January 2010

Essays on Online Browsing and Purchase

Ciju Nair
Washington University in St. Louis

WASHINGTON UNIVERSITY

Olin School of Business

Dissertation Examination Committee:
Tat Chan, Chair
Cheema Amar
Narasimhan Chakravarthi
Sebastian Galiani
Seetharaman Seethu
Xie Yang

ESSAYS ON ONLINE BROWSING AND PURCHASE

By

Ciju T.R. Nair

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2010
Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Essays on Online Browsing and Purchase

by

Ciju T. R. Nair

Doctor of Philosophy in Business Administration

Washington University in St. Louis, 2010

Professor Tat Chan, Chairperson

Essay One: Modeling Online Browsing and Purchase of Airline Tickets

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, we use a joint framework to study consumers' online browsing and purchase of airline tickets in a unique dataset of household-level dynamic click stream panel data. We use a three-stage model to study (i) the choice of the first website visited, (ii) the duration of browsing on travel websites before making a purchase (iii) the choice of the website where consumers will make the purchase, and how a later stage choice is affected by decisions in previous stages. We simultaneously estimate these three models which constitute a non-linear discrete-continuous equation system using a simulation-based econometric technique. We find significant effects of expected level of expenditure, prior browsing experience, prior

purchase experience in determining consumer browsing and purchase behavior. We are able to quantify the differences in attractiveness of a website in getting consumers to first visit them and compare it with the conversion effectiveness of a website in terms of getting consumers who visit to make purchases. A significant impact of choice of the first site visited and browsing duration on choice of the purchase site indicates the importance of modeling these decisions simultaneously. Our results can help managers identify the major determinants of consumer browsing and online purchase behavior, some of which cannot be observed in a brick-and-mortar environment.

Essay Two: Modeling Online Multi-category Purchase in Travel

In this paper we investigate online purchase behavior at the basket level and model the multi-category purchases in the travel product category. While prior research in marketing has looked at browsing or individual category purchase decisions, we study consumers' online purchase of airline, car rental and hotel purchases together using a unique dataset of household-level dynamic click stream panel data. We use a two-stage model to study (i) the propensity of consumers to purchase a combination of products as a basket and (ii) the choice of the website where consumers will make those purchases. We then estimate the propensity of consumers to purchase a particular combination of products in their basket from different websites. This behavior constitutes a high dimensional system of multinomial equations which are then solved using a simulation-based econometric technique. We find significant effects of site preference, loyalty, prior browsing and demographic variables in determining consumer multi-category purchase

behavior. Our results can help managers identify the major determinants of multi-category purchase as well as provide insights into cross promoting as well as upselling other products to consumers who visit their website.

Key Words: airline, car rental, hotel, travel, multinomial choice, purchase, behavior, multi stage models, online, browsing

ACKNOWLEDGEMENTS

Though the following dissertation is an individual work, I could never have reached the heights or explored the depths without the help, support, guidance and efforts of a lot of people. Firstly, I would like to thank my advisors Dr. Tat Chan and Dr. Amar Cheema for instilling in me the qualities of conducting scientific exploration on large marketing datasets to further our understanding about marketing and translating insights in meaningful ways for practitioners. I would like to thank my committee members Dr. Chakravarthi Narasimhan, Dr. Ying Xie, Dr. Seethu Seetharaman and Dr. Sebastian Galiani for their guidance over the years. I would also like to thank faculty members past and present of Washington University for their help and support.

I would like to take this opportunity to thank Dean Mahendra Gupta and members of the faculty at Olin Business School for whom I worked over the years but with whom I forged a very special bond, for their encouragement and helpful advice. Thanks also go out to participants in the PhD seminar series at Washington University in St. Louis, doctoral candidates in the Olin PhD program and participants at INFORMS Marketing Science 2006 conference for their valuable suggestions. Special thanks to the Center for Research in Economics and Strategy (CRES) at the Olin School of Business for the grant to purchase the ComScore database used in this research.

Finally, I would like to thank my parents, T.K Radhakrishnan Nair and Dr. S. Rugmini Radhakrishnan, for their constant encouragement and dedication. Special thanks to my wife Dr. Anuradha Nair and son Krishnachandra Nair who provided continued motivation and stood by me despite financial hardships during this endeavor.

This dissertation is dedicated to

My

Parents, Wife and Son

TABLE OF CONTENTS

LIST OF TABLES

Essay-One

Essay-Two

LIST OF FIGURES

INTRODUCTION

Consumers' pre-purchase activation and path to conversion has intrigued academicians and marketers over the years. Of late the availability of clickstream data and new data sources that capture granular advertising data from traditional media (TV, Radio, Print, OOH etc.) and digital media assets (display, paid and natural search, text links and content networks) helps us explore the impact of marketing investments and quantify its impact on business performance and consumer decision making in more detail. The marketing funnel (see Figure A) is a key conceptual framework that is routinely used by practitioners to deconstruct the marketing activation and identify key issues.

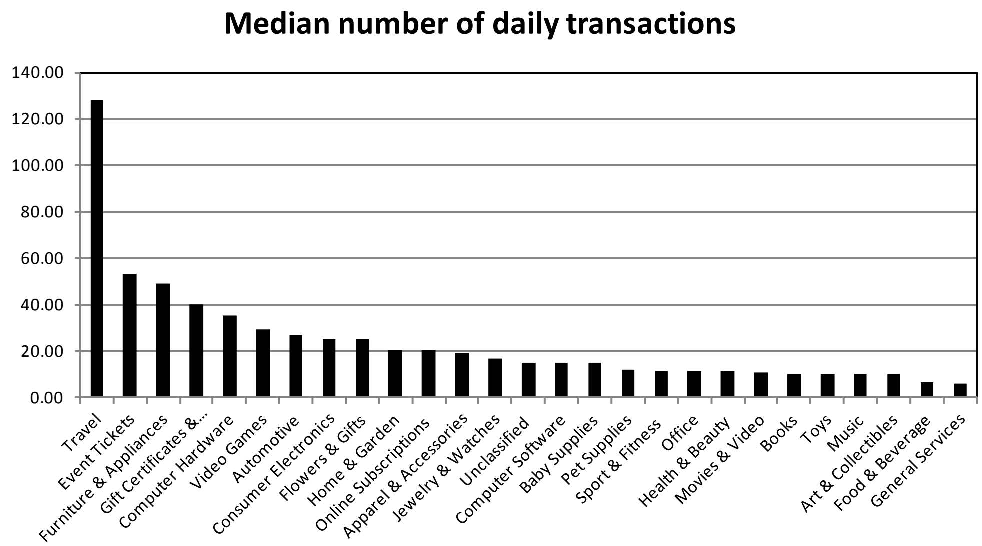
Figure A: Marketing Funnel

Capturing all these effects along the marketing funnel and explaining the impact of interactions across these various stages is an area that requires seminal work and in this dissertation the author takes a few first steps using clickstream data to explore the consumer browsing and purchase behavior in airline ticket purchases using a joint framework. The author also extends this work into a multi-category framework to explain purchases in the travel category (air, hotel and car rental). In the first essay we use a three stage model to study (i) the choice of the first website visited (ii) the duration of browsing on travel websites before making a purchase and (iii) the choice of the website where consumers will make the purchase, and how a later stage choice is affected by decisions in previous stages. In the second essay of the dissertation we investigate online purchase behavior at the basket level and model the multi-category purchases in the travel product category. Any analysis using single-category data provides only a partial view of consumer behavior that ignores possible dependencies between consumer purchase outcomes across multiple products in the basket. This leads to a biased understanding of consumer purchase decisions as it pertains to basket purchases and transactions. We use a two-stage model to study (i) the propensity of consumers to purchase a combination of travel products as a basket and (ii) the choice of the website where consumers will make those purchases.

Significant insights are obtained that help us understand the consumer search process as well as the impact of service provider brands compared to travel portals when it comes to making purchases in the travel category. The scope of this dissertation and the modeling efforts in this empirical research are limited by any limitations contained in the ComScore clickstream dataset from Wharton Research Data Services

Amongst the 27 categories that the ComScore data set is comprised of Travel with 128.28 transactions (see Figure B) constitutes the highest median number of daily transactions during the study period in any category.

Figure B: Median number of daily transactions



Travel also constitutes nearly 50% of total online spend for households who bought travel in our dataset. ComScore (2007) also estimates non-travel spending market in the US to be about $102bil in 2006 (online travel is a $70bn market in the US). Households also spend twice the time on travel websites where they can make a purchase compared to information only websites. For these reasons we investigate the travel category in more depth in this dissertation.

An analysis of the travel browsing and purchase data provides a few useful insights that are relevant to understanding this dataset. We notice 68% of households (Figure C) make one or two travel purchases during the six month period we study (Jul-Dec 2002). A closer look at the basket of travel products reveals single product purchases constitute 82% of all purchases with Air travel purchases being the most dominant at constituting almost half of all transactions (Figure D).

Figure C: Travel purchase transactions made by households

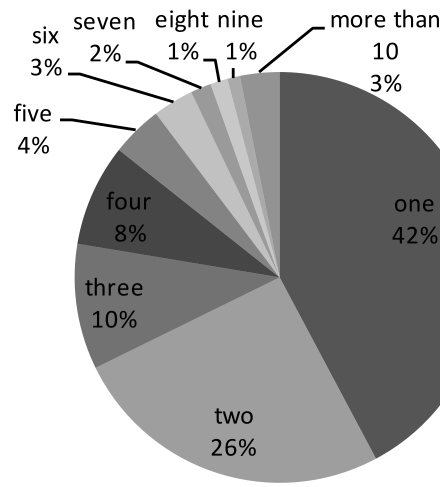
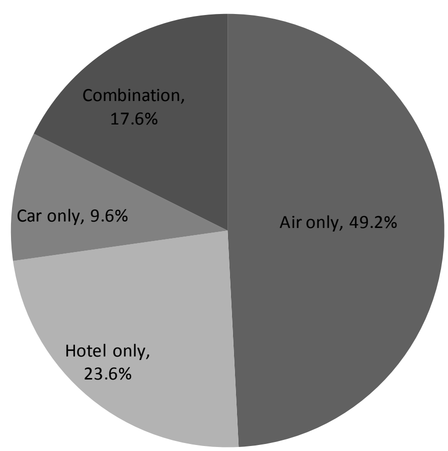
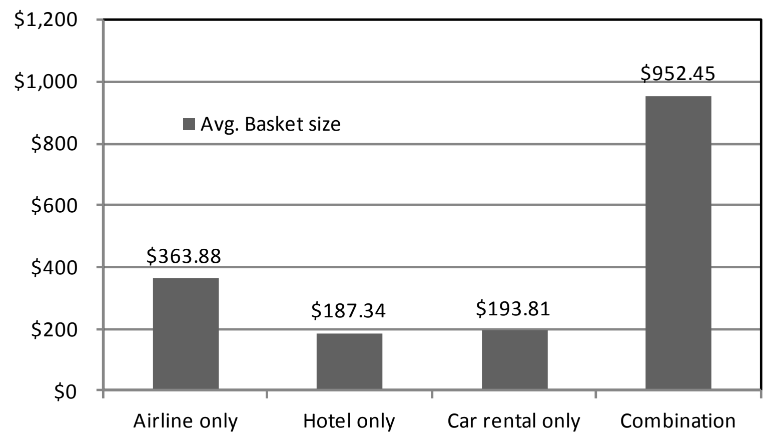


Figure D: Purchase shares by type of basket purchase



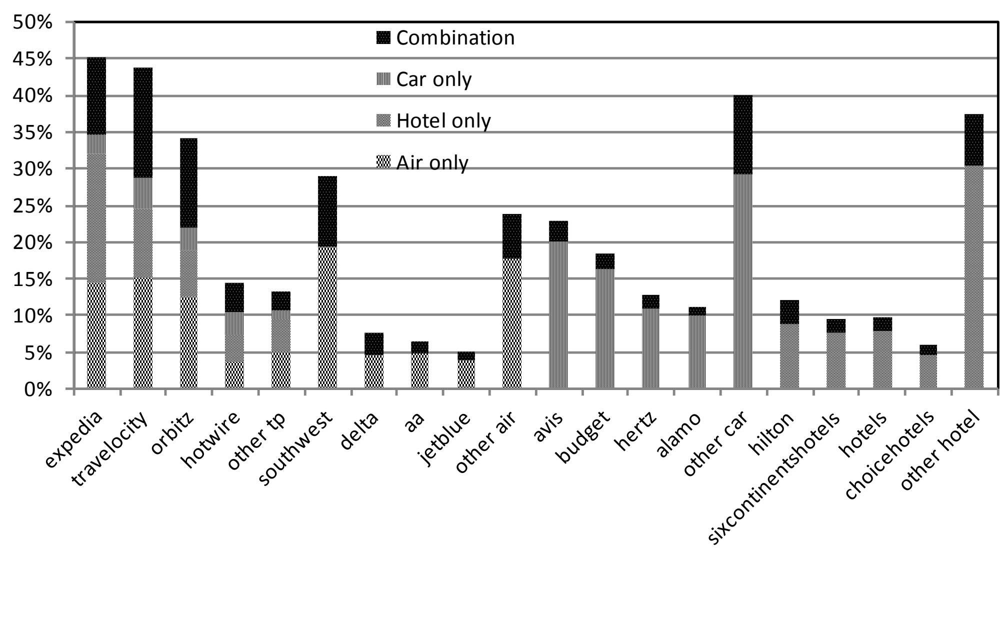
Though combination or basket purchases constitute only 18% of all travel transactions the average basket value is many times more compared to single product purchases for combination purchases (see Figure E).

Figure E: Average Basket value by type of travel purchase



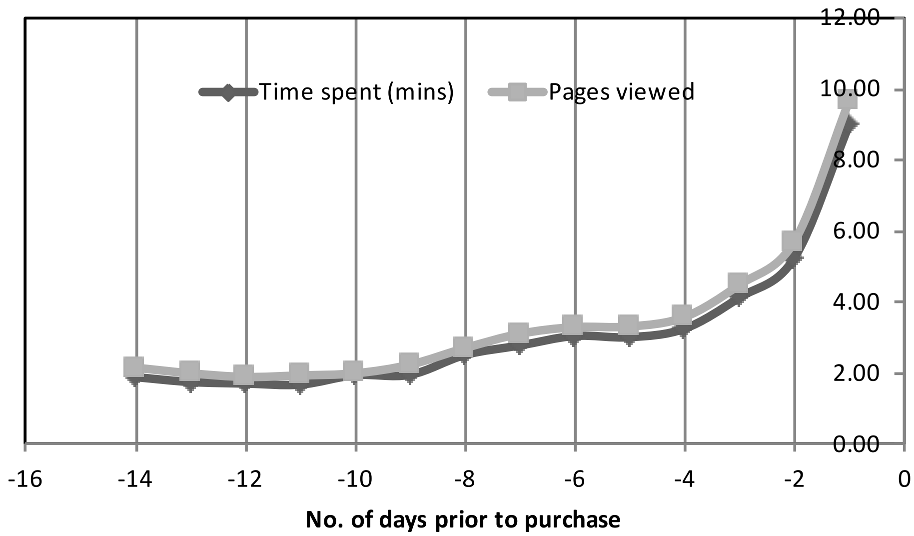
We also find a majority of combination purchases occur on travel portals (see Figure F) compared to airline or hotel or car rental websites.

Figure F: Purchase shares by major websites



We also find the market is more consolidated among the leading four travel portals and the market is very thinly spread across the other travel portals when compared to the big four players. However a large number of transactions occur on other air, hotel and car rental sites indicating that the market is not consolidated amongst the leading players. This could possibly be due to the market being fragmented across a large number of specialist affiliate travel sites that focus on single product sales.

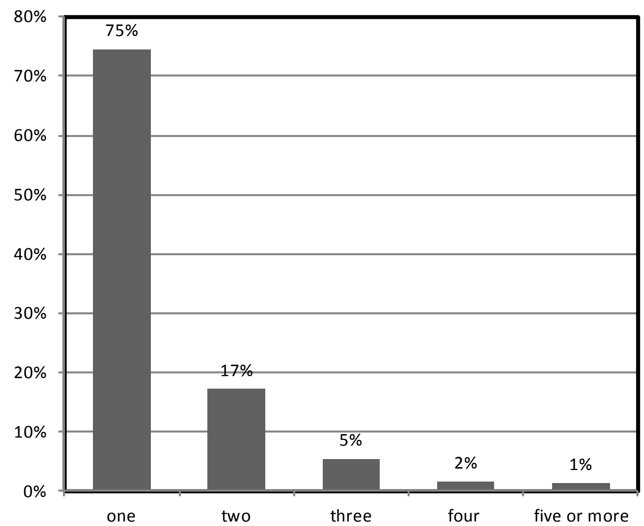
Figure G: Activity no. of days prior to purchase



An analysis of consumer browsing behavior in terms of time spent and pages viewed indicates that consumer browsing is lowest two weeks prior to search and exponentially increases as it gets closer to purchase (see Figure G). This information is managerially relevant and can be used in setting tracking windows for digital cookies. However, different levels of involvement and prolonged purchase cycles could impact this differently across product categories.

A majority of the households visited only one website in the two weeks prior to the transaction (see Figure H). However nearly one-fourth of the households visited two or more websites where they could make a purchase indicating very few websites constituted the household consideration set (max was 9 websites).

Figure H: Size of consideration sets



This concludes the synopsis of the data and the rest of this dissertation is organized into two essays. The first essay focuses on online browsing and purchase of airline tickets in the travel category while the second extends the framework to include browsing and purchase of basket transactions of air, hotel and car rental products.

Essay One:

Modeling Online Browsing and Purchase of Airline Tickets

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 the internet in purchases is also increasing. For instance, nearly a third of the $200 billion travel market purchases were made online by consumers in the US in 2005 (Economist 2005). Travel as a category has also grown significantly 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 sites visited prior to purchase, 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. 2004, 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, across multiple websites, explicitly estimating how the former impacts the latter.

Weitzman (1979) proposes a framework comprised of a selection and stopping rule when it comes to explaining optimal search behavior. The Selection Rule implies "If an option is to be pursued, it should be that unexplored option with highest reservation price" and the Stopping Rule states "Terminate search whenever the maximum sampled reward exceeds the reservation price of every unexplored option". Kim, Albuquerque and Bronnenberg (2009) extend this work by combining Weitzmans's rules with a choice rule that the consumer relies on to choose the maximum utility alternative. Our work differs from that of Kim, Albuquerque and Bronnenberg (2009) in that we do not impose a theoretical framework to model the search process but are more interested in explicitly incorporating the so called behavioral search/browsing metrics observed in data to predict purchase. Also though our work predicts browsing and purchase behavior across websites (travel portals and service provider websites) for a product the non-availability of travel destination information is a challenge when it comes to using prices to infer reservation price or price expectations of consumers when it comes to making choice decisions.

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 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 also focus on investigating the sub-processes affecting each of these three stages, more specifically, effects of expected level of expenditure, prior browsing experience, prior purchase experience and brand strength in addition to 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, 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. We use these results to inform us of differences in consumer behavior across different travel portals and also between travel portal and airline websites.

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. Section 6 concludes, with a discussion of some of the limitations of the present research, and outlines opportunities 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, contingent on visiting the store they will be more likely to buy. In contrast, online shoppers are less likely to buy after visiting an online store. 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). However, this difference in cost of search does not necessarily imply that consumers will have perfect price or product information on the internet because searching online also requires 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 on and compare across). It is therefore important to understand consumer browsing behavior in the online context, which is different from the offline context, and its impact on purchase. 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 a large body 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). This literature suggests that state dependence exists across households irrespective of demographics and diminishes over time. 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 differences in consideration set formation 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. Consistent with research on state dependence one may expect that past experience and browsing or purchase experience will dictate the formation of such consideration set. One of the most important indicators of the consideration set is where consumers start from, i.e., the first website that they visit. In America, 54% of consumers start with a travel portal, such as

Expedia, Travelocity or Orbitz, 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 and Sidestep (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. The first site visited may indicate that consumers have a stronger preference for it than for other sites they can buy from (Fazio et al. 2000). This first site may thus reveal consumers' preference before they encounter present information. Furthermore, information on the first site 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. Hence, we expect experienced consumers would first visit a travel portal as they are information aggregators and provide more information for the experienced consumer to process.

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 studies using automobile purchases have shown pre-purchase search to be minimal (Beatty and Smith 1987). This result 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 levels of category experience. 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. We expect this relationship to be non-linear and have an inverted U shape as found in previous research.

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 consumers who have no prior purchase experience.

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 stay with where they are likely to get good deals. 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

Websites with strong brand names are more likely to be visited first. We use brand intercepts to capture effects of brand strength on the choice of the first site visited. An alternative interpretation is that these intercepts imply different levels of unobserved 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. Other research also finds that older consumers typically have less patience to search (Ward and Lee 2000) hence we expect older consumers to first visit travel portals as they have a search friendly format. We also expect high income consumers to be less price sensitive compared to low income consumers, thus we believe high income consumers would indulge in less search than lower income consumers and will also be more likely to first visit an airline website than the latter. 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.

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 Chintagunta 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 those on 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. On average we expect the airline website brands to have higher brand strength as opposed to travel portals as they are service providers and typically invest more on brand building, use of reward programs and other marketing initiatives. Also stronger brands are expected to have higher conversion effectiveness as opposed to weaker brands.

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 in both online and traditional supermarkets 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. Consistent with the detrimental effect of excessive choice (e.g. Gourville and Soman 2005) older consumers, who typically have less cognitive resources to process information, may have a lower likelihood of purchasing from a travel portal than from airline sites. 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. We expect the effect of broadband to mirror that of high income consumers as only the high income consumers were able to afford broadband in our dataset.

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. It is likely that consumers who first visit a travel portal would exhibit a different browsing behavior as opposed to those who first visit an airline site. Consumers who first visit a travel portal can be expected to search less as travel portals being information aggregators provide a lot of information and options on the same page. On the contrary due to self selection consumers who typically indulge in more search might start by first visiting a travel portal. Hence apriori it is difficult to hypothesize as to how the search duration would be impacted depending on the first site visited. 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 on travel portals. 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 which captures 96% of all search that consumers indulge in); (2) we only study the household's browsing in the travel category (travel portals and airline websites) during

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

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 period. Table 1 reports some summary statistics of the data by household for the 1832 households during the six month period from July 2002 to December 2002. In this dataset, households on average make approximately two online purchases in the travel category during the six month period (the median number of purchases is two). Households have a mean spending of nearly $600 on a travel purchase and spend nearly three hours on average searching on websites selling travel products to make two purchases on average during the six month period.

Table 1: Summary statistics

1832 households (Jul 02 - Dec 02)	Mean	Std. Dev.
Time spent (minutes)	165.65	205.08
Number of pages viewed	165.28	212.73
Number of unique airline websites visited	9.66	9.20
Total number of airline websites visited	13.84	16.56
Purchases in airline category	1.68	1.29
Expenditure in airline category (US$)	570.09	661.55
Purchases in all categories	10.20	17.61
Expenditure in all categories (US$)	1018.22	2839.47

Travel category forms a significant portion of online purchases made by consumers with the mean being two out of ten 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 2002 – September 2002) 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 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.

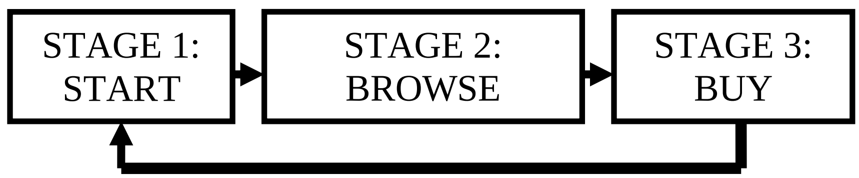
4. Model Specification and Estimation

---

[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).

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.

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 (see Table 3) browsing on travel portals on average is more than that on airline websites. However, it appears that there is not much difference between travel portals, which we investigate further using our model to see if post controlling for the various factors that affect each stage whether the first site visited impacts browsing duration. Moreover this indicates that it is important to incorporate choice of the first site, the decision variable in the first stage in the second stage of the model, to understand how it affects the browsing duration. For simplicity we only model 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. 2004).

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

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 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 site visited than from sites they subsequently visit.

Table 3: Comparison of browsing on travel portals and airline sites

First site visited	Average number of pages viewed
Expedia	69.46
Orbitz	69.31
Hotwire	72.51
Other travel portals	75.96
All airline websites	40.84

Table 4: Effect of first site visited 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	8.7	6.1	2.8	2.7	**36.5**	43.1	32.9
Grand total	15.0	11.9	4.8	3.4	26.3	38.6	100.0

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 site visited than from sites they subsequently visit.

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 treat all airline websites together as an outside option.

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 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).

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

[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 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.

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.

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

We assume that these are determined by a list of factors including customer demographics (age, income, connection speed) Z_{it}, prior category experience H_{it}, expected level of expenditure P_{it} and the site-specific prior browsing experience S_{ij}. Prior category experience H_{it} 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 browsing experience S_{ij} 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 ($>\$356$)

[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.

expenditures[5]. Furthermore, I_{ijt} 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 site visited 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 site visited 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

$$F_{ijt}^{*} = \alpha_{ijt}^{f} + \beta^{f} Z_{it} + \tau_{j}^{f} H_{it} + \gamma^{f} S_{ij} + \lambda_{i}^{f} P_{it} + \rho^{f} I_{ijt} \quad (1)$$

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

$$\alpha_{ijt}^{f} = \alpha_{j}^{f} + \xi_{ij}^{f} + \varepsilon_{ijt}^{f} \quad (2)$$

[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. 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.

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

To ensure proper identification we normalize the latent variable value for all airlines as $F_{i,AIR,t}^{*} = 0 + \varepsilon_{i,AIR,t}^f$.[8] In our estimation model we also incorporate the interactions of demographic variables Z_{it} with all other covariates H_{it}, S_{ij}, 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_{ijt}^f and

---

[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.

[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.

conditional on the random effects ξ_{ij}^{f} and η_i^{f}, we can write down the probability that a household's history of first visits in the whole sample period as below.

$$\begin{aligned}&\Pr(i\text{'s history of first visits}) = \prod_{t=1}^{T_i} \Pr(F_{ijt}^{*} \geq F_{ikt}^{*}, \forall k) \\ &= \prod_{t=1}^{T_i} \frac{e^{\alpha_j^f + \xi_{ij}^f + \beta^f Z_{it} + \tau_j^f H_{it} + \gamma^f S_{ij} + (\lambda^f + \eta_i^f) P_{it} + \rho^f I_{ijt}}}{1 + \sum_{k=1}^{J} e^{\alpha_k^f + \xi_{ik}^f + \beta^f Z_{it} + \tau_k^f H_{it} + \gamma^f S_{ik} + (\lambda^f + \eta_i^f) P_{it} + \rho^f I_{ikt}}}\end{aligned} \qquad (3)$$

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.

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 site visited 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 site visited 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 $F_{ijt} = 0 \text{ or } 1$ and $\sum_{\text{j}} F_{ijt} = 1$, 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:

$$D_{it} = \alpha_{it}^{d} + \beta^{d} Z_{it} + \lambda_{i}^{d} P_{it} + \phi^{d} I_{p=1} + \delta^{d} I_{p>1} + \lambda_{1}^{d} H_{it} + \lambda_{2}^{d} H_{it}^{2} + \chi_{j}^{d}{}' F_{ijt} \quad (4)$$

[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.

This 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, i.e., $\varepsilon_{it}^{d} \sim N(0,\sigma^{2})$. We assume that $\phi_{i}^{d} = \phi^{d} + \eta_{i}^{d}$, 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 site visited 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

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

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 purchase 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 browsing experience as the average daily pages viewed S_{ij} on different sites in the first three months, H_{it} 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 site visited and browsing duration models as covariates in this latent variable function. Therefore we can write down

$$U^{*}_{ijt} = \alpha^{p}_{ijt} + \beta^{p} Z_{it} + \tau^{p}_{j} H_{it} + \gamma^{p} S_{ij} + \lambda^{p}_{i} P_{it} + \rho^{p} I_{ijt} + \chi^{p}\,' F_{it} + \delta^{p} D_{it} \tag{7}$$

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

$$\alpha^{p}_{ijt} = \alpha^{p}_{j} + \xi^{p}_{ij} + \varepsilon^{p}_{ijt} \tag{8}$$

where α^{p}_{j} 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^{f}_{j} in the first stage which measures the strength of a brand in attracting first visits. ξ^{p}_{ij} represents the individual-specific but time-invariant random effect for brands, and ε^{p}_{ijt} 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:

$$\Pr(i\text{'s history of purchases}) = \prod_{t=1}^{T_i} \Pr(U_{ijt}^* \geq U_{ikt}^*, \forall k)$$
$$= \prod_{t=1}^{T_i} \frac{e^{\alpha_j^p + \xi_{ij}^p + \beta^p Z_{it} + \tau_j^p H_{it} + \gamma^p S_{ij} + (\lambda^p + \eta_i^p) P_{it} + \rho^p I_{ijt} + \chi^{p\,\prime} F_{it} + \delta^p D_{it}}}{1 + \sum_{k=1}^{J} e^{\alpha_k^p + \xi_{ik}^p + \beta^p Z_{it} + \tau_k^p H_{it} + \gamma^p S_{ik} + (\lambda^p + \eta_i^p) P_{it} + \rho^p I_{ikt} + \chi^{p\,\prime} F_{it} + \delta^p D_{it}}} \tag{9}$$

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:

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

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_{hdp}^2. 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:

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

where I_j is an identity matrix of dimension J (total number of websites). Subscripts "f", "d" and "p" denote the first site visited, 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

In this section we also share estimates from a model incorporating no-purchase, the results of which are very similar to that of a model that is conditional only on purchase that we discuss in our paper (see Table 7 for a comparison). 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 browsing 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 8) and then focus on the significant effects of covariates on the three stages.

5.1 Model with no-purchase browsing sessions

The motivation for investigating a model that includes no-purchase was primarily based on prior literature that demonstrated the biases that were associated with choice models that used scanner panel data without accounting for no-purchase option.

In our paper we estimate choice conditional on a purchase because unlike grocery store purchase behavior the online purchase behavior is cluttered with numerous visits to the website that don't necessarily translate into transactions. This is explained by the fact that consumers do not visit a retail or grocery store to do window shopping or seek information in general whereas they would visit an online travel portal or airline site to window shop and inform themselves of prices for a long planned vacation or seek information on baggage allowance/restrictions or flight information to pick up family and friend from the airport. It is also virtually costless (time, effort and travel cost) for a consumer to visit a website as opposed to a brick and mortar store. This behavior is very different and at the one end of the continuum; of the fallacies associated with a classic demand model; the other end of which is used to explain unplanned purchases resulting from browsing behavior especially in shopping malls (see Peter and Potts 2000).

Table 7: Comparison of Conditional Model with Model incorporating no-purchase

Effect	Inference		Notes
	Similar	Dissimilar	
First Site Visited and Browsing Duration	X		Browsing duration continues to be highest for users who visit Hotwire first. Propensity to purchase from site first visited is significant and more pronounced when incorporating no purchase transactions. Browsing longer on average increases likelihood of purchase for other travel portals (see Table 8).
Prior Browsing Experience	X		There is a positive effect of category prior browsing experience on the choice of first site to be visited. Its impact on purchase continues to be negative for travel portals except for Expedia (see Table 8). We also find the inverted-U relationship between browsing duration and prior browsing experience in the category. However its not significant (see Table 9 second set of results)
Prior Purchase	X		We find that a prior purchase continues to impact the first site to be visited. Though not significant the estimates indicate it also impacts the browsing duration among light users. However it doesn't seem to lower browsing duration among heavy users (see Table 9 third set of results). Prior purchase also seems to lower the purchase likelihood compared to our finding from the conditional model.
Consumer Demographics	X		While findings are consistent on Age, Income and broadband user effects. The only difference noticed was that purchase likelihood was higher for broadband users (see Table 9 last set of results)

Brand Strength	X		Brand strength continues to have a strong impact on being the first site to be visited as well as positively impact purchase for both Expedia and Orbitz compared to Hotwire or other Travel Portals (see Table 10)
Interactions, Heterogeneities and Model Fit		X	Inferences from the interactions that are significant are consistent with our findings from the conditional model however a lot more of the interactions are not significant in the no purchase model. Higher site heterogeneity estimates (see Table 11) make us believe that we could be adding more noise to the data by incorporating data from no purchase transactions

* note we don't include price effects in the model comparison as we only observe final purchase price in our dataset and the absence of price information in no-purchase transactions is another limitation of this dataset.

Overall we find results to be very similar between a conditional model and one that incorporates no purchase and the differences are summarized above (see Table 7).

5.2 Effect of First Site Visited and Browsing Duration

We find a significant effect of the first site visited on the browsing duration (see the first set of results in Table 9). On average consumers visiting Hotwire first tend to search longer (coefficient 0.77), followed by Expedia and Orbitz, compared to consumers visiting airlines websites first. This indicates a systematic difference in browsing behavior between the 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 the last column of the first set of results in Table 8, with a significant coefficient of 5.22). This strong effect is due to two observations from our data: first, a significant proportion of consumers only visit one website in the whole browsing process, indicating that they may have very limited consideration set, or that the potential benefit of further browsing is perceived as very small. Second, another significant proportion of consumers tend to buy from the website they first visited though they search other websites, indicating a priming effect from the first site visited. Either way it illustrates the benefits to a website if it can attract consumers to first visit it before starting to browse for information on other sites. We will further explore some of the implications later.

Table 8: Effect of first site visited, browsing duration & prior site specific experience on purchase

	Parameters	Stage 1 First site visited w/o no purchase	 w/ no purchase	Stage 2 Browsing duration w/o no purchase	 w/ no purchase	Stage 3 Purchase site w/o no purchase	 w/ no purchase
Stage 1 first Visited site	Expedia			0.686 (0.001)*	0.454 (0.282)*		
	Orbitz	N/A	N/A	0.577 (0.001)*	0.489 (0.290)*	5.223 (0.113)*	8.692 (0.034)*
	Hotwire			0.771 (0.001)*	0.705 (0.348)*		
	Other travel portals			0.374 (0.001)*	0.516 (0.299)*		
Stage 2 browsing duration	Expedia					0.281 (0.005)*	0.339 (0.039)*
	Orbitz	N/A	N/A	N/A		0.444 (0.009)*	0.392 (0.044)*
	Hotwire					0.293 (0.002)*	0.927 (0.045)*
	Other travel portals					0.559 (0.007)*	0.804 (0.045)*
Category prior experience	Expedia	0.008 (0.001)*	0.005 (0.002)*			0.008 (0.001)*	0.003 (0.001)*
	Orbitz	0.007 (0.001)*	0.003 (0.002)*	N/A		-0.001 (0.001)*	-0.001 (0.001)*
	Hotwire	0.007 (0.001)*	0.001 (0.003)			0.008 (0.001)*	0.001 (0.001)*
	Other travel portals	-0.072 (0.002)*	-0.023 (0.003)*			-0.121 (0.002)*	-0.005 (0.001)*

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

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 (see the second set of results in Table 8). This could be because travel portals are in general offering better deals compared with airline websites. This result suggests that by browsing longer consumers are more likely to find cheaper tickets from travel portals and hence end up buying there.

5.3 Effect of Prior Browsing Experience

We measure category prior browsing 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 the data. In general there is a positive effect of category prior browsing experience on the choice of renowned travel portals as the first site to be visited. However, it decreases the propensity to first visit other smaller travel portals as well as airline websites (see column three in the last set of results in Table 8). The effect of category prior browsing experience on the choice of travel portals for final purchase is mixed (see the last column). Though category prior browsing 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. Having price data from these travel portals would have helped us provide a more informed explanation.

We also investigate the effects of site specific and category prior browsing experience on browsing and purchasing. First, we study how site-specific prior browsing experience affects first site visit and purchase decisions. This variable is measured 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 six of Table 9).

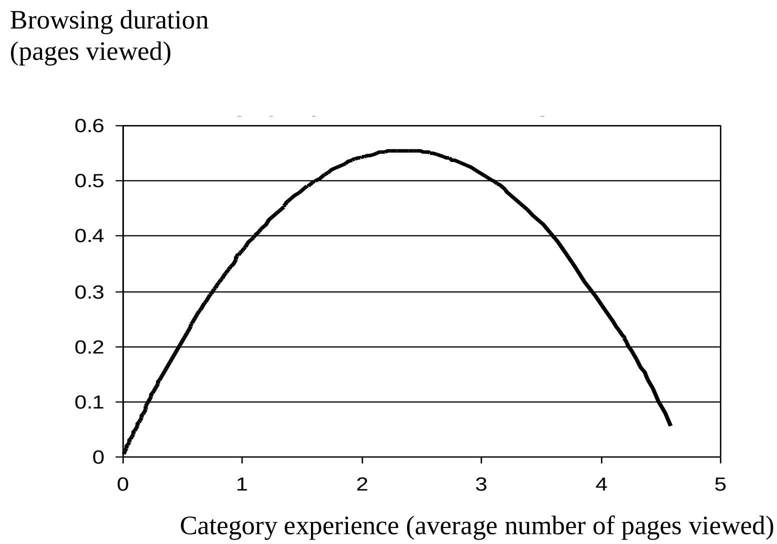
Table 9: Effect of prior browsing experience (site specific and category), prior purchase and demographics

Parameters	Stage 1 First site visited		Stage 2 Browsing duration		Stage 3 Purchase site	
	w/o no purchase	w/ no purchase	w/o no purchase	w/ no purchase	w/o no purchase	w/ no purchase
Site specific prior browsing experience	1.134 (0.021)*	0.001 (0.001)*	N/A	N/A	1.957 (0.024)*	-0.013 (0.004)*
Category prior browsing experience	N/A	N/A	0.426 (0.002)*	0.030 (0.072)*	N/A	N/A
Squared category prior browsing experience			-0.095 (0.001)*	-0.018 (0.057)*		
Prior purchase (one)	1.824 (0.045)*	0.992 (0.019)*	0.017 (0.001)*	0.194 (0.186)	0.154 (0.002)*	-0.352 (0.045)*
Prior purchase (more than one)			-0.244 (0.003)*	0.286 (0.224)		
Age	-0.005 (0.001)*	-0.038 (0.009)*	0.067 (0.001)*	0.179 (0.175)	-0.058 (0.002)*	-0.233 (0.012)*
Income	-0.032 (0.012)*	-0.074 (0.001)*	-0.054 (0.001)*	0.097 (0.130)	0.351 (0.010)*	-0.403 (0.015)*
Broadband users	0.116 (0.002)*	-0.246 (0.019)*	-0.017 (0.001)*	-0.038 (0.083)	-0.788 (0.014)*	0.086 (0.027)*
Medium level of expenditure	0.708 (0.014)*	N/A	0.465 (0.002)*	N/A	2.345 (0.043)*	N/A
High level of expenditure	0.689 (0.016)*	N/A	0.794 (0.001)*	N/A	0.457 (0.008)*	N/A

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

Second, we study how category experience (as defined above) affects how long consumers will browse before making a purchase decision. This is different from the effects on which site to be visited first and purchase as discussed above which are site-specific effects. Prior research has suggested that there may be a non-monotonic relationship between category prior browsing experience and browsing duration. We therefore include a squared term of prior browsing experience in the duration model (see column three in Table 9).

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



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). While online search effects have not previously been investigated in this detail there is evidence in prior research that online search though minimal in general, 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.

5.4 Effect of Prior Purchase

We create an indicator which takes the value 1 if a website is where the last purchase had occurred. This variable has a significantly positive effect on first site visited decision (see column two in Table 9), 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 six in Table 9), indicating that the likelihood of current purchase increases for a website if the last purchase occurred on it.

Similar to the non-monotonic relationship between category prior browsing experience and browsing, again we create two indicators for light and heavy users (as relative to those consumers without any purchase history in the first three months) in the browsing duration 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 9), again indicating an inverted U relationship between prior purchase experience and browsing duration.

5.5 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 9), 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 9). These results are consistent with the cost-benefit evaluation story and indicate consumers behave rationally. Interestingly, while higher expected expenditure (compared to low level) leads to a higher probability of purchasing from travel portals (see column four in Table 9), 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.

5.6 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 9). 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.

Table 10: Effect of brand strength on first site visited, browsing duration and purchase website

	Parameters	Stage 1 First site visited		Stage 2 Browsing duration		Stage 3 Purchase site	
		w/o no purchase	w/ no purchase	w/o no purchase	w/ no purchase	w/o no purchase	w/ no purchase
Brand Strength	Expedia	-1.437 (0.020)*	0.089 (0.034)*	2.776 (0.002)*	2.701 (0.686)*	-3.806 (0.019)*	-1.870 (0.079)*
	Orbitz	-1.624 (0.030)*	-0.345 (0.034)*			-4.239 (0.167)*	-1.827 (0.091)*
	Hotwire	-2.673 (0.021)*	-0.796 (0.034)*			-4.168 (0.056)*	-4.372 (0.093)*
	Other travel portals	-2.033 (0.006)*	-0.558 (0.034)*			-4.283 (0.064)*	-4.663 (0.089)*

* indicates $p < .05$. Compares estimates from our full conditional model with estimates from model with no purchase within parentheses

5.7 Effect of Brand Strength

Under our model set-up brand strengths in the first site visited model may be different from that in the purchase decision model, implying that a website attracting a lot of visits (e.g., through heavy advertising online or offline) may not be capable of converting these visits to final purchases (e.g., consumers may not find the deals attractive compared with other online options). To understand this difference in brand strength in terms of attracting first visit vs. converting them to purchases is more important in the online environment than offline, as the cost of visiting retail stores is much higher than visiting other websites. Table 10 reports the estimation results. We find that Expedia and Orbitz are significantly more attractive than Hotwire and other travel portals in attracting first visits (see column two). However, when it comes to final purchase decisions (see column four), Hotwire has larger (though not significant) brand strength than Orbitz, and that Expedia is still the strongest brand. These results show that different major travel portals have different attractiveness for first visits and final purchase decisions.

To understand better the implications of these estimates we compute the elasticities of first site visited and purchase probabilities as brand strengths change. The results are reported in Table 11. We find that by improving attractiveness of first visited by 1 percent induces a response which varies from about 1 percent more visits to Expedia to 2.5 percent more visits to Hotwire. However these increases in first site visits only imply a 0.3 percent increase in final purchase probability for Expedia and 0.1 percent for Hotwire. Increasing brand strength for the purchase decision stage (conversion effectiveness) by 1 percent will lead to about 2.4 percent increase in purchase probability for Expedia and about 3.2 percent for Hotwire.

Table 11: Elasticities of first site visited, browsing duration and choice of purchase site

Elasticities of first site visited and purchase site		Brand strength (attractiveness)	Brand strength (conversion effectiveness)
First site visited shares	Expedia	1.03	
	Orbitz	1.25	
	Hotwire	2.49	
Purchase shares	Expedia	0.28	2.39
	Orbitz	0.30	2.81
	Hotwire	0.09	3.17

Since we do not have data on marketing variables we cannot say much on how to improve attractiveness (first site visited) and conversion effectiveness (purchase), however, these results may still be useful for managers. For example, Hotwire may use its estimates as a base and re-estimate the model again after new policies are introduced. Comparisons of the estimates before and after help to evaluate how effective the new policies are in improving attractiveness for consumers' first visit vs. final purchases. Comparing elasticities is also useful in providing managerial insights. For example, it seems more important for Hotwire to invest in converting visitors to final purchases (e.g., through offering better deals or providing better online services) than in attracting website visits (e.g., through heavy advertising online or on TV).

5.8 Interactions, Heterogeneities, 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 browsing experience and the age of the consumer to be significantly positive indicating the effect of prior browsing 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 browsing 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.

[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.

Table 12: Site and price heterogeneity covariance parameters

Site heterogeneity			Price heterogeneity	
Parameter	Estimate w/o no purchase	Estimate w/ no purchase	Parameter	Estimate
σ_f	0.002 (0.001)***	0.016 (0.006)	σ_{mf}	0.321 (0.004)***
σ_{fd1}	0.001 (0.001)**	0.004 (0.024)	σ_{hf}	0.017 (0.001)***
σ_{fd2}	0.001 (0.001)*	0.007 (0.027)	σ_{mfd}	0.198 (0.001)***
σ_{fd3}	0.001 (0.001)***	0.007 (0.028)	σ_{hfd}	0.002 (0.001)***
σ_{fd4}	-0.001 (0.001)***	0.005 (0.028)	σ_{md}	0.250 (0.001)***
σ_d	0.001 (0.001)***	0.009 (0.031)	σ_{hd}	0.001 (0.001)***
σ_{fp}	0.001 (0.001)***	0.171 (0.005)	σ_{mfp}	0.001 (0.001)
σ_{pd1}	-0.001 (0.001)***	0.043 (0.080)	σ_{hfp}	0.047 (0.002)***
σ_{pd2}	0.001 (0.001)***	0.076 (0.090)	σ_{mdp}	-0.001 (0.001)
σ_{pd3}	0.008 (0.001)***	0.079 (0.093)	σ_{hdp}	0.007 (0.001)***
σ_{pd4}	0.001 (0.001)***	0.053 (0.091)	σ_{mp}	0.630 (0.011)***
σ_p	0.763 (0.001)***	1.847 (0.064)	σ_{hp}	0.272 (0.005)***

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

There is also a significant interaction between connection speed and site specific prior browsing 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 browse 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 13: 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 (Ln. 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

We do find significant heterogeneity in the behavior of consumers (the heterogeneity parameter estimates are summarized in Table 12). 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 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 covariance's are small in magnitude and can be interpreted based on our discussion in section 4.4.

Finally, we evaluate how good the model fits with data. We compute the expected share of first site visited and purchases on different websites as well as browsing duration based on the estimation results, and compare them to the actual shares and duration in data. Table 13 summarizes the results. Our three stage model clearly performs better in terms of predictive power and explaining the data. The hit rate of first site visited and purchase site is at 96.4%. As a comparison we estimate another model which does not account for the dynamic impact of first site visited or browsing duration separately and find that the hit rate of this model is only 56.4%. This demonstrates how our model explains the data better and can be used to improve prediction efficiency.

5.9 A Simulation Experiment

To better understand implications on how a website manager could employ this model to predict consumer browsing and purchase decisions and hence decide on suitable marketing policies we conduct a simulation experiment with hypothetical consumers to investigate the effect of last purchase on first site visited, browsing duration and choice of purchase site. We use 100 simulations to generate the random effects for 100 hypothetical users with mean values for all covariates except that for last purchase as we change which website they bought in the previous transaction to see its impact. Column 2 in Table 14 reports the expected first visit and purchase probabilities of various websites when there are no prior purchases. Columns 3 to 5 are the changes in the expected probabilities when prior purchase happened at Expedia, Orbitz and Hotwire, respectively. Since prior purchase experiences not only directly affect current purchase probability but also indirectly through the impact on first site visit probability, we first simulate the first site choice for each user and then plug in the choice of first site visited into the purchase decision model. Results show that when there is no prior purchase experience the model predicts that 65% of consumers would visit airline websites first and 67% end up buying from an airline website (see column 3 from Table 14). However when we change the last purchase experience on a particular website we find that consumers have a higher probability of first visiting and then buying from a website where last purchase occurred (see columns 2 to 6 from Table 14). For example, the probability of buying from Hotwire increases from 4 to 18 percent if our hypothetical consumers bought from the same website last time. This clearly indicates a significant dynamic future impact on revenues for a current transaction. Managers with relevant information can predict purchase probabilities better and take appropriate decisions in real time. For example, it may be more effective for a website to target new and current customers differently, given that the probabilities of first visit and final purchasing are different between these two groups of consumers.

Table 14: Effect of last purchase using hypothetical users

Policy experiment (effect of last purchase)		No prior purchase information	Change in Predicted Values		
			Last purchase on Expedia	Last purchase on Orbitz	Last purchase on Hotwire
First site visited shares	Expedia	0.16	0.33	-0.05	-0.02
	Orbitz	0.13	-0.05	0.29	-0.02
	Hotwire	0.04	-0.02	-0.01	0.14
	Other travel portals	0.01	-0.01	0.00	0.00
	All airlines	0.65	-0.26	-0.22	-0.09
Purchase shares	Expedia	0.15	0.36	-0.06	-0.03
	Orbitz	0.12	-0.06	0.34	-0.03
	Hotwire	0.04	-0.02	-0.02	0.18
	Other travel portals	0.01	0.00	-0.01	0.00
	All airlines	0.67	-0.28	-0.25	-0.11

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 later 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 browsing 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 site visited 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.

Another limitation of our model is that it discriminates between buyers and surfers without taking into account the information that they were exposed to due to limitations of the data set, as some consumers will browse on various websites but leave without making a purchase. 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 airline or hotel 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

Comscore Inc (2007), Retail E-Commerce Climbs 23 Percent in Q2 Versus Year Ago, *Press release*, July 30.

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

Earl Peter E. and Jason Potts, (2000), Latent Demand and the Browsing Shopper, *Managerial and Decision Economics*, 21,(3/4), April,111-122

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

Fazio,Eussell H., Carol J. Williams and Martha C. Powell, (2000), Measuring Associative Strength: Category-Item Associations and Their Activation from Memory, *Political Psychology*, 21,(1),7-25

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

Kim, Jun B., Paulo Albuquerque and Bart J. Bronnenberg (2009) Online Demand Under Limited Consumer Search, *forthcoming at Marketing Science*

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, (2004), 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

Weitzman, Mark, (1979), Optimal Search for the Best Alternative, *Econometrica*, 47, May, 641-54

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*,

Essay Two:

Modeling Online Multi-category Purchase in Travel

1. Introduction

Brand choice and store choice models have been extensively explored in depth by marketing researchers in the retail packaged goods context (for a brief review of an extended framework see Ben-Akiva et al 1999). This stream of literature has also spawned interest in multi-category or basket purchases (Harlem and Lodish 1995, for a review see Seetharaman et al 2005), retail chain level demand (Baltas 2005), store choice models (Keng & Ehrenberg 1984, Rust & Donthu 1995, Bell and Latin 1998), category characteristics and promotional elasticities (Narasimhan, Neslin and Sen 1996) and understanding purchase incidence at the grocery store in more detail (Manchanda, Ansari and Gupta 1999). Scant attention has been paid to category demand in prior literature, despite its importance for retailers. Only a small body of research has considered the effects of individual brands on category sales and there has been so far little work at the category level in understanding demand or consumer behavior especially in the online environment.

A recent extension of the online stream of literature into retail space has been furthering our understanding of shopping paths at the grocery store (Bradlow, Hui and Fader 2008). We believe this was to a large extent influenced by path analysis literature using clickstream data that became prevalent with the increasing dominance of the internet (Montgomery et al. 2004). In this paper we pursue this in the reverse direction and extend the work in multi-category and basket purchases in the retail space to online purchases thus contributing to the stream of literature on online basket purchases whilst drawing motivation from the past literature that exists in the retail packaged goods context.

In our work on basket purchases we steer clear of the concept of bundling (Chung and Rao 2003) as those constitute a class of travel products typically sold as vacation packages. We look at basket of travel purchases wherein the consumer makes multiple purchase transactions to constitute a single or combination of travel products.

2. Conceptual Development

We consider the consumer process of making a purchase comprising a basket of travel products in two stages. In the first stage the consumer decides on what combination of travel products to purchase and in the second stage embarks on making the purchase(s) from a particular travel portal or website. The first stage is assumed to be influenced by prior browsing and purchase experience undertaken for travel trips while the second stage is assumed to be influenced by the first stage as well as browsing experience prior to making a purchase. In addition to prior shopping behavior we also take into account consumer demographic factors as they have also been shown to contribute in affecting consumer choices (Ainsle and Rossi 1998). It is also a common practice in research to assume various stages in the consumer decision making and purchase process (Olshavsky and Granbois 1979). It's also possible that consumers have a budget or internal reference point (Bell and Bucklin 1999) for the various products in the category as part of their purchase decision.

2.1 Factors Affecting Basket Choice and Browsing

We expect consumer preferences to exhibit differences between choice of various basket combinations based on a few factors summarized below.

2.1.1 Basket preference

Consumers undertake travel for various reasons and depending on the travel needs that exist within the population its possible that there exists a base level of preference for various baskets or travel product combinations. We are especially interested in determining the base level preference for various combinations of travel products as no prior work has shed light on this. We believe that single product basket purchase would dominate the multiple product basket purchases as most travel is primarily driven by business or leisure trips from point A to point B or related to hotel stays in a particular geography being visited. We also hope to learn more about consumer preferences for road travel compared to air travel based on consumer preferences for car rental and air products and their combinations with hotel products.

2.1.2 Consumer Demographics

We also investigate the impact of consumer demographics on basket purchases for travel products. We expect larger households to make basket purchases in order to avoid inconvenience when traveling. Its possible older consumers would have less disposable income and could possibly be less likely to make basket purchases. Also those who have broadband connections are less price sensitive and hence could be more prone to making basket purchases online.

2.1.3 Category Experience

We also investigate the impact of prior browsing history for various product combinations. Its possible consumers exhibit some state dependence in their browsing habits and we hope to learn if consumers who browse more for a basket of products are likely to end up making a basket purchase or any purchase in general.

2.1.4 Prior Purchase

It is also possible that just like prior browsing experience that prior purchase of a basket of products could positively impact similar basket purchase or any purchase in general due to some state dependence or inertia effects.

2.1.5 Interaction Effects

Using interaction terms we are also interested in investigating the combined impact of demographic impact and prior browsing behavior on purchases. These interaction or combined effects could be very useful for managers to come up with a target profile of consumers who are more likely to make a purchase and tailor their media plans to effectively activate these target consumers. We also investigate the variance amongst consumer preference for various basket combinations but do not assume any expected relationship apriori.

2.2 Factors Affecting Site Choice and Purchase

The factors affecting basket purchase also have an impact on site choice and we investigate the impact of these factors further.

2.2.1 Site Preference

It is a known fact that consumers tend to frequent some stores as opposed to others and some of the factors impacting store choice have been documented in prior store choice literature. (Keng & Ehrenberg 1984, Rust & Donthu 1995, Bell and Latin 1998). We expect similar behavior in the online space and investigate the base level preference for various sites and their combinations when it comes to a basket purchase. We intend to tease out this effect from the intercepts associated with each site choice combination and delineate the differences of making a basket purchase on one site or a combination of multiple sites. We are also interested in seeing if there is a preference to make a basket purchase on travel portals compared to other sites.

2.2.2 Consumer Demographics

We also investigate the impact of demographic factors on basket purchases, especially their role in predicting website choice while pursuing a basket transaction. Prior studies have indicated price sensitiveness to be lower amongst broadband consumers and higher amongst larger households (Nair, Chan, Cheema 2009 working paper).

2.2.3 Category Experience

We believe consumers who have a lot of prior browsing history on a particular site could exhibit state dependence and favor such sites more compared to others when making a basket purchase. It is also possible for this prior experience to have a different influence depending on whether it's a travel portal or not. Overall we expect the impact of prior browsing history in travel to significantly and positively influence the likelihood of making a basket purchase from all site choice combinations.

2.2.4 Prior Purchase

We also believe prior purchase could result in purchase loyalty in subsequent purchases wherein consumers are more likely to make basket purchases from the same site.

2.2.5 Interaction Effects

The interaction effects of demographics with prior travel browsing history and prior purchase on same site combinations are also very interesting and worth investigating because it has targeting implications that manager's can act on to maximize likelihood of basket purchases. In addition to accommodating various interaction effects we also account for site heterogeneity exhibited by consumers to better understand differences between sites when it comes to basket purchases

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[11] that are a representative sample of the US population in 27 product categories. In this study we extend our earlier work by modeling purchases in three categories airlines, car rental and hotel, however, we restrict ourselves to modeling the purchase behavior of consumers and do not explicitly model the browsing behavior. A total of 8937 households in the travel category fit the criteria required for our analysis. To

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

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 which captures 96% of all search that consumers indulge in); (2) we only study the household's purchase in the travel category (travel portals, airline, car rental and hotel 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 period.

Travel category forms a significant portion of online purchases made by consumers with the mean being two out of ten purchases amongst the 27 product categories. As we are interested in studying purchase behavior that is related to a purchase across multiple categories we focus only on those travel websites that also provide an option for consumers to make such a purchase (airline, car rental or hotel). Specifically we investigate purchase behavior on travel portals (such as Expedia, Orbitz, and Hotwire) and category specific websites (like Southwest, Delta, and American airlines; Hertz, Avis, Budget, Alamo and Enterprise for car rentals and Hilton, Hotels.com, Choicehotels.com, Sixcontinentshotels.com and Mariott for hotels) where consumers have an option to purchase the travel product online.[12]

We used the browsing and purchase behavior of households in the first three months of data (July 2002 – September 2002) as the household's prior experience on

[12] 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). We also excluded transactions on websites which were auction sites, search engines and payment gateways such as ebay.com, lycos.com, netbooker.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).

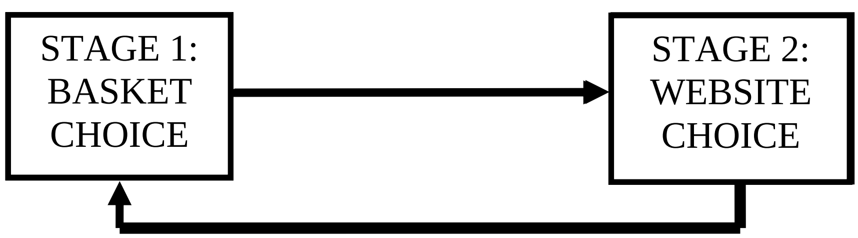
travel websites. We then use the transaction sessions for airline, car rental and hotel purchases in the last three months for model estimation.

In this study we focus on purchase behavior for (i) airline tickets (ii) car rental and (iii) hotel purchases that are either bought alone or in conjunction with another product.

4. Model Specification and Estimation

We propose a two stage model of purchase behavior for travel products. In our model consumers choose the products at the basket level first followed by the choice of the website where they can buy the basket of products. Hence the two stages we model are (i) choice of products at the basket level and (ii) choice of purchase site. This framework is pictorially depicted in Figure 1.

Figure 1: Proposed two-stage model of basket level purchase behavior



In the first part of this section we outline the model used to study the factors that predict basket composition. Initial data analysis revealed it is important to model choice of what products comprise the consumers' basket as that would considerably reduce the number of options that are available to the consumer in subsequent stages when it comes to choice of the purchase site. Understanding what factors affect the composition of products at the basket level would help us to also gain insights on the influence the basket composition has on which site consumers would finally end up making their purchase on. The different product combinations that comprise the basket level decision of the consumer are as summarized in Table 1.

Table 1: Basket level choices

Travel Category Purchases	Unique Households	Transactions (6 Mos.)	Transactions (Last 3 Mos.)	Transaction sets
Airline Only	6029	8074	3754	3754
Car Rental Only	1072	1578	734	734
Hotel Only	2921	4054	1799	1799
Car and Hotel	145	364	181	76
Air and Hotel				210
Air and Car Rental	856	2620	1233	158
Air, Car and Hotel				65
Total	8937	16690	7619	6796

In the second part of this section we model the website choice, in particular, we focus on the choice of purchasing from a particular travel portal. We focus on travel portals because most travel portals provide the opportunity of purchasing airline, hotel and car products on the same site though this has now also become more prevalent on airline websites too. The other compelling reason is that it adds to the parsimony of our model which can be easily extended to incorporate other websites too.

4.1 Modeling Basket Level Choice

We study the basket level choice behavior of consumers prior to making a purchase by using a random coefficients approach of the traditional multinomial logit model (for example see Gudagni and Little 1983) which we explain in detail later. We treat each combination of the three major travel products airline, car rental and hotel bookings as a choice that the consumer makes to constitute the basket. As we are interested in what constitutes the basket and not the actual sequence of purchases we club airline purchases followed by a car rental purchase and car rental purchase followed by airline purchase together.

Table 2: Website choice

Website Type	Product Category		
	Airline	Car Rental	Hotel
Travel Portals		Expedia Orbitz Travelocity Hotwire Other Travel portals	
		Other sites*	
* Sampling of other sites (includes other travel portals and all other airline, car rental and hotel websites)	Southwest	Avis	Hilton
	Delta	Budget	Sixcontinentshotels.com
	American	Hertz	Hotels.com
	Jetblue	Alamo	Choicehotels.com
	US Airways	Enterprise	Mariott

Let a discrete variable $B_{ict} = 1$ indicate that consumer i purchases combination c at time period t, and $B_{ict} = 0$ otherwise. For $B_{ict} = 1$, combination c has to exist in consumer i's consideration set (which includes all possible options) and then c has to dominate other combinations in this consideration set in terms of propensity to purchase that combination under cost-benefit evaluation. We assume that these are determined by a list of factors including customer demographics (age, income, and connection speed) Z_{it}, prior category experience H_{it}, expected level of expenditure P_{it} and prior combination-specific browsing experience S_{ij}. Prior category experience H_{it} is measured as the proportion of combination specific c pages viewed to the total pages viewed in the first three months on all websites selling travel products, and prior combination-specific browsing S_{ij} is measured as average daily pages viewed on combination specific websites c in the first three months. Furthermore, I_{ict} is an indicator variable that denotes whether or not consumer *i 's* last purchase was a specific combination c. This variable may affect the probability of c being in *i 's* consideration set and may create inertia such that i may be more likely to purchase the same combination during the next purchase cycle. Finally, basket level choice is also affected by i's preference for a particular combination c that is independent from the above factors as well as accounts for any marketing activities for a specific combination c which is not observed in our data. This is termed as "basket preference" which is individual-and-time-specific in our model.

We assume that there is a latent variable B^*_{ict} that generates the basket choice decisions. $B_{ict} = 1$ if and only if $B^*_{ict} \geq B^*_{idt}$, for all other combinations d. We specify the function of this latent variable as

$$B^*_{ict} = \alpha^f_{ict} + \beta^f Z_{it} + \tau^f_c H_{it} + \gamma^f S_{ic} + \rho^f I_{ict} \qquad (1)$$

In the above equation the superscript "*f*" denotes the first stage of our two stage model. The variable α^f_{ict} represents the latent basket preference for a particular combination. We use a random effects approach to model this variable as the follows

$$\alpha^f_{ict} = \alpha^f_c + \xi^f_{ic} + \varepsilon^f_{ict} \qquad (2)$$

where α_c^f represents the mean basket preference for a particular combination that will be estimated as parameters, ξ_{ic}^f represents the individual-specific but time-invariant random effect for a combination, and ε_{ict}^f is the individual-and-time-specific idiosyncratic shock that we assume to be i.i.d. type one extreme value distribution.

As we will explain later, we however do not allow these random effects to be correlated with the random effects in the second stage of the model as we believe that the basket preferences are more intrinsic and are not that easily changed by marketing interventions on a particular website and hence can be modeled independent from the choice of the purchase site which we model in the second stage. Since we explicitly model all possible combinations for the choice of the websites on which the basket is purchased the dimensionality of parameters is very large and accounting for all the correlations across each stage adds to further complexity.

To ensure proper identification we normalize the latent variable value for all hotel and car purchases as $B_{i,HC,t}^{*} = 0 + \varepsilon_{i,HC,t}^{f}$.[13] In our estimation model we also incorporate the interactions of demographic variables Z_{it} with all other covariates H_{it}, S_{ic}, and I_{ict}. Let T_i be the total number of household i 's purchases observed in data. Correspondingly there are T_i combination purchases and each combination of purchase is defined as a set. Under the type one extreme value distribution for ε_{ict}^f and conditional on the random

[13] 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 prefer a combination relative to their preference for hotel and car combination purchase.

effect ξ_{ic}^{f} , we can write down the probability that a household's basket preference for a particular combination in the whole sample period as below.

$$\Pr(i\text{'s history of basket preference}) = \prod_{t=1}^{T_i} \Pr(B_{ict}^{*} \geq B_{idt}^{*}, \forall d)$$
$$= \prod_{t=1}^{T_i} \frac{e^{\alpha_c^f + \xi_{ic}^f + \beta^f Z_{it} + \tau_c^f H_{it} + \gamma^f S_{ic} + \rho^f I_{ict}}}{1 + \sum_{c=1}^{C} e^{\alpha_d^f + \xi_{id}^f + \beta^f Z_{it} + \tau_d^f H_{it} + \gamma^f S_{id} + \rho^f I_{idt}}} \qquad (3)$$

4.2 Modeling Choice of Purchase Site

In the second stage of the model we model the choice of the purchase site for each combination in the basket level choice. We illustrate the methodology for a combination purchase of which pertains to buying airline tickets A and hotel bookings H . Consumer i can purchase both these products in his/her basket from the same website or from different websites. The different combinations of making these purchases from the same travel portal $\{ j , j \}$ or different travel portals $\{ j , k \}$ or from a travel portal and one of the other sites (other travel portal, airline, car rental and hotel sites) $\{ j , other \}$ or make both purchases from one of the other sites $\{ other , other \}$ can be expressed as below.

$\{ j , j \}$: (6 options)

$\{ j , k \}$: (15 options)

The utility for these 21 choice combinations can be summarized as in Table 3. In the above expression $V_{j,t}^{A}$ is the indirect utility from making the airline purchase from website j , $V_{j,t}^{H}$ the indirect utility from making a hotel purchase from website j and the utilities when an airline purchase and hotel purchase are made from one of the other sites they are expressed as $V_{o,t}^{A}$ and $V_{o,t}^{H}$. The probability of both airline and hotel purchase being on a travel portal can now be written as

Table 3: Utility function for a combination of Airline and Hotel purchase

Hotel / Air	Travel Portal	Other Site
Travel Portal	$V_{j,t}^{A}+V_{j,t}^{H}+u_t+\varepsilon_{j,t}^{A}+\varepsilon_{j,t}^{H}$	$V_{j,t}^{A}+V_{o,t}^{H}+v_t+\varepsilon_{j,t}^{A}+\varepsilon_{o,t}^{H}$
Other Site	$V_{o,t}^{A}+V_{j,t}^{H}+v_t+\varepsilon_{o,t}^{A}+\varepsilon_{j,t}^{H}$	$V_{o,t}^{A}+V_{o,t}^{H}+u_t+\varepsilon_{o,t}^{A}+\varepsilon_{o,t}^{H}$

The utility for these 21 choice combinations can be summarized as in Table 3. In the above expression $V_{j,t}^{A}$ is the indirect utility from making the airline purchase from website j, $V_{j,t}^{H}$ the indirect utility from making a hotel purchase from website j and the utilities when an airline purchase and hotel purchase are made from one of the other sites they are expressed as $V_{o,t}^{A}$ and $V_{o,t}^{H}$. The probability of both airline and hotel purchase being on a travel portal can now be written as

$$\begin{aligned}
\text{Prob[j,j]} &= \text{Prob [} V_{j,t}^{A}+V_{j,t}^{H} \geq \max(V_{j,t}^{A}+V_{o,t}^{H}, V_{o,t}^{A}+V_{j,t}^{H}, V_{o,t}^{A}+V_{o,t}^{H})] \\
&= \text{Prob [} V_{j,t}^{A}+V_{j,t}^{H} \geq V_{j,t}^{A}+V_{o,t}^{H} \text{] .} \\
&\quad \text{Prob [} V_{j,t}^{A}+V_{j,t}^{H} \geq V_{o,t}^{A}+V_{j,t}^{H} \,/\, V_{j,t}^{A}+V_{j,t}^{H} \geq V_{j,t}^{A}+V_{o,t}^{H} \text{] .} \\
&\quad \text{Prob [} V_{j,t}^{A}+V_{j,t}^{H} \geq V_{o,t}^{A}+V_{o,t}^{H} \,/\, V_{j,t}^{A}+V_{j,t}^{H} \geq \max(V_{j,t}^{A}+V_{o,t}^{H}, V_{o,t}^{A}+V_{j,t}^{H})]
\end{aligned} \tag{4}$$

Since $\varepsilon_{j,t}^{A}$ and $\varepsilon_{j,t}^{H}$ are assumed to be independent hence the conditional term in the second part of the equation drops out and the above equation reduces to

$$\begin{aligned}\text{Prob[j,j]=Prob [} & \varepsilon^{H}_{j,t} \geq V^{H}_{o,t} + v_t + \varepsilon^{H}_{o,t} - V^{H}_{j,t} - u_t \text{] .} \\ & \text{Prob [} \varepsilon^{A}_{j,t} \geq V^{A}_{o,t} + v_t + \varepsilon^{A}_{o,t} - V^{A}_{j,t} - u_t \text{].} \\ & \int 1 \text{ [} \varepsilon^{A}_{j,t} + \varepsilon^{H}_{j,t} \geq V^{A}_{o,t} + V^{H}_{o,t} + u_t + \varepsilon^{A}_{o,t} + \varepsilon^{H}_{o,t} \text{-} V^{A}_{j,t} - V^{H}_{j,t} - u_t \text{] .} \\ & dF(\varepsilon^{A}_{j,t} / \varepsilon^{A}_{j,t} \geq V^{A}_{o,t} + v_t)] \text{ .} dF(\varepsilon^{H}_{j,t} / \varepsilon^{H}_{j,t} \geq V^{H}_{o,t} + v_t)]\end{aligned} \tag{5}$$

Now let us consider the case where airline is bought from website m and hotel is bought on website n. Then the 21 choices that are available to purchase a combination of airline tickets and hotel booking are as depicted in Table 4.

Table 4: Possible website combinations for Airline and Hotel purchase

Air m \ Hotel n		1 Expedia	2 Travelocity	3 Orbitz	4 Hotwire	5 Other TP	6 Other Site
1	Expedia	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	Travelocity	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	Orbitz	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	Hotwire	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	Other TP	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	Other Site	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

As before, now the probability of a consumer i purchasing airline and hotel from website m and website n respectively can be written as

Prob[m,n]=Prob [A] . Prob [B] . Prob [C]

$$\begin{aligned}\text{where A is [} & \varepsilon^{H}_{n,t} \geq \max_{k}(V^{H}_{k,t} + v_t + \varepsilon^{H}_{k,t}, \forall k \neq n) - V^{H}_{n,t} - u_t \text{]} \\ \text{B is [} & \varepsilon^{A}_{m,t} \geq \max_{k}(V^{A}_{k,t} + v_t + \varepsilon^{A}_{k,t}, \forall k \neq m) - V^{A}_{m,t} - u_t \text{] and} \\ \text{C is [} & \varepsilon^{A}_{m,t} + \varepsilon^{H}_{n,t} \geq \max_{k,l}(V^{A}_{m,t} + V^{H}_{n,t} + u_t + \varepsilon^{A}_{o,t} + \varepsilon^{H}_{o,t}, \forall k \neq m, \forall l \neq n)\text{-}V^{A}_{m,t} - V^{H}_{n,t} - u_t \text{]}\end{aligned} \tag{6}$$

Note that in the above equations u_t is the component of utility that consumers derive from making both the airline and hotel purchase from the same website and is explicitly modeled as a function of loyalty and {other independent variables}.

$$\begin{aligned} &u_{mm,t} = \alpha_0 + \beta_1 loyalty_m + \beta_2 loyalty_j + \text{other indep variables} \\ &u_{mn,t} = \alpha_1 + \beta_2 loyalty_j + \text{other indep variables} \\ &\text{note:} \\ &loyalty_m \text{ is an indicator variable that takes value 1 if both purchases are on same website} \\ &loyalty_j \text{ is an indicator variable that takes value 1 if both purchases are on a travel portal} \end{aligned} \tag{7}$$

The indirect utility functions for purchasing an airline $V_{j,t}^A$ or hotel $V_{j,t}^H$ from a particular website website j is also expressed as a function of {independent variables - including consumer demographics Z_{it}, prior browsing experience as the average daily pages viewed S_{ij} on different sites in the first three months, H_{it} 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} }. This concludes our discussion for just one combination AH out of the possible 7 combinations as tabulated in Table 1. The above expressions can easily be extended and similar equations obtained for each of the different combinations.

4.3 Model Estimation

The major difficulty in model estimation comes from dimensionality and to simplify the model we assume that the random effects in each stage and the random effects for purchases in different categories (e.g. $\varepsilon_{j,t}^A$ and $\varepsilon_{j,t}^H$) are also independent of each other. We assume the random effects to be normally distributed, as

$$\begin{aligned}&\varepsilon_{m,t}^{A} \sim N(0,\sigma_A^2)\\&\varepsilon_{n,t}^{H} \sim N(0,\sigma_H^2)\\&\varepsilon_{m,t}^{A}+\varepsilon_{n,t}^{H} \sim N(0,\sigma_A^2+\sigma_H^2)\end{aligned} \tag{8}$$

To solve the second stage of the model we use the simulated maximum likelihood approach for estimation. Let $\Psi_i = \{\varepsilon_{j,t}^{A}, \varepsilon_{j,t}^{H}, \varepsilon_{j,t}^{C}; \forall j;\ u_t;\ v_t\}$ be the vector of random effects in the equation system with the assumed distribution $F(\Psi,\Omega)$, where Ω is the set of all parameters to be estimated. The unconditional likelihood can now be expressed as follows:

$$L_i = \int\left(\underbrace{\Pr(A|\Psi_{\mathrm{i}})}_{\text{equation (6)}}\cdot\underbrace{\Pr(B|\Psi_{\mathrm{i}})}_{\text{equation (6)}}\cdot\underbrace{\Pr(C|\Psi_{\mathrm{i}})}_{\text{equation (6)}}\right)dF(\Psi_i;\Omega) \tag{9}$$

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 algorithm used to estimate the second stage of the model is as shown below

Step 1: Take 1000 draws of ε_t^s (6x6) for each element of Table 4

Step 2: Evaluate A as $\frac{1}{1000}\sum_{s=1}^{1000}\text{Prob}(\varepsilon_{n,t}^{H} \geq \max_k(V_{k,t}^{H}+v_t+\varepsilon_{k,t}^{H}, \forall k \neq n) - V_{n,t}^{H} - u_t)$

Step 3: Similarly evaluate B as

$$\frac{1}{1000}\sum_{s=1}^{1000}\text{Prob}(\ \varepsilon_{m,t}^{A} \geq \max_k(V_{k,t}^{A}+v_t+\varepsilon_{k,t}^{A}, \forall k \neq m) - V_{m,t}^{A} - u_t)$$

Step 4: Evaluate C for each ε_t^s,

$$\text{Prob(m,n)}= \frac{1}{1000}\sum_{s=1}^{1000}[\ \text{Prob}(A^s).\text{Prob}(B^s)\{\geq 1\} + \text{Prob}(C^s)\{<1\}]$$

The corresponding simulated version of (9) can be expressed as

$$\hat{L}_i = \frac{1}{ns}\sum_{s=1}^{ns}\left(\underbrace{\Pr(A|\Psi_i^s)}_{\text{equation (6)}}\cdot\underbrace{\Pr(B|\Psi_i^s)}_{\text{equation (6)}}\cdot\underbrace{\Pr(C|\Psi_i^s)}_{\text{equation (6)}}\right) \tag{10}$$

Note: The superscript s in the above expressions implies simulated values.

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

In this section we discuss results for the first stage of the model identifying factors impacting basket choice followed by results from the second stage of the model which sheds more light on the website choice for the various basket combinations. Though our modeling and estimation methodology can be easily extended to a basket with more than two products we limit our estimation to a basket with two products as our model has 79 parameters to be estimated and there are very few observations (195) for basket transactions with all three travel products in our data

5.1 Factors Impacting Basket choice

We infer the preference of various basket combinations from the intercepts associated with each basket combination and find that on average most consumers tend to have a preference to purchase a single product (air or hotel) and not a basket of travel products. This could be primarily driven by business or leisure airline trips from point A to point B (with car rental or hotel transactions conducted offline) or hotel stays in a particular geography being visited (with airline tickets purchase being done via corporate ticketing or offline). Amongst the basket of travel products consumers tend to have a higher preference on average for a basket of products compared to a car rental only product.

Table 5: Base level basket preference and demographic factors

	Parameters	Estimates
Base level preference for various Basket Combinations	Air and Hotel Combo	-0.157 (0.063)*
	Air and Car rental Combo	-0.232 (0.103)
	Air, Hotel and Car rental Combo	-0.146 (0.073)*
	Airline Only	1.956 (0.052)*
	Hotel Only	1.048 (0.050)*
	Car rental Only	-0.254 (0.053)*
Demographic factors	Household Size	0.107 (0.474)
	Age	0.057 (0.396)
	Income	0.045 (0.347)
	Child Present	0.180 (0.510)
	Connection Speed	0.087 (0.075)

* indicates p < .05. Standard errors are in parentheses

Amongst the basket of products the combination of all three i.e. air, hotel and car rental products has a higher base level preference compared to any two way combination of travel products with the exception of car and hotel. The higher preference for car and hotel basket leads us to believe there is significant preference for road travel amongst consumers. Note that since our model is conditional on a purchase being made - the outside option is a basket of car rental and hotel product (see Table 5 first set of results). Though we find directional evidence that consumers who make basket purchases on average are more likely from a larger household, older, have a broadband connection, higher income and more likely to have a child in the family we don't find the demographic parameters to be significant (see Table 5 second set of results).

We also investigate the impact of prior browsing history for various product combinations and find that on average consumers who browse more for a basket of products are likely to end up making a purchase, and this likelihood is more pronounced towards purchasing a single product as opposed to a basket of products.

Table 6: Prior purchase and browsing history (category and basket specific)

	Parameters	Estimates
Prior browsing history in basket	Air and Hotel Combo	0.001 (0.000)*
	Air and Car rental Combo	0.000 (0.000)
	Air, Hotel and Car rental Combo	0.001 (0.000)*
	Airline Only	0.001 (0.000)*
	Hotel Only	0.002 (0.000)*
	Car rental Only	0.003 (0.000)*
Prior browsing history in travel		0.288 (0.376)
Prior purchase of same basket/combo		-0.039 (0.074)

* indicates p < .05. Standard errors are in parentheses

Though these inferences are directionally insightful we find that these results are significant but for Air and Car rental combo (see Table 6 first set of results). We also find a positive impact of browsing within the travel category to positively impact purchase (see Table 6 second set of results) and this is consistent with our findings from an airline only model (see Nair, Chan and Cheema, 2009 working paper). Though not significant we find directional evidence that purchasing a basket of products in the past doesn't seem to increase the likelihood for repeat purchase of same product combo (see Table 6 third set of results).

Table 7: Interaction Effects

	Interactions	Estimates
Prior Browsing history in travel	Age	0.587 (0.023)*
	Income	0.622 (0.060)*
	Connection Speed	-1.608 (0.117)*
Prior Purchase of same basket/combo	Age	0.127 (0.027)*
	Income	0.622 (0.060)*
	Connection Speed	-0.226 (0.106)*

* indicates $p < .05$. Standard errors are in parentheses

Based on the various interaction effects we incorporate in our modeling effort we find significant evidence that consumers who browse more and are either older or have higher incomes have a higher propensity to make a purchase(see Table 7 first set of results). It is also evident that consumers who made a prior purchase and were either older or have higher income were significantly predisposed to make a purchase (see Table 7 second set of results). The results for interaction of connection speed with prior browsing history and purchase were significant and this pattern could be driven by offline deals that are sought by these consumers. These are insightful for managers to create a target profile of consumers who are more likely to make a purchase.

We find higher variance amongst consumer preference for Air & Car rental basket combination (0.2472), Car rental only (0.1283) and Air, Hotel and Car rental combination (0.0132) purchases in that order (see Table 8).

This could primarily be driven by difference in preferences between leisure and business travelers and is something that needs to be investigated in future studies with richer data that distinguishes between these user types. We also find the covariance between Hotel only purchase with Air & Hotel combo (-0.0040) and Airline only (-0.0011) purchase to be negative indicating consumers could be considering these transactions as substitutes.

Table 8: Basket choice Variance-Covariance parameter estimates

Parameter	σ_{AH}	σ_{AC}	σ_{AHC}	σ_{A}	σ_{H}	σ_{C}
σ_{AH}	0.0020*					
σ_{AC}	0.0001	0.2472*				
σ_{AHC}	0.0009*	0.0023	0.0132*			
σ_{A}	0.0010*	0.0040*	0.0026*	0.0009*		
σ_{H}	-0.0040*	0.0093*	0.0037*	-0.0011*	0.0110*	
σ_{C}	0.0004	0.1109*	0.0050*	0.0027*	0.0067*	0.1283*

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

This could be possibly due to the typical behavior exhibited by business travelers i.e. do a day trip (airline only), or choose to fly in and stay overnight at a hotel near final destination and use a cab to get to the work location or rent a car to drive to and out of the work location same day. We also find that the covariance of Hotel Only purchase with Air and Car rental only purchase to be higher and positive (0.0093) indicating they could be complementing each other and consumers might be using these separate purchases in lieu of a combination of Air, Hotel and Car rental combo purchase.

5.2 Factors Impacting Website choice for various basket purchases

We infer the base level preference for various site choice combinations when it comes to a basket purchase from the intercepts associated with each site choice combination and find that on average most consumers tend to have a preference for purchasing a basket of travel products from different sites one of them being what we define as other site (mostly airline site).

Table 9: Base level site choice preference for basket purchases

	Parameters	Estimates
Base level site choice preference for basket purchases	Expedia only	-5.0318 (10.300)
	Expedia & Travelocity	-12.214 (0.254)*
	Travelocity only	-7.400 (0.782)
	Expedia & Orbitz	-5.045 (0.054)*
	Travelocity & Orbitz	-5.864 (0.418)*
	Orbitz only	0.357 (0.078)*
	Expedia & Hotwire	-8.077 (0.000)*
	Travelocity & Hotwire	-7.169 (0.000)*
	Orbitz & Hotwire	-7.148 (0.062)*
	Hotwire only	-7.798 (0.000)*
	Expedia & Other portal	-6.508 (0.000)*
	Travelocity & Other portal	-7.194 (1.880)
	Orbitz & Other portal	-7.087 (0.826)
	Hotwire & Other portal	-9.547 (0.000)*
	Other portal only	-5.888 (0.799)
	Expedia & Other site	0.989 (0.000)*
	Travelocity & Other site	1.073 (0.113)*
	Orbitz & Other site	0.819 (0.022)*
	Hotwire & Other site	0.171 (0.000)*
	Other portal & Other site	-0.432 (0.287)

* indicates p < .05. Standard errors are in parentheses

This could be primarily driven by air travel being an important part of any travel plan as well as the loyalty connected with airline rewards programs. Amongst those consumers who choose to complete their basket of travel product purchases on the same portal we find the base level preference to be highest for Orbitz followed by Expedia, Other travel portal, Travelocity and Hotwire in that order (see Table 9). The base level preference is lowest for site combo of Expedia & Travelocity when it comes to making a basket of travel purchases. Note that since our model is conditional on a purchase being made the outside option is a basket purchase made on Other sites only.

Table 10: Demographics and Prior browsing history (site specific)

	Parameters	Estimates
Demographic factors	Household Size	-0.753 (0.001)*
	Age	-0.309 (0.001)*
	Income	-0.579 (0.001)*
	Child Present	1.514 (0.024)*
	Connection Speed	-0.164 (0.023)*
Prior browsing history by site choice combinations	Expedia only	-0.621 (0.001)*
	Expedia & Travelocity	-2.102 (0.002)*
	Travelocity only	-1.215 (0.003)*
	Expedia & Orbitz	-10.074 (0.001)*
	Travelocity & Orbitz	- 1.598 (0.001)*
	Orbitz only	-4.701 (0.006)*
	Expedia & Hotwire	-5.290 (0.001)*
	Travelocity & Hotwire	-1.173 (0.001)*
	Orbitz & Hotwire	-0.488 (0.001)*
	Hotwire only	-4.970 (0.001)*
	Expedia & Other portal	-1.429 (0.001)*
	Travelocity & Other portal	-2.557 (0.001)*
	Orbitz & Other portal	0.115 (0.036)*
	Hotwire & Other portal	-0.626 (0.001)*
	Other portal only	-0.061 (0.001)*
	Expedia & Other site	-0.817 (0.001)*
	Travelocity & Other site	-6.604 (0.001)*
	Orbitz & Other site	16.437 (0.001)*
	Hotwire & Other site	-0.516 (0.001)*
	Other portal & Other site	-8.883 (0.001)*

* indicates $p < .05$. Standard errors are in parentheses

Though we didn't find demographic factors to be significant and only directional when it came to basket purchases, we find them to be significant in predicting website choice while pursuing a basket transaction. A child's presence in the household positively influences basket purchases across all site choice combinations while consumers who have a larger household or are older or have higher income or broad band connections are less likely to complete basket purchases across all site choice combinations (see Table 10 first set of results). Consumers who have a prior browsing history on two specific combinations of sites (i) Orbitz and Other sites (mostly airline sites) and (ii) Orbitz and Other travel portals exhibit a higher likelihood of making a basket purchase across all site choice combinations. Consumers who have prior browsing experience on Expedia and Orbitz (-10.074 see Table 10 second set of results) being the least likely to make a basket purchase across all site choice combinations. Others who exhibit similar behavior worth noting are those with prior browsing experience on Other travel portals and Other sites(mostly airline), Travelocity and Other site (mostly airline), Expedia and Hotwire.

Table 11: Prior purchase, browsing history (category specific) and Interaction Effects

	Interactions	Estimates
Prior browsing history in travel		3.937 (1.560)*
Prior purchase from same choice combo		2.237 (0.861)
Prior Browsing history in travel	Age	-2.071 (0.028)*
	Income	-0.501 (0.080)*
	Connection Speed	7.847 (2.231)
Prior Purchase of same site combo	Age	10.333 (0.017)*
	Income	7.367 (0.024)*
	Connection Speed	4.638 (2.056)

* indicates p < .05. Standard errors are in parentheses

We also find a significant impact of overall prior browsing history in travel to significantly and positively influence the likelihood of making a basket purchase from all site choice combinations. Note however that though there is a positive impact of prior purchase on the various site choice combinations on a basket purchase form the same site choice combinations this result is not significant (see Table 11 first set of results).

The interaction effects of demographics with prior travel browsing history and prior purchase on same site combinations were also investigated and the impact of connection speed though positive was not found to be significant. However older consumers or those with higher income had a significant difference in the way they influenced basket purchases across all site combinations decreasing the likelihood when the prior browsing history was higher – learning impact; and increasing the likelihood when prior purchase history was higher – transaction impact (see Table 11 second set of results).

In our models we also incorporate intercepts to measure the base level of preference (more like brand equity) to make a basket of purchases separately (i) from the same website and (ii) from travel portals to see if they have an impact collectively as a group. We also use an indicator variable to separate out the impact of loyalty of making the basket of purchases from (i) the same website (more like state dependence or stickiness) or (ii) from travel portals collectively as a group. This is an important nuance of our model the interpretation of which helps us understand the impact these factors have on influencing basket purchases for various site combinations. We find that consumers have a base level preference to make basket purchases from the same website and the large significant coefficient on this suggests that the brand equity or rewards programs on various sites which could be causing this dominates most other factors including loyalty or state dependence effects arising from a basket purchase (see Table 12 first set of results). This could also be an outcome of consumers executing on their basket purchases based on the choices post a search on all other sites. We also find that consumers have a lower base level preference of purchasing their basket from travel portals collectively as a group. This could be because (i) travel portals might be attempting to extract more consumer surplus through bundling and (ii) consumers are successful in finding better deals on basket purchases directly from non-travel portals i.e. service providers like airline sites.

When it comes to loyalty we find that consumers have a loyalty or stickiness for making basket purchases on the same website as well as from a travel portal and both of these results are significant and positive however the collective impact of travel portals compared to non-travel portals is greater than the loyalty that is related to making basket purchases from the same website (see Table 12 second set of results). This could be because (i) non-travel portals offer slightly lower number of alternatives or (ii) the ease of use when it comes to making basket purchases on a travel portal as they are by design geared up to offer many alternatives from which a consumer could make a choice. Note during the period of study non-travel portal sites didn't have much options in this regard.

Table 12: Base level preference and loyalty effect of purchasing entire basket from same site or travel portal

	Parameter	Estimates
Base Level preference to make purchases from	Same Website	10.511 (0.001)*
	Travel Portal	-0.647 (0.001)
Loyalty	Same Website	0.285 (0.001)*
	Travel Portal	0.500 (0.001)*

* indicates $p < .05$. Standard errors are in parentheses

Due to the complexity of the model and the limited number of observations that were available we do not find the estimates of the variance covariance matrix to be significant (see Table 13 for a summary of these results). However we note that these estimates can be used directionally to make some inferences that could be worth diving deeper into when voluminous transaction data for online basket purchases becomes available.

Table 13: Basket choice Variance-Covariance parameter estimates

Parameter	$\sigma_{Expedia}$	$\sigma_{Travelocity}$	σ_{Orbitz}	$\sigma_{Hotwire}$	$\sigma_{Other\ Portals}$	$\sigma_{All\ other\ sites}$
$\sigma_{Expedia}$	24.791					
$\sigma_{Travelocity}$	14.467	4.491				
σ_{Orbitz}	3.482	1.501	0.748			
$\sigma_{Hotwire}$	-13.557	27.685	10.859	18.358		
$\sigma_{Other\ Portals}$	-3.438	-6.649	-15.838	11.665	1.073	
$\sigma_{All\ other\ sites}$	4.104	-3.893	0.290	1.391	5.727	5.052

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

We find the greatest site choice heterogeneity for basket purchases associated with a single site to be on Expedia (24.791) indicating that either Expedia (i) attracts a diverse target audience or (ii) price discriminates amongst its audience better and makes the right basket offerings to the right consumer. The site choice heterogeneity was least pronounced for basket purchases on Orbitz (0.748) indicating profile of consumers making basket purchases on Orbitz to be very similar or those that knew exactly what to get and where on Orbitz. Other site combinations with large heterogeneity were Expedia and Travelocity (14.467) and Hotwire and Travelocity (27.685). Combinations with negative site heterogeneity that are worth mentioning are Expedia and Hotwire(-13.557) and Orbitz and Other travel portals (-15.838). Also note that the top three portals exhibit positive site heterogeneity amongst themselves while exhibiting negative site heterogeneity with other travel portals indicating that the consumers on the top three portals and those on other travel portals exhibit completely different behavior and could have a different target/segment profile.

6. Conclusions

In this paper we develop a two-stage model to study the category purchase propensities followed by propensity to purchase from a travel website. We model (i) the propensity to purchase a given basket by a consumer and, (ii) the choice of the website where consumers will make the purchases that constitute this basket and how these choices are inter related. We find significant effects of site preference, loyalty, prior browsing and demographic variables in determining consumer purchase behavior. We also find that the choice of the first site to where consumer makes a purchase has a significant impact on choice of the purchase site for other products in the basket indicating multi-category efficiencies.

Managers can use these results to identify the major determinants of consumer online behavior for basket level purchases and make appropriate marketing interventions based on this understanding of how consumers approach buying multiple products at the same time from multiple or same website. The correlations between the various travel products provide unique insights into travel habits of consumers in addition to providing bundling opportunities for service providers to better serve consumer needs. In this paper we also tease out consumer preferences in making multiple travel product purchases on a travel portal as opposed to pursuing this on separate service provider websites. This model also provides unique insights on how travel portals such as Expedia and Orbitz can better satisfy consumer needs by providing a basket of complementary products which involve air tickets purchase, car rental and/or hotel bookings and also explain why many airline sites have moved towards selling car rental and hotel products to successfully compete with travel portals. Managers can also use the insights from our demographic indicators to create a profile of target consumers who are more likely to make a purchase. Availability of more detailed demographics in the ComScore data set could aid managers in fine turning their segmentation strategy and develop more detailed target profiles.

References

Ainslie Andrew Ainslie and Peter E. Rossi, (1998), Similarities in Choice Behavior across Product Categories, *Marketing Science*, 17,(2),91-106

Baltas G., (2005), Modelling Category Demand in Retail Chains, *Journal of Personality and Social Psychology*, 56,(11), November,1258-1264

Bell David R. Bell and Randolph E. Bucklin, (1999), The Role of Internal Reference Points in the Category Purchase Decision, *The Journal of Consumer Research*, 26,(2), September,128-143

Bell David R. and James M. Latin, (1998), Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket" Shoppers Prefer EDLP, *Marketing Science*, 17,(1),66-88

Ben-Akiva Moshe and et al, (1999), Extended Framework for Modeling Choice Behavior, *Marketing Letters*, 10:3,187-203

Chung Jaihak and Vithala R. Rao, (2003), A General Choice Model for Bundles with Multiple-Category Products: Application to Market Segmentation and Optimal Pricing for Bundles, *Journal of Marketing Research*, 40,(2), May,115-130

Harlam Bari A. and Leonard M. Lodish, (1995), Modeling Consumers' Choices of Multiple Items, *Journal of Marketing Research*, 32,(4), November,404-418

Hui, Sam K., Eric T. Bradlow and Peter S. Fader, (2009), Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior, *The Journal of Consumer Research*, 36,October

Kau Ah Keng and A. S. C. Ehrenberg, (1984), Patterns of Store Choice, *Journal of Marketing Research*, 21,(4), November,399-409

Manchanda Puneet, Asim Ansari and Sunil Gupta, (1999), The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions, *Marketing Science*, 18,(2),95-114

Montgomery, Alan L., Shibo Li, Kannan Srinivasan and John C Liechty, (2004), Modeling Online Browsing and Path Analysis Using Clickstream Data, *Marketing Science forthcoming*, 23,(4), Fall,579–595

Narasimhan Chakravarthi, Scott A. Neslin and Subrata K. Sen, (1996), Promotional Elasticities and Category Characteristics, *The Journal of Marketing*, 60,(2), April,17-30

Olshavsky Richard W. and Donald H. Granbois, (1979), Consumer Decision Making-Fact or Fiction?, *The Journal of Consumer Research*, 6,(2), September,93-100

Roland T. Rust and Naveen Donthu, (1995), Capturing Geographically Localized Misspecification Error in Retail Store Choice Models, *Journal of Marketing Research*, 32,(1), February,103-110

Setharaman P.B. and et al, (2005), Models of Multi-Category choice behavior, *Marketing Letters*, 16,(3/4),239-254