Modeling Online Browsing and Purchase of Airline Tickets

CIJU T. R. NAIR
TAT Y. CHAN
AMAR CHEEMA*

Manuscript date: July 01, 2006

Please do not cite or quote without permission

* Ciju Nair [contact person] is a Doctoral Candidate in Marketing, and Tat Chan and Amar Cheema are Assistant Professors of Marketing at the John M. Olin School of Business, Washington University, Campus Box 1133, One Brookings Dr., St. Louis, MO 63130. Ciju Nair can be contacted at Ph: (314) 809-1539 Fax: (314) 935-4074 Email: nair@wustl.edu. The authors thank the Center for Research in Economics and Strategy (CRES) at the John M. Olin School of Business for the grant to purchase the ComScore database used in this research. The authors also thank seminar participants at Washington University in St. Louis and at the INFORMS Marketing Science 2006 conference for their comments.

Modeling Online Browsing and Purchase of Airline Tickets

ABSTRACT

Online purchases are increasingly becoming a significant portion of total purchases in most product categories. While prior research in marketing has looked at information search and purchase decisions separately, there are no known empirical studies that model them jointly. In this paper we study consumers' online browsing and purchase of airline tickets by using a joint framework and a unique dataset of dynamic click stream panel data from individual households. We use a three-stage model to study the (i) the choice of the first website visited, (ii) the browsing duration of consumers on travel websites before making a purchase (iii) the choice of the website where consumers will make the purchase, and how a latter stage choice is affected by decisions in the previous stages. We simultaneously estimate these three models which constitute a system of non-linear equations using a simulation-based econometric technique. We find significant effects of expected level of expenditure, prior browsing experience and prior purchase and brand strength in determining consumer browsing and purchase behavior. We also find that the choice of the first site visited and browsing duration has a significant impact on choice of the purchase site indicating the importance of modeling simultaneously. Our results are useful for managers as they help in identifying the major determinants of consumer browsing and online purchase behavior, some of which cannot be observed in a brick-and-mortar environment.

Key Words: airline, travel, multinomial choice, search, browsing, purchase, behavior, multi stage models

1. Introduction

Consumers' pre-purchase information search has a significant effect on purchase decisions and has received significant attention from marketing researchers (for a review of offline information search see Beatty and Smith 1987; Moorthy, Ratchford and Talukdar 1997; Punj and Staelin 1983). The internet is the most recent information source and purchase channel available to consumers. The average US consumer browses for more than two hours each day, increasingly spending more time on the internet and less on other traditional media such as TV and radio (Bouvard and Kurtzmann 2002). The share of internet in purchases is also increasing. For instance, nearly one third of the $200 billion travel market purchases were made online by consumers in the US in 2005 (Economist Sep 2005). Travel as a category has also grown a lot and e-ticketing is now standard practice amongst airline companies. Hence it is important for both academicians and marketers to understand online search and shopping behavior.

The present research studies the phenomenon of pre-purchase browsing on the internet in this large and growing domain of online purchases of airline tickets. We do so by focusing on three stages of consumers' decisions: the choice of the first site to visit, the duration of browsing on various sites visited subsequently, and the choice of the site where purchase finally occurs. Past research in marketing has investigated online browsing and purchases independently (Park and Fader 2004, Johnson et al 2004, Montgomery et al 2006, Bucklin and Sismeiro 2003, Sismeiro and Bucklin 2004) or within a specific website (Moe and Fader 2004). Another stream of literature has focused on the effect of search on consideration set formation (Wu and Rangaswamy 2003) which relates to which website consumers start their search for information. However, to our knowledge no study has attempted to jointly study consumers' information search processes and purchase decisions, and how the former impacts the latter.

In this study we extend prior work by simultaneously studying both browsing and purchase behavior after controlling for demographic characteristics. We also extend the applicability of existing discrete choice random utility models that have traditionally been used to study scanner panel data and propose a unified dynamic framework to explore browsing and purchase behavior across multiple websites. We focus on investigating the effects of expected level of expenditure, prior browsing experience, prior purchase and brand strength on each of these three stages, and exploring the effect each stage has on the subsequent stage. In summary,

we try to understand (a) the factors that affect the choice of the first website that consumers visit prior to making a purchase, (b) the factors affecting browsing duration on websites selling airline tickets, (c) the factors affecting the choice of the purchase site in addition to drawing inferences on how consumer behavior differs between travel portals and airline websites, and (d) the dynamic impact of past browsing experience and purchase decisions on current purchase, and that of the choice of first website and browsing duration on the subsequent purchase site choice.

The rest of the paper is organized as follows, in section 2 we review prior research and provide a theoretical background for the present research. In section 3 we describe the data and in section 4 outline the model used to study browsing and buying behavior. Section 5 summarizes and discusses the results. Finally in section 6 we conclude, acknowledge the limitations of the present research, and outline the opportunities that exist for future research.

2. Conceptual Development

From a cost-benefit perspective, consumer search increases as the benefits of search increase and decreases as the costs of search increase (Newman 1977; Punj and Staelin 1983). The online search behavior of consumers is different in many aspects compared to that in store, primarily because it costs less in terms of time and effort for a customer to visit an online versus offline store. As consumers incur higher costs in the form of time and effort spent to visit an offline store, they will be more likely to buy. In contrast, online shoppers are more likely to visit a website without any intention of buying. The low cost of visiting a website makes the shopper more likely to delay a purchasing decision and search broadly on various other websites. Consistent with this expectation, we observe lower conversion rates (number of visitors who buy) online than offline (Moe and Fader 2004a). Still, it does not necessarily imply that consumers will have perfect price or product information on the internet as searching online involves time and effort. Furthermore, due to limited cognitive resources or browsing knowledge consumers may not be able to search online exhaustively (e.g., they may not know which websites to search and compare.) It is therefore important to understand consumer browsing behavior in the online context which is different from the offline context and its impact on the choice of purchase site. We next discuss prior research that identifies some of the factors that affect the extent of search and browsing behavior.

2.1 Factors Affecting Browsing

A significant amount of research has focused on individual and product characteristics that affect pre-purchase search. These research streams are discussed here to identify the variables that will be relevant for the present research.

2.1.1 Consideration Sets

There is plenty of evidence to suggest that in the offline context consumers do not choose products from a universal set of alternatives, but frequently choose from consideration sets that consist of a subset of options (for a review, see Shocker et al. 1991; Roberts and Lattin 1997). Thus, factors that affect consideration set formation exert a strong influence on the final product choice. The effect of prior experience with brand and product category has spawned an entire stream of literature on state dependence and variety seeking behaviors exhibited by consumers (see Khan 1995; Seetharaman, Ainslie and Chintagunta 1999). Consumers differ significantly in their consideration set formation even after controlling for the observable differences in demographic and experience characteristics. Consistent with this expectation, prior research has demonstrated that models accounting for heterogeneity of consideration sets do better than models that do not (Chiang et al 1999).

In the online context consumers also may have a limited consideration set that consists of a subset of websites that they will visit in their browsing process. For example, in the travel category it is almost impossible for consumers to remember hundreds of travel portals and airline websites that sell air tickets. Usually past experience and browsing or purchasing experience will dictate the formation of such consideration set. One of the most important indicators of the consideration set is where a consumer starts from, i.e., the first website that he or she visits. In America, 54% of consumers start with a travel portal, such as Expedia.com, Travelocity.com or Orbitz.com, according to a study by Nielsen/NetRatings 2005. The websites of travel suppliers, such as airlines and hotels, are visited first by 37% of shoppers and the other 9% start planning their trips on travel search firms such as kayak.com and sidestep.com (Economist 2005). Indeed, prior research has identified that not accounting for the first site visited is one of the limitations of any search model (Moe and Fader 2004).

Limitations in consideration set formation imply that the consumer may not search extensively online despite the low cost of visiting a website (as discussed above). Therefore we

expect the first site visited to exert a strong influence on consumer's choice of purchase site. Furthermore, this first site may reveal consumers' preference before they encounter present information, and may have a disproportionately larger impact on consumer preferences than later sites visited, akin to a primacy effect that has previously been documented in attitude formation (Anderson 1965) and in legal decisions (Lind, Kray, and Thompson 2001). In this paradigm, limited cognitive resources and memory force the consumer to pay greater attention to information that is encountered earlier rather than later in the decision process. In click-stream data we observe the first site visited before making a purchase and hence we can say that this site is not only part of the consideration set but ranks at the top in terms of consumer preference. We believe modeling the choice of first site visited is a crucial step in modeling search and, consistent with prior research we expect it to significantly influence purchase behavior.

2.1.2 Site-specific and Category Experience

Consumers with greater amount of site-specific experience will be more likely to have prior preferences in terms of which website they prefer to browse and purchase from. Thus, the first site visited by consumers may reveal a preference that is stronger for these consumers than for consumers who have low experience. Moreover, prior site-specific experience will decrease the effort required by the consumer to learn the site layout and, if necessary, the effort of setting up a user account. Thus, consumers who have prior experience surfing on a site would be more likely to visit that site first the next time they make a purchase.

Prior category experience is also expected to have a significant effect on the amount of browsing that the consumer does prior to purchase. Prior research has offered different predictions on the effect of this variable on search. On the one hand, consumers with prior category experience know a lot about the category already and may thus search comparatively less as the benefit may not be worth additional effort. On the other hand, greater prior knowledge implies that consumers may have a larger consideration set by knowing where to search for information and hence increase browsing duration. It is also possible that prior knowledge increases the ability to absorb more information and hence increases search efficiency. That is, greater category experience results in consumers seeking more information as they are aware of the right attributes to process. Knowledge might thus decrease cognitive cost of processing while increasing the benefit of seeking more information leading to increase in search.

Prior empirical results have also been mixed, with one set of studies finding that search increases with category experience, another set finding that search decreases with category experience, and yet another finding that there is no relationship at all (see Brucks 1985 for a discussion of these studies). Empirical findings using automobile purchases have shown pre-purchase search to be minimal (Beatty and Smith 1987). This empirical finding is puzzling especially in high involvement categories and has been attributed primarily to measurement issues related to self report biases. Srinivasan and Ratchford (1991) show that there is a negative relationship between prior experience and search as long as other variables are controlled for; however, subjective knowledge tends to increase search as knowledgeable consumers tend to structure the problem in complex ways resulting in increased search.

A common result that resolves this contradiction is a non-monotonic relationship which is able to account for the mixed results observed in prior studies (Moorthy, Ratchford and Talukdar 1997; Bettman and Park 1980; Hempel 1969; Johnson and Russo 1984). This provides an explanation for different search efforts under different category experiences. We attempt to investigate how category experience affects browsing duration in the online context by allowing for this non-monotonic relationship using both a linear and a quadratic term for prior browsing experience.

2.1.3 Prior Purchase

The websites where past purchases occurred may have an impact on being chosen as first site to be visited in the current search. This may be primarily due to switching costs associated with learning site layouts and setting up user accounts at each new website. In the present research we use the term *inertia* to indicate the tendency to first visit the website that consumers previously purchased from. In addition to effort-related switching costs, this inertia could also be caused by marketing activities such as frequent flyer programs or promotional offers which were made available to consumers to buy from same website. Thus, we expect that the choice probability of being the first site to be visited would increase if that site was the one where the previous purchases occurred.

Prior purchase also directly contributes to consumer category experience. On the one hand, consumers may browse less because they have prior category experience or because of inertia discussed above. On the other hand, they may know more about the category and search

may become easier. In the context of prior purchase, however, we note that the knowledge is site-specific so the latter effect may be smaller. Overall we expect consumers with prior purchase will browse less than those without.

2.1.4 Expected Level of Expenditure

Prior research has demonstrated that consumers are more likely to search for information when there is higher risk associated with purchase (Punj and Staelin 1983). This risk may be physical (e.g., car safety), social (e.g., style of clothing) or financial (e.g., the price of the product). We use the expected level of the expenditure as a proxy for the financial risk associated with the purchase. We categorize the expected level of expenditure to be low, medium, or high on the basis of the observed prices. Consistent with prior research on perceived risk, we expect consumers to browse more for purchases that are expected to have a higher level of expenditure.

We also explore how the expected level of expenditure affects choice of the first website to be visited. We expect that with greater financial risk consumers will be more likely to stick to tried and trusted sites where they have got good deals in the past. That is, we expect the likelihood of first visiting a travel portal to be higher as expected level of expenditure increases.

2.1.5 Brand Strength

The stronger the brand the greater the likelihood that the site would be visited first as it would be top of mind. We use brand intercepts to capture effects of brand strength on the choice of the first visited site. An alternative interpretation is that these intercepts imply different levels of marketing activities undertaken by these firms. Prior literature has not explored this effect on choice of first site to be visited as the first site visited is not observed in most empirical studies.

2.1.6 Consumer Demographics

Prior research has also demonstrated that consumer demographics like age and income play an important role in the search process. For example older consumers may be more price sensitive (because they are retired and have lower income), but have lower opportunity cost of time compared to busy young consumers and hence search longer and be less likely to visit the same site they previously purchased from (i.e., exhibit lower inertia) than the latter. On the other hand cognitive capabilities of older consumers may be declining and prior research has shown

younger consumers process more cues and alternatives (Schaninger and Sciglimpaglia 1981) and tend to search more in general (Ward and Lee 2000). Other research also finds that older consumers typically have less patience to search (Ward and Lee 2000). For the income effect, we expect high income consumers to be less price sensitive compared to low income consumers.

High connection speeds would make search easier and hence we expect consumers with higher connection speed to visit a greater variety of websites and to view more pages (i.e. exhibit a high level of browsing). Consistent with this expectation, Yonish, Delhagen & Gordon (2002) find that broadband users search 33% more than narrowband users given higher surfing speeds.

2.2 Factors Affecting Purchase

Marketing literature has primarily looked at single stage choice models to analyze in-store purchases. As the factors affecting in-store choice are also applicable to online purchase behavior, we draw on findings in existing literature to understand expected effects of these factors on purchase of airline tickets online.

2.2.1 Site-Specific and Category Experience

Increased frequency of visits to a website has been found to strongly influence propensity of purchase (Moe and Fader 2004). This has also been found to be true even in the offline setting (see Bellinger et al. 1978, Janiszewski 1998, Jarboe & McDaniel 1987, Roy 1994). This could be because consumers can take informed decisions as product and category knowledge increases (Brucks 1985) or could be because the consumer increases the likelihood of purchase with the amount of effort sunk into the decision (Staw 1976). Hence we expect both site-specific and category experience gained by consumers who spend more time surfing in general and on specific websites to positively impact the likelihood of purchasing from those websites.

2.2.2 Prior Purchase

Evidence for inertia or state dependence among consumers is well documented in marketing literature when it comes to in-store brand choice among consumers (e.g., Seetharaman, Ainslie and Chintaguta 1999). Consistent with the effect of prior purchase on first website visited, we expect the likelihood of purchase to be higher for a particular website if the last purchase happened to be on that website.

2.2.3 Expected Level of Expenditure

The expected level of expenditure indicates the amount of financial risk that the consumer takes when they purchase the ticket, with higher expenditure levels making them more hesitant (Punj and Staelin 1983). Expectations of price levels on travel portals in general tend to be lower than that of airlines implying a higher likelihood of finding a better deal on travel portals. Hence we expect consumers are more likely to buy from travel portals when expected level of expenditure increases.

2.2.4 Brand Strength

Brynjolfsson and Smith (2001) find strong brand effects in consumers' choice of websites to visit from a shopbot listing. Also, web site brand equity would create confidence in buyers to buy from a particular website especially when there are fewer product attribute information available online (Degeratu, Rangaswamy and Wu 2000). We expect to find differences in brand strength across websites, with some having stronger brand images than others. Note that the brand effects on purchase decisions may be different from first site visited. Some websites may be more attractive for browsing first (for the purpose of information search) but may be less successful in converting these visits to final purchases than other websites.

2.2.5 Consumer Demographics

Existing literature does not find much significance in demographic variables to segment consumers when it comes to online purchase (Bhatnagar and Ghose 2004). However, Degaratu Rangaswamy and Wu 2000 do a comparison of online and traditional supermarkets and find that income dampens price sensitivities; hence we believe consumer demographics could play a significant role when it comes to predicting site purchase probabilities. Travel portals provide choice of different airlines in addition to usual itinerary details thus increasing the available set of alternatives and cognitive load required to process this information. Gourville and Soman (2005) show more alternatives could decrease propensity to buy. For this reason older consumers, who typically have less cognitive resources to process information, may have a lower likelihood of purchasing from a travel portal. High income consumers on the other hand are known to be less price sensitive and this coupled with the lower information processing required on airline websites should decrease their likelihood of buying from a travel portal.

2.2.6 First Site Visited and Browsing Duration

In this research we explore two additional process effects on the final decision of which site to purchase from: (a) the effects of first site visited, and (b) amount of browsing. As discussed above, the first site visited choices may indicate inertia from prior experience or higher order website preference, and consequently this site may have a higher likelihood of being the one the consumer finally purchases from. Also, the primacy effect would suggest that the first site visited would have a greater likelihood of persuading the consumer to purchase than subsequent web sites. These reasons suggest that if a site is the first site visited in the process of browsing for the present purchase, it is likely to be the one the final purchase is made from.

Prior research suggests that consumers who browse more will be more likely to buy than those who browse less (Moe and Fader 2004). Hence we expect browsing duration to increase the likelihood of purchase. Also, browsing duration may be correlated with consumers' price sensitivity (more price sensitive consumers will browse longer for information) so purchase site choices of those who browse longer may be different from those who browse less. . Note that in our model the browsing duration is also affected by the first web site that consumer visits. Thus, there is a cascading effect of the first site visited on browsing duration and choice of website to finally make the purchase.

3. Data

We use the ComScore clickstream dataset available from the WRDS database for our analysis. This dataset comprises of surfing and transaction details of 100,000 households[1] that are a representative sample of the US population in 27 product categories. In this study we restrict ourselves to the airline category and focus on browsing and purchase behavior of airline tickets as it is one of the categories with the highest number of online purchases. A total of 1832 households in the travel category fit the criteria required for our analysis. To ensure that a household's browsing is only related to a specific observed purchase we use the following three conditions: (1) we only focus on the household's browsing seven days prior to a purchase (the browsing period); (2) we only study the household's browsing in the travel category (travel portals and airline websites) during that seven day period; and (3) on top of that, we only choose households that have had no surfing on travel websites for seven days prior to the browsing

[1] Hereafter we will use "households" and "consumers" interchangeably with the same meaning.

period. Table 1 reports some summary statistics of the data by household for the 1832 households during the six month period from July 2001 to December 2001. In this dataset, households on average make approximately three online purchases in the travel category during the six month period (the median number of purchases is two).Households have a median spending of nearly $300 on a travel purchase and spend a little over four hours on average searching on websites selling travel products to make three purchases on average during the six month period.

Table 1: Summary statistics

	Mean	Median	Std. Dev.
Time spent (minutes)	255.84	155	349.81
Number of pages viewed	259.76	151	418.35
Number of unique websites visited	13.65	9	14.54
Total number of websites visited	20.77	12	28.91
Purchases in travel category	2.87	2	3.89
Expenditure in travel category (US$)	560.53	294	1187.02
Purchases in all categories	9.99	4	17.95
Expenditure in all categories (US$)	856.74	452	2088.95

Travel category forms a significant portion of online purchases made by consumers with the median being two out of four purchases amongst the 27 product categories. As we are interested in studying browsing behavior that is related to a purchase we focus only on those travel websites that also provide an option for consumers to purchase airline tickets. Specifically we investigate browsing and purchase behavior on travel portals (such as Expedia, Orbitz, and Hotwire) and airline websites (such as Southwest, Delta, and American) where consumers have an option to purchase the ticket online.[2]

We used the pages viewed by households in the first three months of data (July 2001 – September 2001) as the household's prior experience on travel websites. We then use the online browsing and purchase sessions in the last three months for model estimation. To study the

[2] We excluded from the analysis those households that were very heavy users (whose purchases exceeded the 99.9th quantile both in terms of amount as well as number of transactions) in the airline category. We also excluded transactions on websites which were auction sites, search engines and payment gateways such as ebay.com, lycos.com, and authorize.net (these constituted less than 5% of the recorded travel purchases). Multiple purchases bought by a household were clubbed together if they occurred at the same time on a particular website (e.g. spouses buying airline tickets).

impact of expected level of expenditure that consumers incur to purchase airline tickets we classify the value of purchase as low, medium and high based on the distribution of prices. We use a median split widely used in marketing literature and use indicators for low (less than 33rd quantile), medium (33rd to 66th quantile) and high expected level of expenditure (higher than the 66th quantile).

In this study we focus on browsing and purchase behavior for airline tickets because of two reasons (i) airlines constitute 52% of (number of) purchases in the travel category and (ii) car rental (12% of purchases) and hotel (25 %) purchases are typically made in conjunction with an airline purchase. Investigating browsing and purchase behavior at the basket level (i.e., including hotel and car rental purchases) could be an interesting future study.

4. Model Specification and Estimation

We propose a three stage model of consumer browsing and purchase behavior and jointly estimate the combined model. The three stages we model are (i) choice of first website visited (ii) duration of browsing on travel websites and (iii) choice of purchase site. This framework is pictorially depicted in Figure 1.

Figure 1: Proposed three-stage model of consumer online browsing and purchase behavior

In the first part of this section we outline the model used to study the choice of the website that is visited first in the purchase process. Initial data analysis revealed a strong correlation between the website at which the most recent purchase was made and the choice of the first website to be visited for the present purchase process (see Table 2). For example, almost 60 percent of households that visit Expedia and Orbitz as the first site in the purchase process purchased from these sites the last time they bought an airline ticket. Similarly, almost 85 percent of the households who visited airline websites first, had purchased their most recent

ticket from airline websites. Understanding how households choose the first site to browse in a purchase process (in addition to inertia) can help us gain insights on the final purchase decision as the first site visited is likely to influence the decision more than subsequently visited sites.

Table 2: Relationship between last purchase and first visited site (%)

Site from which last purchase was made	Site visited first in the present purchase process					
	Expedia	Orbitz	Hotwire	Other travel portals	All airline websites	Grand total
Expedia	**57.9**	10.5	1.3	2.6	27.6	3.9
Orbitz	14.6	**56.1**	1.2	2.4	25.6	4.2
Hotwire	6.7	23.3	**26.7**	3.3	40.0	1.5
Other travel portals	18.2	4.5	0.0	**27.3**	50.0	1.1
All airline websites	6.3	6.0	1.3	2.1	**84.3**	19.6
No last purchase	20.1	13.9	5.1	7.1	53.9	69.7
Grand total	18.4	14.0	4.4	5.9	57.4	100.0

In the second part of this section we model the browsing duration, in particular, we focus on pages viewed by the consumer while browsing prior to a purchase. It is evident from the data that the first site visited tends to get a disproportionate amount of time in terms of total browsing duration. For example, Table 3 shows that if the browsing starts with Expedia or Orbitz, more than 50 percent of the total browsing is on these sites. However, if browsing starts with Hotwire or with other travel portals, the percent of the total browsing in the site visited first is much lower. To model the browsing duration we incorporate choice of the first site, the decision variable in the first stage, to understand how it affects the browsing duration. For simplicity we only study the total browsing duration (category browsing and not site-specific browsing) at this stage. However modeling the path consumers take is crucial for marketing interventions that are related to design of banner ads or promotions (see Montgomery et al 2006).

Table 3: Comparison of browsing on first visited site and all other sites

First site visited	Average number of pages viewed on			Percentage of browsing on first site visited
	First site visited	Subsequent sites	All sites	
Expedia	41.95	27.50	69.46	60.40
Orbitz	35.45	33.86	69.31	51.15
Hotwire	27.09	45.42	72.51	37.36
Other travel portals	10.90	65.06	75.96	14.34
All airline websites	120.68	153.70	274.38	43.98

In the third and last part of this section we discuss the models used to determine the relationship between first website visited, the browsing duration and the purchase website. Table 4 summarizes the impact of the choice of first site visited on the choice of the purchase site. It clearly demonstrates that households are much more likely to buy from the first visited site than from sites they subsequently visit. Additionally, this primacy effect appears stronger for airline websites than for other portals.

As the second and third stages involve the earlier stage decisions, we jointly estimate the three stages as a non-linear simultaneous equation system. Though we estimate a few alternate specifications we outline below the general case where we include individual travel portals and individual airline websites.

Table 4: Effect of first visited site on purchase (%)

Site from which current purchase is made	Site visited first in the current purchase process						
	Expedia	Orbitz	Hotwire	Other travel portals	All airline websites	No first site visited	Grand total
Expedia	**37.2**	10.9	5.7	2.3	13.8	30.1	17.9
Orbitz	10.5	**34.6**	7.1	3.4	9.0	35.3	13.6
Hotwire	17.2	17.2	**20.7**	8.0	14.9	21.8	4.5
Other travel portals	17.4	15.7	3.3	**9.9**	13.2	40.5	6.2
All airline websites	15.3	10.7	5.0	4.8	**64.2**	0.0	32.9
Grand total	15.0	11.9	4.8	3.4	26.3	38.6	100.0

4.1 Modeling Choice of First Site Visited

To study the choice behavior of the first website that consumers visit indicating the start of browsing and information search we use a random coefficients approach of the traditional multinomial logit model (for example see Gudagni and Little 1983) which we explain in detail later. We classify websites that consumers choose to visit first into travel portals and airline websites. We pick the top three travel portals and separately club all other travel portals and all airline websites together.[3] A website is defined to be the first website visited prior to a purchase

[3] We do not observe from data the departure and arrival airports of flights. As airlines do not fly every route and browsing and purchase behaviors may be mainly determined by whether or not a specific route is served by an airline (e.g., one may not visit Southwest Airline's website when flying to the JFK Airport in New York), we choose to group all airline websites together. In comparison, one can buy tickets flying every route served by

if it is the first website that is visited within a seven day window prior to a purchase with no surfing history on travel websites for at least seven days prior to that first visit. On average we find that airlines tend to have a higher conversion rate compared to travel portals (see Table 5).

Table 5: Website conversion rates [4]

Site	Number of visitors	Number of transactions	Transaction share (%)	Conversion (%)
Expedia	37508	2378	19.0	6.3
Orbitz	26613	1564	12.5	5.9
Hotwire	10990	686	5.5	6.2
Other travel portals	11822	655	5.2	5.5
All airline websites	54687	5280	42.2	9.7

Let a discrete variable $F_{ijt} = 1$ indicate that consumer i visits website j first at time period t, and $F_{ijt} = 0$ otherwise. For $F_{ijt} = 1$, website j has to exist in consumer i's consideration set (which may not include all possible options) and then j has to dominate other websites in this consideration set in terms of information search under cost-benefit evaluation. We assume that these are determined by a list of factors including customer demographics (age, income, connection speed) Z_{it}, prior category experience H_{ijt}, expected level of expenditure P_{it} and the site-specific prior experience S_{ijt}. Prior category experience H_{ijt} is measured as the proportion of pages viewed on website j to the total pages viewed in the first three months on all websites selling travel products, and site-specific prior experience S_{ijt} is measured as average daily pages viewed on website j in the first three months. We also incorporate the effect of expected level of expenditure P_{it} by classifying the final purchase price into three categories of low ($\leq \$108$), medium (>\$108 and $\leq \$356$) and high (>\$356) expenditures[5]. Furthermore, I_{ijt}

different airlines from most of the travel portals. The behavior captured in our first site visited model is whether or not a household will start information search with travel portals or any airline websites and, in the first case, which travel website the household is more likely to choose.

[4] In our dataset Travelocity has a 100% conversion rate. This is because we do not observe any history or search for purchases made on Travelocity. We therefore exclude Travelocity from our analysis.

[5] Note that this cannot be interpreted as the price effect. We only observe from our data the final purchase price; however, we do not have the price information from other websites where consumers visited but did not purchase.

is an indicator variable that denotes whether or not consumer i's last purchase was at j. This variable may affect the probability of j being in i's consideration set and may create inertia such that i may be more likely visit the same website first during the next purchase cycle. Finally, first visited site choice is also affected by i's preference for or familiarity of website j that is independent from the above factors as well as j's marketing activities which are unobserved from our data. This is termed as "brand strength" which is individual-and-time-specific in our model.

We assume that there is a latent variable F_{ijt}^* that generates the first visited site decisions. $F_{ijt} = 1$ if and only if $F_{ijt}^* \geq F_{ikt}^*$, for all other website k. We specify the function of this latent variable as

$$F_{ijt}^* = \alpha_{ijt}^f + \beta^f Z_{it} + \tau^f H_{ijt} + \gamma^f S_{ijt} + \lambda_i^f P_{it} + \rho^f I_{ijt} \tag{1}$$

In the above equation the superscript "f" denotes the model of first visited website. The variable α_{ijt}^f represents the latent website brand strength. We use a random effects approach to model this variable as the follows

$$\alpha_{ijt}^f = \alpha_j^f + \xi_{ij}^f + \varepsilon_{ijt}^f \tag{2}$$

where α_j^f represents the mean brand strength that will be estimated as parameters, ξ_{ij}^f represents the individual-specific but time-invariant random effect for brands, and ε_{ijt}^f is the individual-and-time-specific idiosyncratic shock that we assume to be i.i.d. type one extreme value distribution.

We assume that $\lambda_i^f = \lambda^f + \eta_i^f$, where η_i^f is a time-invariant and individual-specific random variable which captures the consumer heterogeneity in response to expected expenditure level.[6] We allow ξ_{ij}^f and η_i^f to be correlated among themselves.[7] As we will explain later, one distinct aspect of our estimation model is that we also allow these random effects to be correlated

---

Hence we cannot identify how prices offered from different websites affect the browsing and final purchase behavior. Instead, consumers usually have a perception of how expensive a ticket will be, e.g., flying from New York to Los Angeles will be more expensive than flying from New York to Boston, and this will affect how much time and effort they invest in information search as well as purchase site decisions. We use the above categorization that is based on the final purchase price as a proxy for such a perception.

[6] For simplicity we assume that only brand intercepts and the price coefficients are heterogeneous across consumers in all three stage of decision-making.

[7] Such correlations can be identified though the panel structure in our data.

with the random effects in the other stages of the model. Hence the dimensionality of parameters is very large considering the all the correlation coefficients in our three-stage model.

To ensure proper identification we normalize the latent variable value for all airlines as $F^*_{i,AIR,t} = 0 + \varepsilon^f_{i,AIR,t}$.[8] In our estimation model we also incorporate the interactions of demographic variables Z_{it} with all other covariates H_{ijt}, S_{ijt}, P_{it}, and I_{ijt}. Let T_i be the total number of household i's purchases observed in data. Correspondingly there are T_i first visits. Under the type one extreme value distribution for ε^f_{ijt} and conditional on the random effects ξ^f_{ij} and η^f_i, we can write down the probability that a household's history of first visits in the whole sample period as below.

$$\Pr(i\text{'s history of first visits}) = \prod_{t=1}^{T_i} \Pr(F^*_{ijt} \geq F^*_{ikt}, \forall k)$$
$$= \prod_{t=1}^{T_i} \frac{e^{\alpha^f_j + \xi^f_{ij} + \beta^f Z_{it} + \tau^f H_{ijt} + \gamma^f S_{ijt} + (\lambda^f + \eta^f_i) P_{it} + \rho^f I_{ijt}}}{1 + \sum_{k=1}^{J} e^{\alpha^f_k + \xi^f_{ik} + \beta^f Z_{it} + \tau^f H_{ikt} + \gamma^f S_{ikt} + (\lambda^f + \eta^f_i) P_{it} + \rho^f I_{ikt}}} \quad (3)$$

4.2 Modeling Browsing Duration

We quantify search as the pages viewed by consumers on travel selling websites seven days prior to purchase of a product in the travel category. To check model robustness we also use time spent on travel websites and find very similar results. However, we believe that pages viewed is a more reliable measure since it is less prone to contamination or noise compared to time spent where users could typically open a page and then leave it while they attend to other chores and are not necessarily in front of the computer. We choose seven days prior to purchase to be on the safer side though a significant portion of the search occurs only three days prior to purchase (see Table 6). We exclude the pages viewed on the day of purchase (day 0) because a large proportion of it is related to transaction completion and would only add noise to the actual browsing duration. If there were multiple transactions in this seven day window, we exclude all but the first which is not preceded by any other purchase, from our analysis.

[8] Because of this normalization the coefficients corresponding to all of the variables in equation (1) have to be interpreted as the difference in probabilities that consumers visit travel portals relative to airline websites.

We use the log of browsing duration as the endogenous variable in our model and assume this is affected by the following factors. First, we include the customer demographics (age, income, connection speed) Z_{it} as covariates. Next we also incorporate a prior browsing experience H_{it} which is measured by the average number of daily pages viewed in the travel category in the first three months. As we discussed above, there may be a non-monotonic relationship between browsing duration and prior browsing experience. To capture this nonlinear relationship we incorporate H_{it} and its squared term as covariates in our model.[9] As expected expenditure level will also affect the time and effort a consumer invests in search for information, we include the indicators P_{it} ("medium" and "high" levels) used in the first visited site model in this stage. To distinguish the behavior difference between "light" and "heavy" users we use two indicators $I_{p\leq 1}$ (indicator takes value of 1 for users with zero or one purchase in the first three months) and $I_{p>1}$ (indicator takes value of 1 for users with more than one purchase in the first three months) to represent light and heavy users, correspondingly. Finally, as discussed above, the first visited site choice seems to have an important impact on the browsing duration. The vector of discrete choice variables $F_{it} = (F_{i1t},...,F_{iJt})'$ in the first stage, where $F_{ijt} = 0$ or 1 and $\sum_{j} F_{ijt} = 1$, are included as covariates in the browsing duration model.

Let D_{it} be the log of browsing duration of i in period t, and α_{it}^{d} be the individual-and time-specific intercepts in the model representing consumer heterogeneity in browsing behavior. The browsing duration model is specified as the follows:

$$D_{it} = \alpha_{it}^{d} + \beta^{d} Z_{it} + \lambda_{i}^{d} P_{it} + \phi^{d} I_{p\leq 1} + \delta^{d} I_{p>1} + \lambda_{1}^{d} H_{it} + \lambda_{2}^{d} H_{it}^{2} + \chi^{d}{}' F_{it} \tag{4}$$

The latent variable α_{it}^{d} is specified as

$$\alpha_{it}^{d} = \alpha^{d} + \xi_{i}^{d} + \varepsilon_{it}^{d} \tag{5}$$

where α^{d} represents the mean intercept in the model to be estimated, ξ_{i}^{d} represents the individual-specific but time-invariant random effect for browsing duration, and ε_{it}^{d} is the individual-and-time-specific idiosyncratic shock that we assume to be i.i.d. normally distributed,

[9] Note that this variable is different from H_{ijt} defined in the first visited site model. Here category browsing experience is not site specific.

i.e., $\varepsilon_{it}^{d} \sim N(0,\sigma^2)$. We assume that $\theta_i^d = \theta^d + \eta_i^d$, where η_i^d is a time-invariant and individual-specific random variable which captures the consumer heterogeneity in response to expected expenditure level. Similar to the first visited site model, we allow ξ_i^d and η_i^d to be correlated among themselves. These random effects are also allowed to correlate with random effects in other stages.

Table 6: Browsing behavior prior to purchase

Days prior to purchase	Time spent (minutes)	Average number of pages viewed
-1	8.97	9.65
-2	5.23	5.69
-3	4.12	4.48
-4	3.24	3.57
-5	3.01	3.31
-6	3.04	3.29
-7	2.77	3.10
-8	2.51	2.70
-9	1.95	2.23
-10	1.94	1.99
-11	1.67	1.93
-12	1.70	1.89
-13	1.74	1.99
-14	1.88	2.15
-15	1.62	1.76

Let T_i be the total number of household i's purchases observed in data. Correspondingly there are T_i browsing durations. Under the distribution assumption for ε_{it}^d, i's history of browsing duration is generated through a normal process conditional on ξ_i^d and η_i^d. The likelihood function of observed history of browsing duration is the following

$$\begin{aligned}&\Pr(i\text{'s history of browsing duration})\\&=\prod_{t=1}^{T_i}\frac{1}{\sigma\sqrt{2\pi}}\exp(-[(D_{it}-(\alpha^d+\xi_i^d+\beta^d Z_{it}+(\lambda^d+\eta_i^d)P_{it}+\phi^d I_{p\le 1}+\delta^d I_{p>1}+\lambda_1^d H_{it}+\lambda_2^d H_{it}^2+\chi^d{}'F_{it}))^2/2\sigma^2]\end{aligned} \tag{6}$$

4.3 Modeling Choice of Purchase Site

Let a discrete variable $U_{ijt}=1$ indicate that consumer i finally purchases from website j after his or her browsing at time period t, and $U_{ijt}=0$ otherwise. Again we assume that there is

a latent variable U_{ijt}^{*} that generates the first visited site decisions. $U_{ijt}=1$ if and only if $U_{ijt}^{*} \geq U_{ikt}^{*}$, for all other website k. Similar to earlier stages, we assume that U_{ijt}^{*} is a function of a list of factors including consumer demographics Z_{it}, expected level of expenditure indicators P_{it}, prior experience as the average daily pages viewed S_{ijt} on different sites in the first three months, H_{ijt} as a proportion of pages viewed on website j to the total pages viewed on all travel websites that the customer visited in the first three months, and whether or not consumer i *'s* last purchase was at j, I_{ijt}.

As discussed before, we expect that the choice of the first website F_{ijt} and the actual browsing duration D_{ijt} may be important determinants for the final purchase site decisions. Hence we also include the decision variables F_{ijt} and D_{ijt} in the first visited site and browsing duration models as covariates in this latent variable function. Therefore we can write down

$$U_{ijt}^{*} = \alpha_{ijt}^{p} + \beta^{p} Z_{it} + \tau^{p} H_{ijt} + \gamma^{p} S_{ijt} + \lambda_{i}^{p} P_{it} + \rho^{p} I_{ijt} + \chi^{p}{}' F_{it} + \delta^{p} D_{it} \tag{7}$$

where α_{ijt}^{p} represents the individual-and-time-specific random effect on the purchase site decision. Similar to earlier specifications we model this variable as the follows

$$\alpha_{ijt}^{p} = \alpha_{j}^{p} + \xi_{ij}^{p} + \varepsilon_{ijt}^{p} \tag{8}$$

where α_{j}^{p} represents the mean brand intercept that will be estimated as parameters. This parameter measures the brand strength in converting website visits to purchases, which is different from α_{j}^{f} in the first stage which measures the strength of a brand in attracting first visits. ξ_{ij}^{p} represents the individual-specific but time-invariant random effect for brands, and ε_{ijt}^{p} is the individual-and-time-specific idiosyncratic shock that we assume to be i.i.d. type one extreme value distribution. As before, to ensure proper identification we set the intercept for all airlines to zero. Hence the parameters corresponding to all covariates have to be interpreted as the relative difference with those consumers purchasing on airline websites. In the estimation model we also incorporate interactions of the demographic variables with all other covariates.

We assume that $\lambda_{i}^{p} = \lambda^{p} + \eta_{i}^{p}$, where η_{i}^{p} is a time-invariant and individual-specific random variable which captures the consumer heterogeneity in response to expected expenditure

level. We allow ξ_{ij}^{p} and η_{i}^{p} to be correlated among themselves and also correlated with other random effects in the earlier stages.

Let T_i be the total number of household i's purchases observed in data. Under the distribution assumption for ε_{ijt}^{p} and conditional on random effects ξ_{ij}^{p} and η_{i}^{p}, we can write down the probability of i's purchase history as the following traditional multinomial logit probability function:

$$\Pr(i\text{'s history of purchases}) = \prod_{t=1}^{T_i} \Pr(U_{ijt}^{*} \geq U_{ikt}^{*}, \forall k)$$

$$= \prod_{t=1}^{T_i} \frac{e^{\alpha_j^p + \xi_{ij}^p + \beta^p Z_{it} + \tau^p H_{ijt} + \gamma^p S_{ijt} + (\lambda^p + \eta_i^p) P_{it} + \rho^p I_{ijt} + \chi^{p\prime} F_{it} + \delta^p D_{it}}}{1 + \sum_{k=1}^{J} e^{\alpha_k^p + \xi_{ik}^p + \beta^p Z_{it} + \tau^p H_{ikt} + \gamma^p S_{ikt} + (\lambda^p + \eta_i^p) P_{it} + \rho^p I_{ikt} + \chi^{p\prime} F_{it} + \delta^p D_{it}}} \tag{9}$$

4.4 Model Estimation

Conditional on random effects in the three stages, we have a non-linear simultaneous equation system of (3), (6) and (9), where as the latter two equations involve endogenous variables F_{ijt} and D_{it} from the earlier stages. The major difficulty in model estimation comes from the fact that the random effects in each equation are likely to be correlated with each other within the equation and across equations. For example, a household with a higher ξ_{ij}^{f} in first visiting website j may also exhibit a higher ξ_{ij}^{p} in finally purchasing from j. Similarly, a household with a larger η_{i}^{d} for expected expenditure level in the browsing duration equation may also have a larger η_{i}^{p} in the purchase site decision. To solve this problem we use simulated maximum likelihood method to estimate this simultaneous equation system.

Let $\Psi_i = \{\xi_{ij}^{f}, \xi_{i}^{d}, \xi_{ij}^{p}; \forall j;\ \eta_{i}^{f}, \eta_{i}^{d}, \eta_{i}^{p}\}$ be the vector of random effects in the simultaneous equation system with the assumed distribution $F(\Psi;\ \Omega)$, where Ω is the variance-covariance matrix to be estimated. Equations (3), (6) and (9) are conditional on Ψ_i. These can be expressed under an integrated framework and transformed into the unconditional likelihood as follows:

$$L_i = \int \left(\begin{array}{l} \underbrace{\Pr(i\text{'s history of first visits}|\Psi_i)}_{\text{equation (3)}} \cdot \\ \underbrace{\Pr(i\text{'s history of browsing durations}|\Psi_i)}_{\text{equation (6)}} \cdot \\ \underbrace{\Pr(i\text{'s history of purchases}|\Psi_i)}_{\text{equation (9)}} \end{array} \right) dF(\Psi_i;\Omega) \tag{10}$$

We estimate this likelihood using the simulated maximum likelihood method. We draw Ψ_i^s, *s=1, …,ns,* where *ns* is the number of simulated draws, following the distribution of *F* (which we will explain later). The corresponding simulated version of (10) can be expressed as

$$\hat{L}_i = \frac{1}{ns}\sum_{s=1}^{ns} \left(\begin{array}{l} \underbrace{\Pr(i\text{'s history of first visits}|\Psi_i^s)}_{\text{equation (3)}} \cdot \\ \underbrace{\Pr(i\text{'s history of browsing durations}|\Psi_i^s)}_{\text{equation (6)}} \cdot \\ \underbrace{\Pr(i\text{'s history of purchases}|\Psi_i^s)}_{\text{equation (9)}} \end{array} \right) \tag{11}$$

We assume that Ψ_i is normally distributed as $N(0;\Omega)$, where Ω is the variance-covariance matrix. As discussed above each element of the Ω matrix explicitly accounts for the covariance of the random effects within and across different stages for brand strength (ξ_{ij}'s) and expected level of expenditure (η_i's). We make some simplifying assumptions on the Ω matrix to overcome computational burden and avoid over-parameterization issues. We assume the random effects for brand strength (ξ_{ij}'s) to be independent of the random effects for expected level of expenditure (η_i's) both within and across stages. However, ξ_{ij}'s and η_i's are allowed to be correlated among themselves within each stage as well as across stages. Hence a household with a higher ξ_{ij}^f on the first visited website j may also have a higher ξ_{ij}^p in purchasing from j. For simplicity we assume the covariance of these effects σ_{fp}^2 to be the same across all websites. We also assume same variance for the random effect ξ_{ij}^p at the purchase site decision model across all websites that is denoted by σ_{p}^2. The covariance of η_i^d for expected high level of expenditure in the browsing duration equation and η_i^p in the purchase site decision is captured

by σ^2_{hdp}. For further simplification we assume that the covariance between random effects for different levels of expected level of expenditure is zero. Similar interpretations could be made for other elements of the covariance matrix Ω. Its full structure is as provided below:

$$\Omega = \begin{pmatrix} \begin{bmatrix} \underset{(\text{jxj})}{\sigma^2_f I_j} & \underset{(\text{jx1})}{\sigma^2_{f_j d}} & \underset{(\text{jxj})}{\sigma^2_{fp} I_j} \\ \underset{(\text{1xj})}{\ddot{0}'} & \underset{(\text{1x1})}{\sigma^2_d} & \underset{(\text{1xj})}{\sigma^2_{p_j d}} \\ \underset{(\text{jxj})}{\ddot{0}'} & \underset{(\text{jx1})}{\ddot{0}'} & \underset{(\text{jxj})}{\sigma^2_p I_j} \end{bmatrix} & \underset{((\text{j+1+j})\text{x6})}{zeros} \\ \underset{(\text{6x(j+1+j)})}{zeros} & \begin{bmatrix} \underset{(\text{2x2})}{\begin{pmatrix} \sigma^2_{\text{mf}} & 0 \\ 0 & \sigma^2_{\text{hf}} \end{pmatrix}} & \underset{(\text{2x2})}{\begin{pmatrix} \sigma^2_{\text{mfd}} & 0 \\ 0 & \sigma^2_{\text{hfd}} \end{pmatrix}} & \underset{(\text{2x2})}{\begin{pmatrix} \sigma^2_{\text{mfp}} & 0 \\ 0 & \sigma^2_{\text{hfp}} \end{pmatrix}} \\ \underset{(\text{2x2})}{\ddot{0}'} & \underset{(\text{2x2})}{\begin{pmatrix} \sigma^2_{\text{md}} & 0 \\ 0 & \sigma^2_{\text{hd}} \end{pmatrix}} & \underset{(\text{2x2})}{\begin{pmatrix} \sigma^2_{\text{mdp}} & 0 \\ 0 & \sigma^2_{\text{hdp}} \end{pmatrix}} \\ \underset{(\text{2x2})}{\ddot{0}'} & \underset{(\text{2x2})}{\ddot{0}'} & \underset{(\text{2x2})}{\begin{pmatrix} \sigma^2_{\text{mp}} & 0 \\ 0 & \sigma^2_{\text{hp}} \end{pmatrix}} \end{bmatrix} \end{pmatrix},$$

where I_j is an identity matrix of dimension J (total number of websites). Subscripts "f", "d" and "p" denote the first visited site, browsing duration and purchasing site decisions, correspondingly. Subscripts "h" and "m" denote high and medium level of expected expenditure, correspondingly. In model estimation we restrict Ω to be a positive definite matrix. The number of simulated draws used to calculate the simulated likelihood was 100. The Nelder-Mead simplex algorithm we use is very efficient in estimating these complex models though some sensitivity to starting values was observed.

5. Results and Discussion

Initial investigation revealed two main effects – that of first site visited and browsing duration – on choice of final purchase site. In addition to these effects we also observe significant effects of prior experience, prior purchase, expected level of expenditure, brand strength, demographics and effects of some of the interaction terms. We first discuss the effect of

first site visited and browsing duration on choice of purchase site (as listed in Table 7) and then focus on the significant effects of covariates on the three stages.

Table 7: Effect of brand strength, prior site specific experience, first site visited and browsing duration on purchase website

	Parameters	Stage 1 First site visited		Stage 2 Browsing duration		Stage 3 Purchase site	
Intercept	Expedia	-1.437	(0.020)*	2.776	(0.002)*	-3.806	(0.019)*
	Orbitz	-1.624	(0.030)*			-4.239	(0.167)*
	Hotwire	-2.673	(0.021)*			-4.168	(0.056)*
	Other travel portals	-2.033	(0.006)*			-4.283	(0.064)*
Category prior experience	Expedia	0.008	(0.001)*	N/A		0.008	(0.001)*
	Orbitz	0.007	(0.001)*			-0.001	(0.001)*
	Hotwire	0.007	(0.001)*			0.008	(0.001)*
	Other travel portals	-0.072	(0.002)*			-0.121	(0.002)*
Stage 1 first visited site	Expedia	N/A		0.686	(0.001)*	5.223	(0.113)*
	Orbitz			0.577	(0.001)*		
	Hotwire			0.771	(0.001)*		
	Other travel portals			0.374	(0.001)*		
Stage 2 browsing duration	Expedia	N/A		N/A		0.281	(0.005)*
	Orbitz					0.444	(0.009)*
	Hotwire					0.293	(0.002)*
	Other travel portals					0.559	(0.007)*

* indicates p < .001. Standard errors are in parentheses

5.1 Effect of First Site Visited and Browsing Duration

We find a significant effect of the first site visited on the browsing duration (see the third column in Table 7, labeled "Stage 2"). On average consumers visiting Hotwire first tend to search more (coefficient 0.77) followed by Expedia and Orbitz. As the effect of first visiting airline websites on search duration is normalized to zero (see column two in Table 7, labeled "Stage 1"), these results imply that consumers will on average browse longer before final purchases if they start with travel portals, indicating a systematic difference in browsing behavior between these two consumer types. We also find that the first site visited has significantly large impact on the propensity to finally purchase from the same website (see column four in Table 7, labeled "Stage 3," with a significant coefficient of 5.22). This provides evidence that there is significant loyalty among consumers as they tend to buy from the website they first visited though they search other websites. Of course, this loyalty may also arise from prior purchases at

this site and persists through the present purchase process. Finally, browsing longer on average leads to a higher probability of purchasing from travel portals, and this effect is more pronounced for Orbitz and other travel portals.

5.2 Effect of Prior Experience

We measure category prior experience as the (log of) total number of pages viewed on the travel websites (including travel portals and airline websites) in the first three months of our data.. There is a positive effect of category prior experience on the choice of travel portals. These effects seem to affect all travel portals equally except other travel portals (see column two in Table 7) in predicting the first site visited. The effect of category prior experience on the choice of travel portals for final purchase is mixed (see column four in Table 7). Though category prior experience leads to a higher chance of finally buying from Expedia and Hotwire, the effects on Orbitz and other travel portals are negative. We suspect these results are related to the pricing policies of these websites during our sample period. Had we had the price data from these travel portals we would be able to provide an explanation.

We measure site-specific prior experience as the ratio of pages viewed on a site to the total pages viewed on all travel websites in the first three months of our data. This, in effect, is a measure of the share of a specific site in the total browsing done by the household on travel sites. Site specific category experience seems to positively affect propensity to first visit a website as well as purchase from a website (see columns two and four of Table 8).

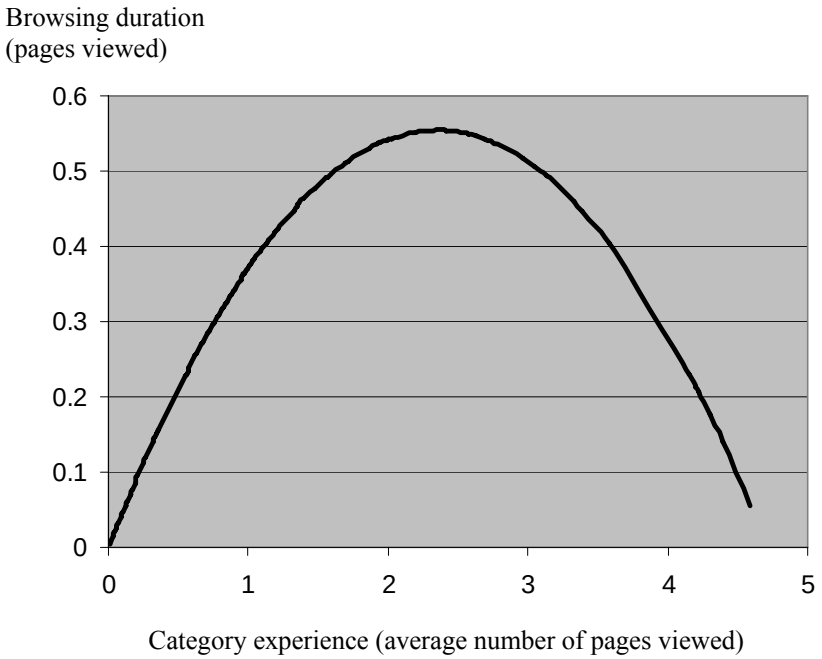
Table 8: Effect of prior experience (category), prior purchase and demographics

Parameters	Stage 1 First site visited		Stage 2 Browsing duration		Stage 3 Purchase site	
Site specific prior experience	1.134	(0.021)*	N/A		1.957	(0.024)*
Category prior experience	N/A		0.426	(0.002)*	N/A	
Squared category prior experience			-0.095	(0.001)*		
Prior purchase (one)	1.824	(0.045)*	0.017	(0.001)*	0.154	(0.002)*
Prior purchase (more than one)	N/A		-0.244	(0.003)*	N/A	
Age	-0.005	(0.001)*	0.067	(0.001)*	-0.058	(0.002)*
Income	-0.074	(0.001)*	-0.054	(0.001)*	0.351	(0.010)*
Broadband users	0.116	(0.002)*	-0.017	(0.001)*	-0.788	(0.014)*
Medium level of expenditure	0.708	(0.014)*	0.465	(0.002)*	2.345	(0.043)*
High level of expenditure	0.689	(0.016)*	0.794	(0.001)*	0.457	(0.008)*

* indicates $p < .001$. Standard errors are in parentheses

As discussed before, prior research has suggested that there may be a non-monotonic relationship between category prior experience and browsing duration. We therefore include a squared term of prior experience in the duration model (see column three in Table 8). The coefficient for the linear term is significantly positive but that for the squared term is negative. This finding is similar to that exhibited in offline search behavior where the relationship between consumers' prior knowledge and amount of information search has an inverted U shape, as illustrated in Figure 2, with moderate knowledge being associated with most search (Bettman and Park 1980, Hempel 1969, Johnson and Russo 1984).

Figure 2: Inverted-U relationship between category experience and browsing duration



While online search effects have not previously been investigated in this detail there is evidence in prior research that search, though minimal, is more pronounced with heavy users searching less (Johnson et. al. 2004).) than light users. One possible explanation for this is that product class knowledge increases search efficiency (Brucks 1985, Srinivasan and Ratchford 1991)), and is consistent with the non-monotonic relationship observed in our data.

5.3 Effect of Prior Purchase

The indicator to capture effects of prior purchase takes the value 1 if a website is where the last purchase had occurred. We find this effect to be significant and positive on first visited site decision (see column two in Table 8), indicating there is a very high probability that the website where last purchase occurred is invariably the first website to be visited before making a purchase. This inertia effect is also significantly positive on final purchase site decisions (see column four in Table 8), indicating that the likelihood of current purchase increases for a website if the last purchase occurred on it.

In the browsing duration model we use two indicators to for light and heavy users (as relative to those consumers without any purchase history in the first three months). Hence these variables indicate consumers' category purchase history, as compared to site-specific purchase history in the first site visited model. We find that the coefficient corresponding to light users is significantly positive but that for heavy users is significantly negative (see column three in Table 8), again indicating an inverted U relationship between prior purchase experience and browsing duration, consistent with the effect of prior category-specific experience discussed previously.

5.4 Effect of Expected Level of Expenditure

A higher level of expected expenditure (compared to low level) leads to a higher probability of first visiting travel portals (see column two in Table 8), perhaps indicating that consumers who purchase expensive tickets expect a larger benefit of getting a deal from travel portals and hence start browsing there. Similarly, higher expected expenditure also leads to a longer browsing duration (see column three in Table 8). These results are consistent with the cost-benefit evaluation story. Interestingly, while higher expected expenditure (compared to low level) leads to a higher probability of purchasing from travel portals (see column four in Table 8), it is consumers with medium level of expected expenditure who are most likely to purchase from travel portals. This may be due to the difference in pricing policies of airlines and travel websites (e.g., travel portals may offer better discount rate for medium priced tickets). Without data on price during the sample period we are not able to bolster this supposition with empirical evidence.

5.5 Effect of Brand Strength

The magnitudes of the site intercepts do reflect their relative importance of being chosen as the first visited site prior to a purchase. We find that the intercepts for Expedia and Orbitz are significantly larger than those for Hotwire and other travel portals (see column two in Table 7), indicating the difference in attractiveness of these websites for first visits. However, when it comes to final purchase decisions (see column four in Table 7), the intercept for Hotwire is larger though not significantly different from that for other travel portals, and that for Expedia is still the highest. These results show that different major travel portals have different attractiveness for first visits and final purchase decisions.

5.6 Effect of Consumer Demographics

Using demographics has been the traditional way of segmenting online consumers. We find that demographics like age, income, and connection speed do help to explain the browsing and purchase behavior of consumers in our model (see Table 8). For example, we find older consumers have a lower likelihood of first visiting or purchasing on a travel portal but tend to search more. In addition, high income consumers typically have lower likelihood of visiting a travel portal first, but have a higher likelihood of making a purchase on a travel portal as opposed to an airline website. Also those with higher income are less likely to search for longer duration perhaps due to higher opportunity cost for time.

We find that broadband users search less and have lower purchase probabilities on a travel portal (compared to airline sites). This is contrary to some existing evidence: for instance, Yonish, Delhagen & , and Gordon (2002) find that broadband users search 33% more compared to narrowband users due to the faster surfing speeds. A possible explanation is that broadband users are also high income consumers (broadband was relatively more expensive during our study period) and hence less price sensitive than narrowband users.

5.7 Effect of Interactions and Model Fit

We also estimate the interactions of demographics on covariates in all three stages. We discuss here those of managerial relevance.[10] We find the interaction between the prior

[10] We do not to report the full set of results in order to conserve space. The comprehensive set of results is available from the authors on request.

experience and the age of the consumer to be significantly positive indicating the effect of prior experience is stronger for older consumers as compared to young consumers. This is consistent with the expectation that older consumers are more reluctant to process new information as compared to younger consumers. The interaction between prior experience and the income of the consumer is significantly negative in the first site visited model implying that experienced high income customers tend to seek newer options rather than first visiting sites where they have most experience with. This is intuitive because as the income levels increase consumers do wish to explore more and take more risks.

There is also a significant interaction between connection speed and site specific prior experience in the first site visited model (Stage 1) indicating broadband users tend to visit first the site they most often visited in the past. There is a significant interaction between connection speed and prior purchase in the browsing duration model (Stage 2) indicating heavy users with higher connection speeds tend to browse more while light users would search less. Finally, the interaction between connection speed and prior purchase being on the same website is significant and positive in the purchase model (Stage 3) indicating broadband users tend to have a higher probability of purchasing from the website where they made their last purchase.

Table 9: Actual and predicted shares and browsing duration

Stage	Site	Actual	Predicted
1. First site visited shares	Expedia	0.1837	0.1843
	Orbitz	0.1402	0.1455
	Hotwire	0.0435	0.0466
	Other travel portals	0.0589	0.0597
	All airline websites	0.5737	0.5640
2. Browsing duration	N/A (Average number of pages viewed)	3.5330	3.5560
3. Purchase site shares	Expedia	0.1791	0.1797
	Orbitz	0.1372	0.1387
	Hotwire	0.0445	0.0462
	Other travel portals	0.0619	0.0612
	All airline websites	0.5773	0.5742

In order to evaluate how good the model fits with data we compute the expected share of first site visited and purchases for different websites as well as browsing duration based on the

estimation results, and compared to the actual shares and duration in data. Table 9 summarizes the results. Our three stage model clearly performs very well in terms of predictive power and explaining the data. The hit rate of first visited site and purchase site is at 96.4%. As comparison we estimate another model where each decision stage is estimated separately and find that the hit rate of this model is only 56.4%. This demonstrates how our model can be used to improve prediction efficiency.

5.8 Site and Price Heterogeneity

We do find significant heterogeneity in the behavior of consumers (the heterogeneity parameter estimates are summarized in Table 10). To conserve space we restrict our discussion to two interesting insights. Overall there seems to be significant price heterogeneity and very little site heterogeneity among consumers. The heterogeneity on the choice of purchase site (σ_p) is very large, implying that websites are viewed differently among different consumers. Also there is greater consumer price heterogeneity in the mid market as opposed to the high end of the market. The remaining covariances are small in magnitude and can be interpreted based on our discussion in section 4.4.

Table 10: Site and price heterogeneity covariance parameters

Site heterogeneity			Price heterogeneity		
Parameter	Estimate		Parameter	Estimate	
σ_f	0.002	(0.001)***	σ_{mf}	0.321	(0.004)***
σ_{fd1}	0.001	(0.001)**	σ_{hf}	0.017	(0.001)***
σ_{fd2}	0.001	(0.001)*	σ_{mfd}	0.198	(0.001)***
σ_{fd3}	0.001	(0.001)***	σ_{hfd}	0.002	(0.001)***
σ_{fd4}	-0.001	(0.001)***	σ_{md}	0.250	(0.001)***
σ_d	0.001	(0.001)***	σ_{hd}	0.001	(0.001)***
σ_{fp}	0.001	(0.001)***	σ_{mfp}	0.001	(0.001)
σ_{pd1}	-0.001	(0.001)***	σ_{hfp}	0.047	(0.002)***
σ_{pd2}	0.001	(0.001)***	σ_{mdp}	-0.001	(0.001)
σ_{pd3}	0.008	(0.001)***	σ_{hdp}	0.007	(0.001)***
σ_{pd4}	0.001	(0.001)***	σ_{mp}	0.630	(0.011)***
σ_p	0.763	(0.001)***	σ_{hp}	0.272	(0.005)***

*** indicates p < .001, ** indicates p < .005, * indicates p < .01. Standard errors are in parentheses

6. Conclusions

In this paper we develop a three-stage model to study the consumer online browsing and purchasing behaviors in the travel category. We model (i) the choice of the first website visited, (ii) the browsing duration of consumers on travel websites before making a purchase, (iii) the choice of the website where consumers will make the purchase, and how a latter stage choice is affected by decisions in the previous stages. We find significant effects of expected level of expenditure, prior browsing experience and prior purchase and brand strength in determining consumer browsing and purchase behavior. We also find that the choice of the first site visited and browsing duration has a significant impact on choice of the purchase site indicating the importance of modeling simultaneously.

Managers can use these results to identify the major determinants of consumer browsing and online purchase behavior. The findings from the browsing duration models (Stage 2) suggest that consumers are not penny wise and pound foolish i.e. consumers spend more time searching for prices when they expect a higher level of expenditure. These consumers are also more likely to start their browsing by first visiting and finally purchasing from travel portal websites. We also find an inverted-U relationship between prior experience and browsing duration.

We find strong state dependence in the browsing and purchasing behaviors such that prior purchase from a website increases the probability that the consumer will first visit that website (inertia) and will finally purchase from the same website. This is consistent with the learning or switching cost explanation, which suggests that consumers do not easily switch to competitors once they have transacted with a specific site. Moreover, we also find that first visited site choice strongly affects the probability that a consumer will finally purchase from the same site. The above results suggest a significant long term benefit for a website once it can attract consumers to visit the website first and especially if it can convert the visit to final purchase through various types of marketing and promotional activities. Our results are also useful for current major travel portals to understand their brand equity in terms of attracting consumers to first visit versus converting them to finally purchase.

With the above important findings, we also acknowledge some limitations in the current research. The major data limitation is that we do not observe what information consumers obtained while browsing. Specifically, we only observe from data the final transaction price but not prices from other competing websites. Hence we are neither able to say much about the price

effect on final purchase decisions nor how consumers search for price information online. Moreover, we do not have detailed transaction information such as the date and places of the flight. As a result, we cannot study some potentially interesting phenomena such as the difference in browsing and purchasing decisions between “last minute” and “planned” purchases.

Related to the above discussion, an interesting avenue of future research will be to collect data not only on consumers’ browsing path but also the information they obtained during the search. Also, in the current data set, combination (or basket) of purchases air, hotel, and car rental need to be explored further in order to understand how consumers approach buying multiple products at the same time from multiple or same website. It is important to understand how by providing a basket of complementary products which involve air tickets purchase, car rental and/or hotel bookings travel portals such as Expedia and Orbitz can better satisfy consumer needs and hence successfully compete with airlines or hotels websites which sell products separately. There also exists an opportunity to incorporate dynamic visit behavior in modeling the browsing duration stage by exploring the sequence of sites visited and the impact it has on purchase. Another important extension would be to develop multistage models that help distinguish buyers from browsers in a more detailed and dynamic manner.

References

Anderson N.H., (1965), Primacy Effects in Personality Impression Formation Using A Generalized Order Effect Paradigm, *Journal of Personality and Social Psychology*, 34,July,1-9

Beatty, Sharon E. and Scott M. Smith, (1987), External Search Effort: An Investigation Across Product Categories, *Journal of Consumer Research*, 14,(1) , June,83-95

Bellinger, D.N., Robertson D.H. and Hirschman E.C., (1978), Impulse Buying Varies By Product, *Journal of Advertising research*, 18,December,15-18

Bettman, James R. and C. Whan Park, (1980), Effects Of Prior Knowledge And Experience And Phase Of The Choice Process On Consumer Decision Processes: A Protocol Analysis, *Journal of Consumer Research*, 7,(3) December,234-248

Bhatnagar A. and Ghose, (2004), A Latent Class Segmentation Analysis Of E-Shoppers, *Journal of Business Research*, 57,758-767

Bouvard,P. and Kurtzmann, W., (2002), The Broadband Revolution: How Superfast Internet Access Changes Media Habits In American Households, *[Online] Arbitron/Coleman Report. Available www.arbitron.com/downloads/broadband.pdf*, December 20

Brucks, Merrie, (1985), The Effects Of Product Class Knowledge On Information Search Behavior, *Journal of Consumer Research*, 12,(1), June,1-16

Brynjolfsson, Erik and Michael D. Smith, (2001), The Great Equalizer? Consumer Choice Behavior At Internet Shopbots, *Sloan Working Paper*, 4208-01, October

Bucklin, Randolph E. and Catarina Sismeiro, (2003), A Model of Web Site Browsing Behavior Estimated on Clickstream Data, *Journal of Marketing Research*, August,249-267

Chiang, Jeongwen , Siddhartha Chib and Chakravarthi Narasimhan, (1999), Markov Chain Monte Carlo And Models Of Consideration Set And Parameter Heterogeneity, *Journal of Econometrics*, 89,March/April,223-248

Degeratu, Alexandru M., Arvind Rangaswamy and Jianan Wu, (2000), Consumer Choice Behavior In Online And Traditional Supermarkets:The Effects Of Brand Name, Price, And Other Search Attributes, *International Journal of Research in Marketing*, 17,55-78

Disckson and Sawyer, (1990), The price knowledge and search of supermarket shoppers, *Journal of Marketing*, 54,(3) July,42-53

Economist, (2005), Flying From The Computer., *Economist.com*, September 29th

Gourville, T John and Dilip Soman, (2005), Overchoice And Assortment Type: When And Why Variety Backfires, *Marketing Science*, 24,(3),382-396

Gudagni, Peter M. and Little, John D.C., (1983), A Logit Model Of Brand Choice Calibrated On Scanner Data, *Marketing Science*, 2,(3) Summer,203-238

Hauser John R. and Birger Wernerfelt, (1990), An Evaluation Cost Model Of Consideration Sets, *Journal of Consumer Research*, 16,March,393–408

Hempel D.J., (1969), Search Behavior And Information Utilization In The Home Buying Process, *Marketing Involvement in Society and the Economy*, 241-249

Janiszewski C., (1998), The Influence Of Display Characteristics On Visual Exploratory Search Behavior, *Journal of Consumer Research*, 25,3,290-301

Jarboe, G.R. and McDaniel C.D., (1987), A Profile Of Browsers In Regional Shopping Malls, *Journal of the Academy of Marketing Science*, 15,Spring,46-53

Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman and Gerald L.Lohse, (2004), On The Depth And Dynamics Of Online Search Behavior, *Management Science*, 50,(3) March,299-308

Khan B.E., (1995), Consumer Variety-seeking Among Goods And Services, *Journal of Retailing and Consumer Services*, 2,2,139-148

Lind EA, Kray L and Thompson L., (2001), Primacy Effects In Justice Judgments: Testing Predictions From Fairness Heuristic Theory, *Organ Behav Hum Decis Process*, 85,(2) July,189-210

Moe, Wendy W. and Peter S. Fader, (2004a), Dynamic Conversion Behavior At E-commerce Sites, *Management Science*, 50,(3) March,326-355

Moe, Wendy W. and Peter S. Fader, (2004b), Capturing Evolving Visit Behavior In Clickstream Data, *Journal of Interactive Marketing*, 18,(I), Winter,5-19

Montgomery, Alan L., Shibo Li, Kannan Srinivasan and John C Liechty, (2006), Modeling Online Browsing and Path Analysis Using Clickstream Data, *Marketing Science forthcoming*,

Moorthy, Sridhar, Brian T. Ratchford and Debabarata Talukdar, (1997), Consumer Information Search Revisted: Theory And Empirical Analysis, *Journal of Consumer Research*, 23,March,263-277

Newman, John W., (1977), Consumer External Search: Amount And Determinants, *Consumer and Industrial Buying Behavior*, New York: Elsevier,ed. Arch Woodside et.al,79-84

Park, Young-Hoon and Peter S. Fader, (2004), Modeling Browsing Behavior At Multiple Websites, *Marketing Science*, 23,(3) Summer,280-303

Prelec, Drazen and Loewenstein, George, (1998), The Red And The Black: Mental Accounting Of Savings And Debt, *Marketing Science*,

Punj, Girish N. and Richard Staelin, (1983), A Model Of Consumer Information Search Behavior For New Automobiles, *Journal of Consumer Research*, 9,(4) March,366-380

Roberts, John H. and James M. Lattin., (1991), Development And Testing Of A Model Of Consideration Set Composition, *Journal of Marketing Research*, 28,November,429–440

Roy, A., (1994), Correlates Of Mall Visit Frequency, *Journal of Retailing*, 70,(2),139-161

Schaninger, Charles M. and Donald Sciglimpaglia, (1981), The Influence Of Cognitive Personality Traits And Demographics On Consumer Information Acquisition, *Journal of Consumer Research*, 8,(2) September,208-216

Seetharaman, P.B., Andrew Ainslie and Pradeep Chintagunta, (1999), Investigating Household State Dependence Effects Across Categories, *Journal of Marketing Research*, 36,(4),488-500

Shocker, Allan D., Moshe Ben-Akiva, Bruno Boccara and Prakash Nedungadi, (1991), Consideration Set Influences On Consumer Decision-Making And Choice: Issues, Models, And Suggestions, *Marketing Letters*, 2,(3),181–197

Sismeiro Catarina and Randolph E Bucklin, (2004), Modeling Purchase Behavior at an E-Commerce Web Site: A Task Completion Approach, *Journal of Marketing Research*, 41,(3) August,306-323

Srinivasan, Narasimhan and Brian T. Ratchford, (1991), An Empirical Test Of A Model Of External Search In Automobiles, *Journal of Consumer Research*, 18,(2) September,233-242

Staw, Barry M, (1976), Knee-deep In The Big Muddy: A Study Of Escalating Commitment To A Chosen Course Of Action, *Organizational Behavior and Human Performance*, 16,27-44

Wu, Jianan and Arvind Rangaswamy, (2003), A Fuzzy Set Model Of Search And Consideration With An Application To An Online Market, *Marketing Science*, 22,(3) Summer,411-434

Yonish,S., Delhagen,K. and Gordon,J., (2002), Where The Broadband Shoppers Are: Wholeview Technographics Research Report, *[Online] Forrester Research. Available www.forrester.com*, December 18

Zellermayer, Ofer, (1996), The Pain Of Paying, *PhD dissertation. Carnegie Mellon University*,