopping cart feature or choose not to use it we also note that shipping costs normally are of a form where the marginal cost for shipping the first book in an order is 3 to 3.5 times larger than the marginal cost to ship additional books. For this reason a customer who sees a savings of \$1 on shipping for the first book can only expect a savings of \$0.30-0.33 on additional books in the order. Thus, a small correlation in item prices within stores would wipe out gains from relying on shipping costs as a signal of savings on additional books in an order.

A third explanation for the increased sensitivity of consumers to shipping prices is that consumers are simply opposed to paying for costs they perceive to be unrelated to the product. A consumer may perceive that a dollar paid to a publisher (and eventually, in part to the author) is different than a dollar paid to a store, a shipper, or to the government (in the case of taxes). Similarly, consumers may object to prices they believe to be "unfairly" high (Kahneman, Knetsch, Thaler 1986) such as handling charges typically added to shipping costs.

Prospect theory (Kahneman and Tversky 1979; Thaler 1985) offers fourth possible explanation. Consumers may be using different reference prices for shipping costs and item prices. For example, consumers may be using a low (possibly zero) reference price for shipping charges and a higher reference for item price, having strong negative reactions to increases in price above their reference price for each price category. A fifth, and closely related, possibility is that consumers evaluate percentage changes in prices — responding more strongly to an increase in shipping cost from $3 to $4 than an increase in item price from $30 to $31.

A sixth possibility is that consumers are planning to make multiple purchases from the retailer over several shopping visits, and are taking into account how lower shipping costs will effect their total purchase price over multiple items.[8]

There may also be other explanations and this finding deserves more study. It would be interesting to focus on differences in consumer response to partitioned prices between a typical Internet retailer's web site where base prices and shipping costs are presented separately and a shopbot where they are presented together. Comparing our results with

[8] DealTime offers a (separate) service for consumers making multiple book purchase at the same time. This service searches for the best deal on the combination of books, even suggesting deals that span two or more retailers. By not including these consumers in our analysis, we automatically control for the possibility that these results are due to consumers evaluating total shipping costs on multiple books.

those of Morwitz, Greenleaf, and Johnson, such an investigation might reveal that retailers should adopt differential pricing strategies with respect to shipping charges for shopbot consumers and web site direct consumers. Similarly, one could analyze price comparison behavior among web shoppers from a prospect-theoretic or cognitive processing context. As noted above, a possible explanation for our results is that consumers respond non-linearly to price changes and have separate mental accounting functions for the different elements of price. Non-linear response is also seen in the importance of an offer's position in the price comparison table reflected in Table 8 columns 2-6 and may be explained by prospect theory or the cognitive processing costs of evaluation additional offers.

3.2. Contractible and Non-contractible Product Characteristics

Another aspect of competitive behavior in Internet markets pertains to how consumers respond to contractible and non-contractible aspects of the product. Contractible aspects of the product bundle include aspects where consumers have clear avenues of recourse if the retailer does not deliver what they had promised such as the characteristics of the physical product or the product's price. Other aspects of the product bundle, such as delivery time, are non-contractible. It is difficult, if not impossible, to force the retailers to deliver a product within the time frame quoted to the consumer.

In the presence of non-contractible product characteristics, economic theory predicts that consumers will use a retailer's brand name as a proxy for their credibility in fulfilling their promises on non-contractible aspects of the product bundle (e.g., Wernerfelt 1988). Moreover, consumers who are more sensitive to non-contractible aspects of the product bundle should disproportionately use brand in their evaluation of product offers.

To investigate how consumers respond to non-contractible aspects of the product bundle we assume that consumers who sort the offer comparison tables based on elements of shipping time (e.g., shipping service, shipping time, and total delivery time) are more sensitive to accuracy in delivery time than consumers who accept the default sorting (total price). We then compare the responses of these two sets of consumers to relevant aspects of the product bundle (Table 5).

Table 5: Sorting Based on Shipping versus Price

	Coefficients
Item Price	-.194 (.001)
Shipping Price	-.371 (.002)
Tax	-.206 (.020)
No Tax (0/1)	.527 (.039)
Average Delivery Time	-.019 (.001)
Delivery “N/A”	-.370 (.015)
Branded Retailers	.292 (.014)
Prior Last Click	.545 (.028)
Prior Click	-.124 (.064)
Differential Coefficients for consumers who sort on shipping	
Sort on Shipping * Item Price	.080 (.014)
Sort on Shipping * Shipping Price	.297 (.019)
Sort on Shipping * Average Delivery Time	-.053 (.011)
Sort on Shipping * Branded Retailers	.985 (.222)

* Standard Errors listed in parenthesis. All results are significant at p<.05. N=39,548 sessions (39,422 sessions sort on total price or item price, 126 sessions sort on shipping time, delivery time, or shipping service).

The selected variables include the differential response of consumers who sort on shipping columns to the product’s item price, shipping price, average delivery time, and a dummy variable identifying whether the product is sold by a branded retailer. These variables were chosen using a likelihood ratio test to compare the restricted model (in Table 5) to an unrestricted model where all variables are allowed to vary between consumers who sort on shipping and consumers who sort on price. The likelihood ratio test failed to reject (p<.05) the null hypothesis that there is (jointly) no difference in the response of consumers who sort on shipping and consumers who accept the default sorting method to tax, the no tax dummy variable, and delivery “N/A.”[9] It is important to note that by separately modeling the response of customers who sort of shipping and customers who accept the default sorting, we control for the effect of the IIA assumption

[9] We note that, with the exception of delivery “N/A,” in each case the restrictions make intuitive sense. There is little reason to believe that consumers who sort on shipping time should respond any differently to the variables relating to tax or retailer loyalty. The fact that there is no statistical difference between the two groups’ response to delivery “N/A” is more surprising as we would expect consumers who care about shipping time to be more sensitive to situations where the retailer is unable to quote an acquisition time.

since. In a nested logit setting where “sort on shipping” is the upper level decision, the lower level regressions would be the same as the ones used for this analysis.

Our results show that consumers who care about accuracy in delivery time are, not surprisingly, less sensitive to item price and shipping price and more sensitive to average delivery time. Strikingly, these consumers are also more than four times more sensitive to the presence of brand in an offer than consumers who sort in price. These results confirm the economic intuition above. Consumers who care about non-contractible aspects of the product bundle appear to use retailer brand as a proxy for credibility.

This result may also explain a comparison of our results for frequent versus infrequent visitors. It is possible that frequent book purchasers are more likely to be sensitive to quality service as a function of their motivation for making the frequent purchases. To analyze this we classify cookies that appear only once in our 69-day sample as representing “infrequent visitors” and cookies that appear multiple times in our sample as representing “frequent visitors”. We present multinomial logit model results for these two groups of consumers in Table 6.

Table 6: Comparison of Frequent and Infrequent Visitors

	Frequent Visitors	*Infrequent Visitors*
Item Price	-.179 (.002)	-.228 (.003)
Shipping Price	-.343 (.003)	-.423 (.004)
Tax	-.422 (.017)	-.473 (.025)
Average Delivery Time	-.018 (.001)	-.019 (.001)
Delivery “N/A”	-.330 (.018)	-.448 (.026)
Branded Retailers	.344 (.017)	.260 (.024)

* Standard Errors listed in parenthesis. Italicized results are insignificant at p<.05. N=26,390 sessions.

Because each model has a unique and unidentified scale parameter (Swait and Louviere 1993), we are unable to directly compare coefficients across model specifications. However, following Smith (2000) it is possible to compare coefficients across model runs after normalizing to a common variable within each specification. Normalizing in this manner cancels the scale parameter and provides a common basis for comparison. In our case, we normalize each coefficient in Table 8 as follows

$$\beta_i' = -\frac{\mu_s \beta_{is}}{\mu_s \beta_{js}} = -\frac{\beta_{is}}{\beta_{js}} \tag{10}$$

where *j* is item price and *s={frequent visitors, infrequent visitors}*. Thus, as in equation (8) we express each coefficient in terms of its dollar value impact on a consumer's evaluation of the product bundle. Our results from this normalization are shown in Table 7 and the resulting standard errors ($s_f / \sqrt{n_f}$) are listed in parentheses.

In each case, we test the null hypothesis that the normalized coefficients are equal using the standard t-test for $\mu_a = \mu_b$ with σ_a and σ_b unknown and $\sigma_a \neq \sigma_b$

Under this test, we reject the null hypothesis for average delivery time and the presence of brand at p=0.05, finding instead that frequent visitors are more sensitive to average delivery time and the presence of brand. We fail to reject the null hypothesis for the normalized coefficients on shipping price, tax, and delivery "N/A." [10] Consumer response to these coefficients is statistically the same for frequent and infrequent visitors. One possible explanation for this finding is that, as in the results in Table 5, frequent purchasers are more sensitive to elements of service quality and this is reflected in using brand as a proxy for this non-contractible element of the product. We also note that this finding does not support the conventional wisdom that regular users of shopbots will, over time, rely on brand less in their purchase behavior.

Table 7: Comparison of Frequent and Infrequent Visitors, Normalized by Item Price

	Frequent Visitors	*Infrequent Visitors*
Shipping Price/Item Price	-1.911 (.024)	-1.853 (.030)
Tax/Item Price	-2.355 (.095)	-2.073 (.111)
Avg. Delivery Time/Item Price	-.101 (.004)	-.083 (.005)
Delivery "N/A"/Item Price	-1.840 (.101)	-1.960 (.117)
Branded Retailers/Item Price	1.916 (.097)	1.136 (.108)

* Standard Errors listed in parenthesis. Italicized results are insignificant at p<.05. N=26,390 for frequent visitors and 13,264 for infrequent visitors.

[10] Applying this test methodology to the unrestricted models for consumers who sort on shipping time and consumers who sort on price yields the same results as expressed in Table 7.

4. Conclusions

Internet shopbots provide a setting for consumer choice that closely resembles the idealized setting commonly assumed in consumer research. By evaluating data from such a setting we are able to assess the importance of pricing and branding strategies in the Internet bookselling market.

We find that consumers react quite differently to a marginal increase in the total price of a book when that increase is ascribed to shipping costs than they do when it is allocated to the base item price: consumers are much more averse to paying shipping charges. We also find that consumers who care about shipping time pay much more attention to branding when they make their decisions than do the consumers who make decisions based on the default sort order, which is not based on shipping time.

Our findings suggest that partitioned pricing strategies that increase demand among web site direct consumers may decrease demand among shopbot consumers. In particular, most of the retailers in our study would appear to be able to costlessly increase demand for their product from shopbots consumers, while keeping the same total price, by simply allocating more of their stated costs to the item price instead of shipping. Our findings strongly suggest that offering “free shipping” would be a successful strategy.

Our results also suggest that consumers use brand name as a signal of reliability in service quality for non-contractible aspects of the product bundle such as shipping. These results may derive from service quality differentiation, asymmetric market information regarding quality, or cognitive lock-in among consumers. While books are a relatively well-specified, homogeneous commodity, the fact that branding is important even here suggests that the branding will be even more important in Internet markets for less homogeneous goods and services, especially when they have important non-contractible characteristics.

Finally, for academic researchers, our results demonstrate the feasibility of using Internet shopping data to better understand consumer behavior in electronic markets. Future research in this regard may be able to extend these results to better understand how web site direct and shopbot consumers respond to partitioned prices, to evaluate the cognitive processing costs of shopbot consumers, and to empirically analyze the application of personalized pricing strategies to shopbot consumers. Moreover, our results suggest that the quantity and quality of data available in Internet markets may introduce a revolution the analysis of consumer behavior rivaling that of the scanner data revolution in the 1980s.

References

Amemiya, Takeshi. 1978. On a two-step estimation of a multivariate logit model. *Journal of Econometrics* **8**(1) 13-21.

Bailey, Joseph P. 1998. Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce. Ph.D. Thesis, Technology, Management and Policy, Massachusetts Institute of Technology, Cambridge, MA.

Ben-Akiva, Moshe, Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press, Cambridge, MA.

Brynjolfsson, Erik, Michael Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* **46**(4)563-585.

Brynjolfsson, Erik, Michael Smith. 2000b. The Great Equalizer? Consumer Behavior at Internet Shopbots. MIT Working Paper.

Bucklin, Randolph E., Sunil Gupta. 1992. Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *Journal of Marketing Research* 29(2) 201-215.

Clay, Karen, Ramayya Krishnan, Eric Wolff, Danny Fernandes. 1999. Retail Strategies on the Web: Price and Non-price Competition in the Online Book Industry. Working Paper, Carnegie-Mellon University, Pittsburgh, PA.

Clemons, Eric K., Il-Horn Hann, Lorin M. Hitt. 1998. The Nature of Competition in Electronic Markets: An Empirical Investigation of Online Travel Agent Offerings. Working Paper, The Wharton School of the University of Pennsylvania, Philadelphia, PA.

Fader, Peter S., Bruce G.S. Hardie. 1996. Modeling Consumer Choice Among SKUs. *Journal of Marketing Research* **33**(November) 442-452.

Goolsbee, Austan. 2000. In a world without borders: The impact of taxes on Internet commerce. *Quarterly Journal of Economics* **115**(2) 561-576.

Greene, William H. 1997. *Econometric Analysis.* Prentice Hall, Upper Saddle River, NJ.

Guadagni, Peter M., John D.C. Little. 1983. A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science* **2**(3) 203-238.

Guadagni, Peter M., John D.C. Little. 1998. When and What to Buy: A Nested Logit Model of Coffee Purchase. *Journal of Forecasting* **17** 303-324.

Kahneman, Daniel, Jack L. Knetsch, Richard Thaler. 1986. Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *The American Economic Review* **76**(4) 728-741.

Kahneman, Daniel, Amos Tversky. 1979 Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263-292.

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105-142.

Morwitz, Vicki, Eric A. Greenleaf, Eric J. Johnson. 1998. Divide and Prosper: Consumers' Reactions to Partitioned Prices. *Journal of Marketing Research*. **35(**November) 453-463.

Smith, Michael D. 2000. Essays on Structure and Competition in Electronic Markets. Ph.D. Thesis, Management Science, Massachusetts Institute of Technology, Cambridge, MA.

Swait, Joffre, Jordan Louviere. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* **30**(3) 305-314.

Thaler, Richard. 1985. Mental accounting and consumer choice. *Marketing Science* **4**(3) 199-214.

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**(3) 458-466.