The Impact of Leisure Travelers' Online Search and Purchase Behaviors on Promotion Effectiveness

Misuk Lee[*]
Mark E. Ferguson[†]
Laurie A. Garrow[‡]
David Post[§]

April 26, 2010

Abstract

We develop a predictive model of individuals' online search to purchase conversion behavior for the leisure travel industry. Our model explicitly considers the link between individuals' search and purchase behaviors. We incorporate both within-visit search effects (search depth and page view durations) as well as inter-visit search effects (the number and timing of repeat visits to the retailer's website). We develop a new continuous-time semi-Markov chain model that uses discrete choice models to predict transition probabilities and proportional hazard models to predict the timing of search and purchase decisions. Our model is flexible in that it allows for time-varying search and purchase dynamics as well as heterogeneity across individuals.

Our results show that higher discounts result in greater visit search depths as well as non-linear increases in revisit and purchase probabilities. From the perspective of evaluating promotion effectiveness, it is critically important to consider the timing of repeat visits, as inter-visit times reveal two distinct customer segments. Specifically, individuals who revisit the retail website within 24 hours are much more likely to purchase. Failure to account for revisit activities leads to dramatic underestimation of the value of a promotion, as purchasing probabilities are three times higher for revisit (or latent demand) visitors versus first-time visitors.

Key words: e-commerce; dynamic customer search; online purchasing conversion; promotion effectiveness; Markov chain models

---

[*] Misuk Lee, College of Management, Georgia Institute of Technology, Atlanta, GA 30332. e-mail: mlee@gatech.edu.

[†] Mark E. Ferguson, College of Management, Georgia Institute of Technology, Atlanta, GA 30332. e-mail: mark.ferguson@mgt.gatech.edu.

[‡] Laurie A. Garrow, School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, e-mail: laurie.garrow@ce.gatech.edu

[§] David Post, SigmaZen GmbH, Bommersheimerweg 31, 61348 Bad Homburg, Germany, e-mail: david.post@sigmazen.de.

1 Introduction

Evaluating promotion effectiveness is a challenging problem. As noted by Barry Smith, former Chief Scientist and Corporate Officer for Sabre Holdings®, "one of the main barriers to accurately evaluating the total demand associated with e-commerce promotions has been the inability and/or the prohibitively high time and cost involved in tracking individual consumers over time" (2010). Without this information, it is difficult to comprehensively quantify all revenue impacts associated with a promotion. As an example, consider an e-mail campaign. Whereas it is usually straight-forward to measure the direct demand, or the number of individuals who entered the retail website by clicking on the email promotion link, it is more difficult to measure the latent demand, or those individuals who saw the advertisement and entered (or perhaps re-entered) the retail website through a non-promotional link. Kevin Geraghty, Vice President of Reporting and Analytics for 360i®, elaborates on the source of this difficulty, noting that "while many firms are using cookies to track repeat customers, the shear volume of search and purchase data results in many firms using simple attribution rules to measure the value of a promotion" (2010).

Due to the underlying data processing challenges associated with tracking customers, there is a research need to develop prediction models for promotions that incorporate latent demand effects; in turn, these prediction models can help managers decide whether or not to run a promotion. In the leisure travel industry, where promotions are common but the cost of promotions are high (because of the high ticket price) and the margins are low, any improvement in measuring the dollar impact of a promotion is particularly important. Conversion rate models that consider only direct demand or use simple attribution rules may be dramatically underestimating a promotion's impact. Further, failure to consider *when* latent demand occurs may lead to lost revenue opportunities due to early termination of a promotion.

The objective of our study is demonstrate the importance of, as well as to investigate the ability for, incorporating latent demand (as revealed through repeat visits to a retail website) into a predictive online conversion model. The resulting predictive online conversion model can then be used to measure the total value of an online promotion without the need for tediously tracking the latent demand, and to help estimate the true value of a potential promotion prior to its implementation. We investigate this question in the context of online leisure travel, using search and purchase data from an airline based in New Zealand. These data are unique in three distinct ways, which makes them particularly well suited to investigate our research questions.

First, the airline offered different promotions (which we will refer to interchangeably as discount rates or relative discount rates) to travelers who were willing to accept uncertainty in their travel dates. However, because the airline was very concerned about whether their new opaque product was cannibalizing revenue, they conducted follow-up customer surveys. Through these surveys, the airline determined that more than 70% of all revenue generated was incremental and the vast majority was for leisure travel (Post 2006). This allows us to to clearly separate and capture leisure travel demand from business travel demand. Since promotions are mostly targeted towards leisure travel, it was important for us to isolate this segment for our research.

Second, the airline tracked both the IP addresses of individuals visiting their site as well as detailed click-stream data of every page that each visitor viewed. This provides us with the ability to examine the link between individuals' search and purchase behaviors as well as to investigate how conversion rates are influenced by both within-visit characteristics as well as inter-visit characteristics.

Third, due to the fact that airline analysts were cautious about setting discount rates too high, the underlying pricing model used to generate discounts effectively resulted in the airline *offering discounts that were uncorrelated with the customers' degree of travel flexibility.* While clearly this situation is undesirable if one wants to determine an optimal pricing model as a function of customers' flexibility characteristics, this situation is incredibly fortuitous in our research context. The low correlation between offered price discounts and travelers' flexibility characteristics reduces potential selection bias and allows us to make stronger causal inferences about how discount levels (or different promotions) result in higher search, purchase, and revisit activities.

The remainder of this paper is organized as follows. First, we present a conceptual model of the factors we hypothesize influence individuals' search and purchase behaviors; this conceptual model is also used to describe how our work relates to and contributes to the existing literature. Next, we describe our data and the continuous-time semi-Markov chain model that we developed to investigate our research questions. Section 5 and 6 present results and model validation, respectively. Section 7 investigates the impact of our findings on the evaluation and design of online promotions. We conclude with describing managerial implications of findings, particularly as they relate to the design and evaluation of promotions.

2   Conceptual Model and Literature Review

At a fundamental level, an investigation into whether it is important to incorporate latent demand into the design and evaluation of promotions involves understanding the linkages between individuals' search and purchase behaviors – both within a single visit to the retailer's site as well as across multiple visits to the retailer's site. Figure 1 portrays the *activities* that we incorporate in our purchase conversion model. The first activity, "search," occurs when an individual enters information at the promotional retailer's site and receives a discount offer based on this information. After receiving the discount offer, the individual may decide to "purchase" the offer, continue "searching" by modifying the input information to receive a new discount offer, or "leave" the promotional retail site (and thus terminate the current visit). Those customers who leave may never return to the promotional retail site within a six-month period (represented by the "no purchase" activity) or may decide to initiate a new visit (reflected by a return to the "search" activity).

Figure 2 illustrates how we incorporate these search activities in our purchase conversion model. Each graph in Figure 2 represents a customer who visits this retail website one or more times. Customer 1 visits the website once and views three distinct pages (resulting in two page-view times) before leaving the site without purchase. The amount of time the customer spends on page
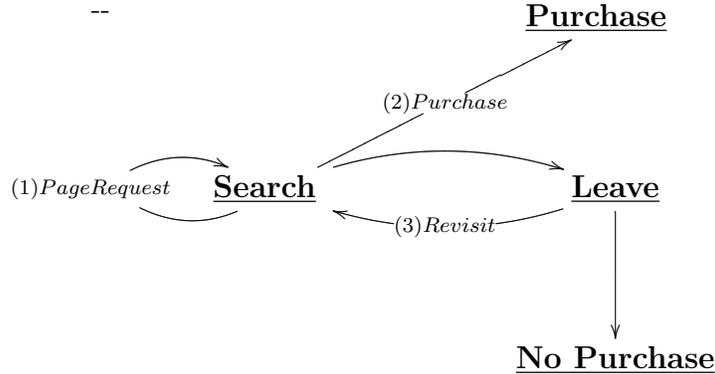
Figure 1: Within-Visit and Across-Visit Search and Purchase Activities

j during visit i is given by ν_{ij}. The cumulative amount of time the customer has spent viewing the first j pages during visit i is given by ω_{ij}. In contrast to Customer 1, Customer 2 shows more engaged behavior. On the first visit, she views four pages and on the second visit she views three pages and leaves the second visit without purchasing.

If we assume that Customer 2 was randomly presented with a higher discount offer, then our first set of fundamental research questions can be expressed as follows: Are higher discount offers associated with the different search and purchase *activities* portrayed by Customers 1 and 2? That is, are higher discount rates more likely to result in purchases? Are higher discount rates more likely to result in increased search activity within a particular visit (as reflected by the depth of search, or higher probabilities of requesting additional pages)? Are higher discount rates more likely to result in increased search activity across visits (as reflected by higher revisiting probabilities)?

So far, we have used Figure 2 to visualize how discount rates and prior search activities within and across visits may influence subsequent search, purchase, and revisit activities. We can also use Figure 2 to visualize how the durations of these activities (specifically, page view durations and inter-visit durations) may vary with time. With respect to page view durations, we hypothesize that page view durations decrease as customers become more adept and learn how to navigate the site. These learning effects may appear within a visit (e.g., see Customer 1) as well as across visits (e.g., see Customer 2), although we expect within-visit learning effects to be stronger than across-visit learning effects.

The importance of inter-visit durations on future search and purchase activities can be seen by comparing Customers 2 and 3. In this case, one of the key difference in the timing of their search patterns is when their second revisit activity to the retail website occurred. The inter-arrival time between visit 1 and 2, given by τ_1, is shorter for Customer 3, leading to the research question: Can the timing between visits be used to segment consumers? In particular, are individuals who revisit the promotional retailer's site within 24 hours (Customer 3) more likely to purchase than those who have longer revisit times (Customer 2)?

Figure 2 can also be used to visualize why it is important to link search and purchase behaviors
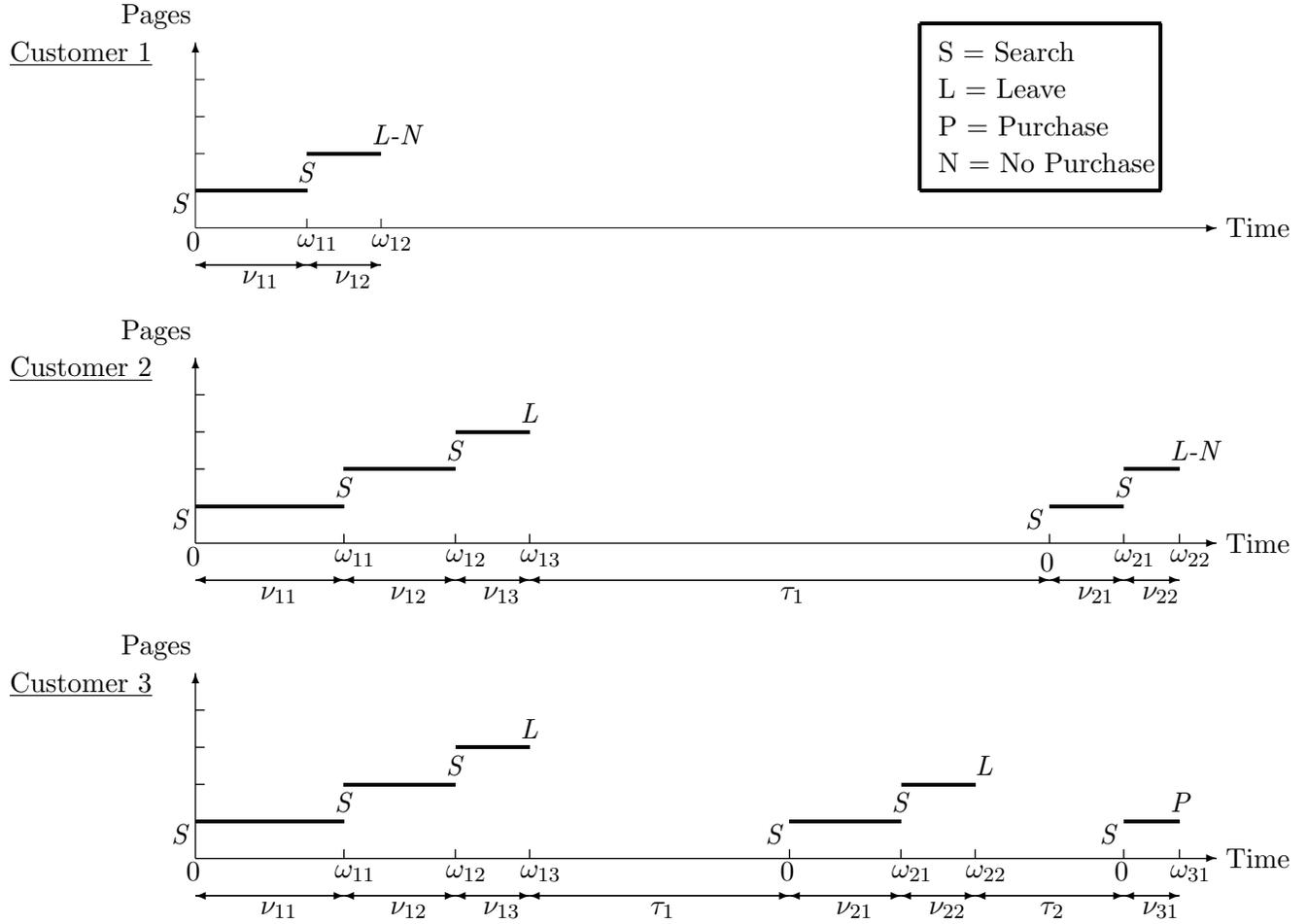
Figure 2: Three examples of customer search and purchase behaviors

in the context of promotions. First, assume that customers were not tracked, and that searches and purchases that occurred after their first visit were not associated with the promotional link. In this case, Customer 3's purchase would be "latent" demand and would not be captured as part of the promotional evaluation metrics. Finally if the dynamics of search and purchase differ across discount rates, it may be important to consider these dynamics when deciding the length of time to run a promotion. In Figure 2, Customer 3 has a longer decision purchase cycle, or a longer amount of time between when the individual first visits the site and eventually decides to purchase. Failure to consider how the dynamics of search and purchase change may subsequently result in early termination of a promotion, i.e., lost revenue.

By framing our study as the investigation of the link between search and purchase behavior at a single retail site (and subsequently quantifying the importance of incorporating this link in the

context of promotion evaluations), we can describe how our study relates to and contributes to the existing literature. Table 1 summarizes our key research hypotheses. We classify our research hypotheses along two key dimensions. The first dimension distinguishes between (1) discount factors and prior search activity factors that influence the *probabilities* of future search and purchase activities; and, (2) duration effects that may influence not only the *probabilities* of future search and purchase activities, but also provide additional behavioral insights related to *learning* and *customer segmentation.* The second dimension distinguishes between whether these factors are measured within a single visit or across multiple visits.

One of the key factors shown in Table 1 that we expect to influence search and purchase probabilities is price. Conceptually, we expect that price will have a large effect on search and purchase behaviors and that higher discounts will lead to more search and purchase activities. Importantly, if revisit and purchase activities increase non-linearly with discount rates, traditional promotional evaluation metrics based on simple attribution rules may be misleading.

Whereas few (if any) studies have investigate the effect of discount rates on search and purchase behaviors, many studies have investigated how individuals' search activities and patterns (either within a visit or across multiple visits) impact conversions (e.g., see Moe and Fader 2004; Chatterjee, Hoffman and Novak 2003; Montgomery et al. 2004). We build upon this prior research by simultaneously investigating the impacts of the number of previous page views, number of prior visits, page view durations, and inter-visit durations on search and purchase probabilities. To the best of our knowledge, our study represents the first to comprehensively examine all of these factors.

An investigation of page view durations and inter-arrival durations is particularly interesting, as the timing of when additional page views are requested and/or when revists occur may provide evidence of learning effects (in the case of page view durations) or consumer browsing intentions (in the case of inter-arrival times between visits). Several researchers have examined whether page view durations and/or total visit durations decrease as individuals become more adept at navigating the retailer's website (e.g., see Johnson, Bellman and Lohse 2003; Bucklin and Sismerio 2003; Johnson et al. 2004). We hypothesize that page view times will decrease within (and possibly across) visits, reflecting learning behavior as individuals become more comfortable with interacting with the retailer's website.

Inter-arrival times may provide insights into individuals' different browsing objectives (which in turn, reflect different customer segments). Several studies have examined the different objectives of shoppers, which may include simply browsing, to gathering information about a product, to shopping with the intention of purchasing, etc. (e.g., see Hoffman and Novak 1996; Hoffman et al. 2003; Moe 2003; Wolfinbarger and Gilly 2001). In our research context, we hypothesize that those individuals who are simply browsing (with little intention of buying) will have longer durations between visits. In particular, we hypothesize that those individuals who revisit the site within 24 hours are more likely to purchase.

The second classification in Table 1 is used to distinguish between whether searching and/or purchasing behaviors are examined within a single visit or across multiple visits. This classification

5

Effects of Price and Prior Searches on Search and Purchase Behaviors	
Price	Are higher discount rates associated with increased probabilities of search (page view requests, revisits) and purchase?
Prior search activities	
Within-visit	Are deeper searches (a larger number of previous page views and long page view durations) associated with increased probabilities of search (page view requests, revisits) and purchase?
Across-visits	Are a larger number of previous visits and short inter-visit durations associated with increased probabilities of revisits and purchases?
Duration Effects on Search and Purchase Behaviors	
Within-visit	Do page view durations decrease as the number of page views increase?
Across-visit	Do visitors with short inter-arrival times represent a distinct customer segment from visitors with long inter-arrival times?

Table 1: Summary of key research hypotheses

is particularly useful for relating our work to prior studies. First, it important to note that there are very few studies that explicitly link online search and purchase behaviors. Three examples of studies that have examined search behavior at a single online retail site include those by Johnson, et al. (2003), Bucklin and Sismeiro (2003), and Moe (2003). The first by Johnson, et al. (2003) found that cumulative visit duration at a single retail site decreased over time. However, their study did not explicitly incorporate intra-visit characteristics (such as individual page view times). The second study by Bucklin and Sismeiro (2003) offers an explanation for these reduced cumulative visit durations. Specifically, Bucklin and Sismeiro developed a model of web site browsing behavior and found that repeat visits were associated with fewer page view requests (but not reduced page view durations). Moe (2003) pursued this question further by developing a model of web site browsing behavior that used page content information to capture customers' underlying objectives (i.e., buying, browsing, searching or knowledge-building).

Additional studies examining search behavior across multiple online retail sites include those by Park and Fader (2003) and Johnson, et al. (2004). Park and Fader (2003) examined how consumer behavior at one retailer's site influenced consumer behavior at a second retailer's site (as captured by if and when the consumer visited this second site). Johnson, et al. (2004) examind search across multiple e-commerce sites to investigate whether lower search costs resulted in more search. Similar to our study, Johnson, et al. (2004) considered both the depth and dynamics of search; however, their study did not incorporate intra-visit characteristics (such as individual page view times) and, as noted earlier, does not explicitly link search and purchase behaviors.

6

Whereas it has been more common in the literature to examine search behavior, a few studies have jointly examined search and purchase behavior (typically in the context of a single retail site). In some studies, such as that by Montgomery, et al. (2004), the goal was to understand how to influence page-to-page transitions via interactive marketing to increase purchases. A second study by Chatterjee, et al. (2003) jointly examined search and conversion behaviors in the context of banner ads; the goal in this study was to understand how conversion rates were influenced by the number of banner ads shown, how frequently the ads were shown, when in the search path they were shown, etc. Finally, Moe and Fader (2004) investigated how conversion rates at Amazon.com were influenced by customers' underlying objectives. Although their study investigated the dynamics of search, they did not incorporate detailed page view times.

Key Research Question	Possible Implication if True	Possible Implication if False
As discount rates ↑ do search and purchase ↑?	Promotion design influences both search patterns and purchases, potentially affecting conversion rates in non-linear ways.	Price-sensitive customers are more likely to search and find higher discounts.
Do deeper searches lead to ↑ search and purchase?	Retailers can predict path-level conversions and interactively market to consumers.	Number of pages and page view durations may be a function of different consumer intentions and make conversion prediction (without page content) difficult.
As the number of visits ↑ and/or inter-visit times ↓, do revisits and purchases ↑?	Short inter-visit durations may enable retailer to run shorter promotions. Increases in revisits may be captured as latent demand and not attributed to promotion.	Large number of purchases from first-time visitors implies current promotion evaluation metrics are likely accurate.
Do page view durations ↓ as the number of pages ↑?	Shorter page view durations reflect learning effects and/or time constraints.	Longer page view durations reflect different customer intentions, with longer page views being associated with more engaged customers.
Do short inter-arrival times represent a distinct customer segment?	Short inter-visit durations may enable retailer to run shorter promotions.	The amount of time required to run a promotion may be large (driving up the cost).

Table 2: Potential implications of research findings on promotion design and evaluation

To the best of knowledge, no previous studies that have directly examined the link between

7

search and purchase behaviors have done so in the context of promotion evaluation and design. Table 2 summarizes potential implications of what our research findings mean in the context of promotion design and evaluation. Different implications are possible, depending on whether the relationships reflected in our research questions are true or false. The implications summarized in Table 2 are not meant to be exhaustive but rather illustrative of how the findings from our particular study can influence promotion evaluation and design. We will return to these possible implications in our conclusions.

In summary, our paper contributes to the literature by directly examining the link between search and purchase behaviors at a single retailer's website in the context of promotion design and evaluation. To the best of our knowledge, our study is the the only one to comprehensively consider the impacts of both within-visit search characteristics as well as inter-visit search characteristics on purchase. From a methodological perspective, we develop a new continuous-time semi-Markov chain model that simultaneously considers these two impacts. Importantly, our methodology is distinct from all of the prior studies noted above in that we use discrete choice models versus statistical distributions to predict the transition probabilities among search, purchase, and revisit activities. The use of discrete choice models enables us to gain a richer understanding of the relative influence of multiple factors on individuals' search and purchase behaviors. The use of timing models enables us to gain additional insights related to learning effects and consumer heterogeneity.

3   Data and Key Modeling Assumptions

Our clickstream data is based on a website promotion offered by an airline that operated routes in Australia, Fiji, and New Zealand. The offered discount level was (in theory) dynamically customized to each individual based on two main factors: (1) how much uncertainty customers were willing to accept in their travel dates; and, (2) how much demand was present for the desired set of possible travel dates. A simple webpage was used to obtain trip information (i.e., destination, the number of travelers, how many nights at the destination) and traveler flexibility information (the earliest possible departure date, the latest possible return date, and prior notice). Prior notice is defined as how many days before the earliest departure date the traveler wants to be informed of the travel details. From the individual's perspective, the promotion can be described as follows: "I want to take a week of vacation and spend six nights in Fiji. I could take this trip any time in March so my earliest possible departure date is March 1 and my latest possible return date is March 31. To arrange for time off work, I would need to know my travel dates at least 14 days in advance of flying." In this example, the individual's flexibility is represented by the number of possible travel date combinations (e.g., taking a one week vacation within a two month period offers more flexibility than a one week vacation within a one month period), and how many days in advance the individual needs before traveling (thus, an individual who can travel with as little as one week prior notice is more flexible than an individual who needs one month of prior notice before traveling).

The promotion was designed so that those customers who were more flexible would receive

higher discounts. Discount rates were determined by first calculating the lowest non-promotional fare the individual could receive that satisfied her desired travel conditions. The promotional price that was subsequently shown to the individual reflected a discount rate *relative* to this lowest non-promotional fare.

From a modeling perspective, it is important to emphasize two key points related to the use of a relative discount rate. First, as noted earlier, airline analysts were cautious about setting discount rates too high, and as a consequence the underlying pricing model used to generate discounts effectively resulted in the airline offering discounts that were uncorrelated with the customers' degree of travel flexibility. The important modeling implication is that it is not possible to explicitly examine the relationship between customers' flexibility attributes and their search and purchase behaviors. However, the effective "random" assignment of discount levels to customers does enables us to investigate a question that is more relevant in the context of promotions, namely whether there is a relationship between the offered relative discount rates and customers' search and purchase behaviors. This leads to our second key point, namely that from a modeling perspective, the use of a relative discount rate to initialize a Markov chain model can be viewed as appropriate based on prior modeling conventions used by the airline industry. That is, while it can be argued that customers may not know the relative discount rate, consumer surveys and work by Travelocity[®] published as part of the 2006 Frankz Edelman Award Competition establish a precedence for using relative discount rates to model airline customers' online search and purchase behaviors. More specifically, according to PhoCusWright and Forrester Research, approximately 60% of online leisure travelers purchase the lowest fare they can find, suggesting that customers are aware of market prices (Harteveldt 2004, PhoCusWright 2004). Likewise, Travelocity[®] describes how they use relative discount levels (defined as the lowest fare offered over a rolling time-horizon) to trigger e-mail campaigns as part of their "Best Day to Buy" campaigns (Smith, et al. 2007). Consequently, our work follows prior airline industry conventions by using relative discount rates to initialize our Markov chain models.

Before describing this Markov chain model, it is important to precisely define how we have processed our online clickstream data and used it for analysis. The online clickstream data for this research was collected over a two-year period, from August 2004 to June 2006, and targeted to customers traveling on one of 21 routes serving New Zealand, Australia, and/or Fiji. The majority of prices shown to customers were between NZ \$300 and \$500, where \$500 was approximately equal to the lowest price offered by the retailer for a traditional round trip itinerary. The average discount rate offered was 47.8% and the standard deviation of all discount offers was 29.3%.

Online clickstream data captures the sequence of web pages viewed by customers. We define four terms to describe how clickstream data was used for the analysis. A *search* is defined as the combination of two pages: the page in which a consumer inputs search information and the page in which the price offered to the consumer is displayed. A *visit* is a sequence of pages that an individual requests within a specific time period to search for a single product (defined as an airline route).

An alternative concept of a visit which has often been used in prior analysis of clickstream data is a *session*. A session is an identity tag that is assigned every time a user launches a new browser. However, in our data, we discovered that many individuals launched several browsers almost simultaneously and searched back and forth over multiple browsers. In order to eliminate multiple browser effects, we do not use the concept of a session for modeling, but rather define a visit as a sequence of searches done within a specific time period regardless of the number of browsers launched. We follow previous modeling conventions (e.g., see Bucklin and Sismeiro (2003) and Moe and Fader (2004)) and assume that a page request started a new visit if it was requested after an idle period of at least 30 minutes.

Numerous authors including Bucklin and Sismeiro (2003), Chatterjee, et al. (2003), Johnson, et al. (2004), and Moe and Fader (2004), point out that when modeling individuals' online search and purchase behaviors, it is important to consider the dynamics of consumer searches across sessions (or visits as we have defined for our application). Indeed, the *purchase decision cycle* for an individual, or the period of time during which she searches prior to making a purchase or no purchase decision, may span days, weeks, or even months over several repeated visits. Further, we expect that the purchase decision cycle for high-priced goods (including airline tickets) may be longer and thus more important to model than for lower-priced goods. We model individuals' searches throughout the entire purchase decision cycle (search-exit-revisit-purchase), which we assume spans up to six months. Our data indicate that this is a reasonable assumption, as more than 90% of purchase cycles fall within one month.

The final definition that was used to process our data relates to *cookies*. A cookie is used as a proxy for an individual consumer. As noted by Moe and Fader (2004), one limitation typically associated with studies that used clickstream data collected during the early- to mid- 2000's is that it was difficult to obtain characteristics that could be used to identify a particular user. However, in contrast to the U.S. where retailers were initially reticent to trace cookies stored in customers' computers due to privacy concerns, New Zealand retailers were much more open to tracking cookies. This provides us with the opportunity to use cookies to identify and track individual customers.

Table 3 presents summary statistics for the clickstream data based on the data processing assumptions and definitions noted above. Most customers visit the site only once. On average, an individual searches three times throughout all visits. The purchase conversion rate per decision cycle is approximately 3%. However, customers that ended up purchasing searched, on average, much more than three times.

4 Methodology

In this section, we develop a continuous-time semi-Markov chain that uses discrete choice models to predict transition probabilities and proportional hazard models to predict the timing of search and purchase decisions. Our model is flexible in that it allows for time-varying search and purchase dynamics, as well as customer heterogeneity. Our methodology is developed in several sections.

Data View	Frequency
Searches	42,554
Visits	20,354
Purchase Cycles	15,237
Cookies	12,588
Purchases	474

Table 3: Summary statistics for clickstream data

First, we describe how we use dynamic discrete choice models to represent *transition probabilities* for the state spaces. These dynamic discrete choice models are subsequently combined with Poisson models to represent *transition rates*. Next, we apply this general dynamic choice model to our specific problem context which relates transition probabilities according to the visit number, number of page view requests, and navigation state. We then use proportional hazard models to represent the timing of search and purchase decisions for page views and the inter-arrival times between visits. We incorporate two separate baseline hazard models into our revisiting formulation, as the probability of purchase is found to be highly correlated with *when* a revisit activity occurs. The mixture of these baseline functions captures the observation that those customers who revisit the promotional site within 24 hours exhibit much higher purchase probabilities.

4.1 Development of a Dynamic Choice Model

We model individuals' search and purchase behaviors using a semi-continuous Markov chain model. Transition probabilities among the state spaces are derived based on discrete choice models. Formally, let S be the set of possible states in the Markov chain. Assume that there is a latent utility associated with each state which is given as:

$$U_{ij} = \Gamma_i X_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim G(0,1)$$

where U_{ij} is the utility that an individual in state i would receive if she selects j as the next state, X_{ij} is a vector of independent variables associated with state i,j, and Γ_i is a vector of parameters to be estimated. Assuming ϵ_{ij} is iid Gumbel distributed, the one-step *transition probability* from state i to j is given by a multinomial logit (MNL) model (McFadden 1978):

$$P_{ij} = \frac{e^{\Gamma_i X_{ij}}}{\sum_{m \in A_i} e^{\Gamma_i X_{im}}}$$

where A_i is the set of possible next states from state i. The *transition rate* from state i to state j, denoted as $\lambda_{ij} = \lambda_i P_{ij}$, is simply the product of a Poisson rate out of each state, λ_i, and the one-step transition probability, P_{ij}. That is, intuitively, since Poisson rates can be multiplied by constants without affecting the underlying assumptions needed for a Markov chain model, we can combine discrete choice and Poisson models to capture transition rates. By doing so, we are

11

able to obtain richer behavioral interpretations for the underlying search characteristics that are influencing search and purchase probabilities; these behavioral insights are typically more difficult (if not impossible) to uncover when using one or more statistical distributions.

Thus far, we have assumed that transition rates are time-homogeneous. We can relax this assumption to incorporate time-dependent transition probabilities. Formally, we extend our model to a semi-Markov chain by assuming that the underling latent utility becomes a function of time t:

$$U_{ij}(t) = \Gamma_i X_{ij} + \alpha_{ij} t + \epsilon_{ij}, \quad \epsilon_{ij} \sim G(0,1)$$

where α_{ij} is the weight parameter associated with time t. Conditional on a transition being made at time t, the one-step transition probability is given as:

$$P_{ij}(t) = \frac{e^{\Gamma_i X_{ij} + \alpha_{ij} t}}{\sum_{k \in A_i} e^{\Gamma_i X_{ik} + \alpha_{ik} t}}.$$

Note that this conditional probability sums to one over all possible transitions. That is, $\sum_{k \in A_i} P_{ij}(t) = 1$. We assume that the rate at which the process leaves state i is constant, i.e. $\lambda_i = \sum_{j \in A_i} \lambda_{ij}(t) = \sum_{j \in A_i} \lambda_i P_{ij}(t)$.

The transition probability from state i to state j now depends on how long an individual has spent in state i. We define $q_{kij}(t)$ as the probability that decision maker k moves into state j after spending less than time t in state i. This probability can be expressed as:

$$q_{kij}(t) = \int_0^t \frac{e^{\Gamma_i X_{ij} + \alpha_{ij} s}}{\sum_{m \in A_i} e^{\Gamma_i X_{im} + \alpha_{im} s}} \lambda_i e^{-\lambda_i s} ds. \tag{1}$$

The derivation of this probability and its associated likelihood function are contained in Appendix 1.

4.2   Dynamic Choice Model Application to Search and Purchase Behavior

We can apply our dynamic choice model to investigate individuals' search and purchase behaviors by relating states i and j to an individual's navigation state as well as the visit number and number of pages. First, we assume that consumer k has latent utility U_{knva} in state a on discrete viewing occasion n of visit v, where there are totals of K consumers, A states, V_k visits of consumer k, and N_{kv} viewings for the vth visit of consumer k. Then we formulate the utility as:

$$U_{kvn} = \Gamma_{vn} X_{k,v,n-1} + \epsilon_{kvn}, \quad \epsilon_{kvn} \sim G(0,1) \tag{2}$$

where $U_{kvn} = (U_{kvnS}, U_{kvnP}, U_{kvnL}, U_{kvnN})$ is a vector of the utilities associated with the four navigation states shown in Figure 1 (Search, Leave, Purchase, and No Purchase), and Γ_{vn} is a vector of parameter estimates associated with X_{kvn-1}, a vector of independent variables denoted as:

$$X_{kv,n-1} = (1, d_{kv,n-1}, \tau_{kv,n-1}, \nu_{kv,n-1}, \omega_{kv,n-1}, v, n-1)'$$

where $d_{kv,n-1}$ is the discount rate, $\tau_{kv,n-1}$ is the inter-arrival time for resuming search (revisit time), $\nu_{kv,n-1}$ is the inter-arrival time for a page-to-page transition, and $\omega_{kv,n-1}$ is the cumulative page view time of visit v. The cumulative page view time is found as:

$$\omega_{kvn} \quad = \quad \omega_{kv,n-1} + \nu_{kvn}.$$

Note that U is a function of factors that vary with time, i.e., we model individuals' decisions using a semi-Markov chain.

Using maximum utility theory, we associate the individual's selection on the next state with the latent utilities as follows:

$$Y_{kvna} \quad = \quad \begin{cases} 1 & \text{if } U_{kvna} = max_b\{U_{kvnb}I(a(k,v,n-1),b),0\} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $a(k,v,n) \in \{S,P,L,N\}$ is the state of individual k on page n of visit v, i.e. $a(k,v,n) = b \in A$ such that $Y_{kvnb} = 1$. Matrix $I(i,j)$ indicates that the corresponding transition from i to j is allowed. $I(a(k,v,n-1),b)$ indicates whether the transition from current state $a(k,v,n-1)$ to state b is possible. For example, the transition from "Search" to "Search" is allowed, i.e. $I(S,S) = 1$, but the direct transition from "Search" to "No Purchase" is not allowed, i.e. $I(S,N) = 0$.

We define individuals' utilities U_{kvn} as linear-in-parameter functions of search characteristics $X_{k,v,n-1}$. Assuming iid Gumbel error terms, we model the consumer choices Y_{kvna} with a series of time-dependent multinomial logit models. By fitting Y_{kvna} on $X_{k,v,n-1}$, we estimate the effects of individuals' search characteristics on their search and purchase behaviors.

4.3 Page View and Inter-Visit Times

In addition to deriving transition probabilities, we need to add a timing model to represent the influence of page view and inter-visit times on search behavior. To accomplish this, it will be convenient to express our model in more explicit Markov model notation. Note that individuals' decisions, Y_{kvn}, are based on the current state $a(k,v,n-1)$ and underlying utilities U_{kvn}. Individuals' utilities, U_{kvn}, are determined by the current search, $X_{kv,n-1}$. Consequently, the search result $X_{kv,n-1}$ based on the current state $a(k,v,n-1)$ leads to the next decision Y_{kvn}, which, in fact, indicates the next state $a(k,v,n)$.

If we redefine an individual's search state as $(X_{kv,n}, a(k,v,n))$, the next state depends only on the present state and not on previous states. Thus, we may define search and purchase decisions using Markov chains. To see this, let us decompose $(X_{kv,n}, a(k,v,n))$. The present search result $X_{kv,n-1}$ is characterized by a discount rate, search depth (pages, visits), and time between searches (intra- and inter- visit durations). Search depth and $a(k,v,n)$ define the state where individual k is on the discrete occasions. The intra- and inter- visit durations characterize inter-arrival times of individuals' decision transitions.

Specifically, let $M_k(t,d_k)$ denote a Markov chain corresponding to individual k's state at time t parameterized by discount rate d_k. As other continuous Markov chains, $M_k(t,d_k)$ is characterized by the (1) state space; (2) transition probabilities; and, (3) inter-arrival times.

4.3.1 Page-to-page Transitions

The discrete state of our Markov chain is search depth and search state, i.e., (v, n, a). As shown in Figure 3, transitions between states are derived from three sequential binary decisions: (1) whether to request an additional page or to exit the site; (2) conditional on existing, whether to purchase or leave once ending the search; and, (3) conditional on leaving, whether to revisit after ending the search or never purchase. For example, if an individual requests an additional page, the number of page views increases by one with the state unchanged ("Search"). When an individual comes back later, the number of visits increases by one and she transits from the "Leave" state to the "Search" state. Formally, the possible transitions are given as:

$$
\begin{aligned}
\text{Page request} &: (v, n, S) \to (v, n+1, S) \\
\text{Purchase} &: (v, n, S) \to (v, n, P) \\
\text{Exit without purchase} &: (v, n, S) \to (v, n, L) \\
\text{No purchase} &: (v, n, L) \to (v, n, N) \\
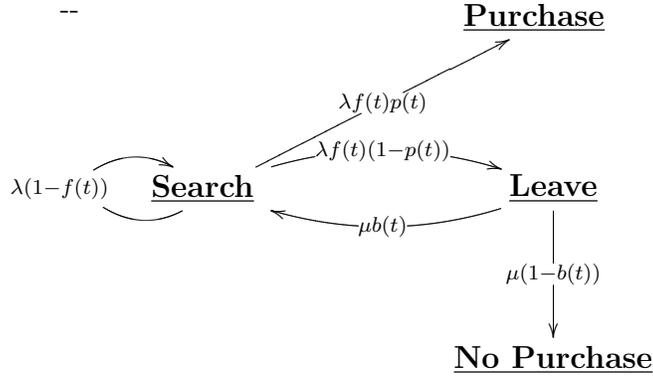\text{Revisit} &: (v, n, L) \to (v+1, 1, S).
\end{aligned}
$$

Figure 3: Page-to-page transitions and transition rates

The transition probabilities can be derived from Equation 1. However, note that each transition probability has a specific meaning depending on the state and can be defined as:

- $f(t)$: probability of ending search

- $p(t)$: probability of purchase conditional on ending search

- $b(t)$: probability of revisiting conditional on leaving.

Using these definitions, we can express page-to-page transition probabilities (illustrated on

Figure 3) as:

$$
\begin{aligned}
P\{M(t) = (v, n+1, S)|M(0) = (v, n, S)\} &= 1 - f(t) \\
P\{M(t) = (v, n+1, P)|M(0) = (v, n, S)\} &= f(t)p(t) \\
P\{M(t) = (v, n, L)|M(0) = (v, n, S)\} &= f(t)(1 - p(t)) \\
P\{M(t) = (v, n, N)|M(0) = (v, n, L)\} &= 1 - b(t) \\
P\{M(t) = (v+1, 1, S)|M(0) = (v, n, L)\} &= b(t).
\end{aligned}
$$

4.3.2   Time Durations

There are two types of time durations between transitions. Page view times, ν_{kvn}, correspond to transitions from the "Search" state. Inter-arrival times for resuming search (or revisit times), τ_{kvn}, are associated with the "Leave" state.

Figure 4 shows the distribution of page view durations. Note that the majority of individuals spend less than five seconds viewing each page and that the frequency of page view times decreases non-linearly as page view times increase. We model the conditional page view times with state-dependent heterogeneous exponential distributions.
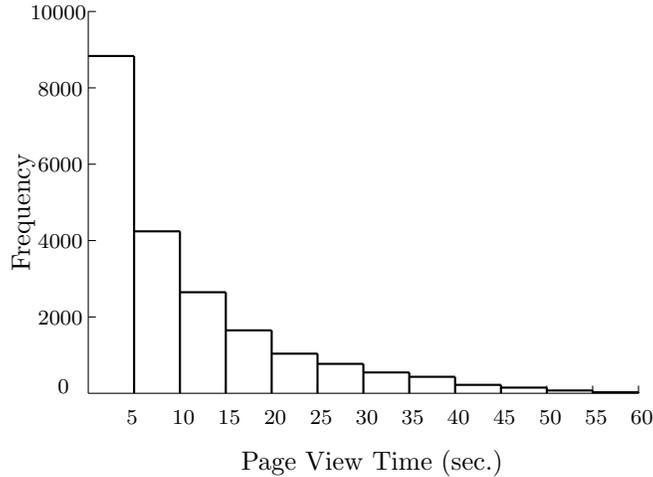

Figure 4: Distribution of page view durations

We define the probability that an individual spends more than time t at each page as:

$$
P\{M(t + s) = (v, n, S)|M(s) = (v, n, S)\} = e^{-\lambda t}.
$$

We compare this model with one that allows the inter-arrival rate λ_{vn} depend on the number of page views n and visits v. That is, we extend this model by allowing the inter-arrival rate λ to depend on the number of page views n and visits v, i.e., λ_{vn}. This enables us to investigate the effects of search depth on page view durations. Specifically, we explore whether page view durations decrease within or across visits – a phenomena that would support the presence of learning effects.

Revisit Times

The second type of time durations between transitions involves revisit times, or the time until an individual decides to come back after leaving the retailer's website. Figure 5 implies that we may model the conditional inter-arrival time for consumer k to revisit using exponential distributions. As in the case of page view time, we first estimate a model that assumes the inter-arrival rate for revisits is constant μ.
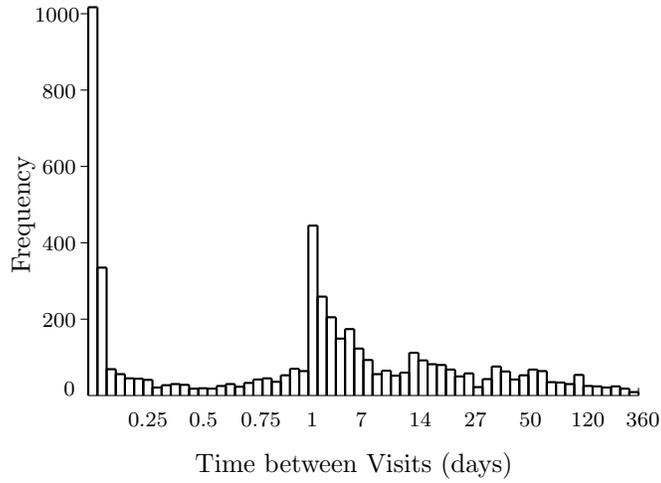

Figure 5: Distribution of inter-visit durations

However, Figure 5 clearly shows that inter-visit durations are very disperse, ranging from less than one hour to six months. We therefore extend the model to allow heterogeneous arrival rates across consumers. Formally, we assume that μ is Gamma-distributed with parameters θ and δ. Then, the distribution of any inter-arrival time τ is:

$$\begin{aligned} P\{\tau < t | \mu\} &= 1 - e^{-\mu t} \\ P\{\tau < t\} &= \int_{-\infty}^{\infty} \left(1 - e^{-\mu t}\right) dF(\mu) \\ &= 1 - \left(t/(t+\theta)\right)^{-\delta}. \end{aligned}$$

However, individuals with short inter-visit durations may have distinct motivations from individuals with long inter-visit durations. Some may be highly goal-oriented in the sense that they are revisiting the site with the intention of making an immediate purchase. Others may be building knowledge about the product. Several researchers including Chatterjee, et al. (2003), Hoffman and Novak (1996), and Moe (2003) discuss two distinct categories of consumption behavior: goal-oriented (directed) and experiential (exploratory) consumer behaviors. They suggest that consumers who revisit after relatively short durations are likely to be more goal-oriented than consumers who revisit after relatively longer intervals. Importantly, goal-oriented consumer behaviors are more likely to result in immediate purchases.

16

To investigate the presence of distinct customer segments, we assume two consumer populations: one that has short inter-visit times and a second that has long inter-visit times. Specifically, we assume that with probability α, consumer k is a short-duration type, whose arrival rate is Gamma distributed with parameters θ_s and δ_s. Other consumers have a longer revisit time, of which the arrival rate is Gamma distributed with θ_l and δ_l:

$$\mu \sim \begin{cases} Gamma(\theta_s, \delta_s) & \text{w.p. } \alpha_s \\ Gamma(\theta_l, \delta_l) & \text{w.p. } 1 - \alpha_s. \end{cases}$$

Then, the distribution of any inter-arrival time follows as:

$$P\{\tau < t\} = \alpha_s \left(1 - \left(t/(t + \theta_s) \right)^{-\delta_s} \right) + (1 - \alpha_s) \left(1 - \left(t/(t + \theta_l) \right)^{-\delta_l} \right).$$

These revisit rates are conditional in the sense that they depend on whether individual k decided whether or not to revisit the retailer's site. Therefore, as in our models of page view rates, we find the unconditional transition rates after leaving the site (see Figure 3). These unconditional revisit cycles are exponentially distributed with rate μb. Again, notice that transition probability b is dependent on each consumer's search characteristics, as is the unconditional revisit cycle.

5   Results

This section summarizes our results. We present our results following the same structure we used to present our methodological development. That is, we first describe the effects of price and prior search activities on search and purchase probabilities that are obtained from our dynamic discrete choice models. We then interpret results related to within-visit durations (learning effects) and across-visit durations (customer segmentation) that are obtained from our timing models.

5.1   Dynamic Discrete Choice Model Results

Table 4 reports the results from our dynamic discrete choice models, and specifically the transition state probabilities associated with the utility function defined by Equation (2). Parameter estimates for Γ_{vn} and p-values are given for independent variables X_{kvn}. When interpreting results, it is important to keep in mind that whereas $f(t)$ represents the probability of ending search, that $p(t)$ and $b(t)$ are conditional probabilities, i.e., they represent, respectively, the probability of purchase *conditional* on ending search and the probability of revisiting *conditional* on leaving. Estimation results are reported on a sample of data, enabling us to compare the predictive performance with a hold out sample. Due to the small number of purchases, we used unevenly balanced samples for estimation. For example, in order to predict $p(t)$, we randomly sampled equal numbers of purchasing and non-purchasing observations - with the equal sample size of 332 each.

There are two key motivations for why we used weighted sampling. First, using choice frequencies that were approximately equal for the estimations resulted in parameter estimates that were

more stable (and interpretable) across different specifications. Second, the use of balanced choice frequencies enabled us to design a validation that was "fairer" in terms of comparing within-sample and out-of-sample predictions. That is, unweighted samples resulted in validations that had a very high prediction accuracy (as the majority of purchase-no purchase predictions would be accurate due to the fact that there are only 474 (3.1%) purchases out of 15,237 purchase decision cycles.) Importantly, from a theoretical perspective, the use of unbalanced choice frequencies for our estimations will result in consistent parameter estimates. Formally, Manski and Lerman (1977) and Lerman and Manski (1978) show that the exogenous sampling maximum estimator (ESML) may be used with unbalanced (or choice-based) samples for multinomial logit models that include a full set of identifiable constants. In this case, the ESML estimator provides consistent parameter estimates with the exception of the constants. However, the bias in constants can be quantified (and removed) by using the estimation and data sample shares. For these reasons, we report model results using weighted sampling.

Effects	Variables	$f(t)$: Pr(End Search)	$p(t)$: Pr(Purchase)	$b(t)$: Pr(Revisit)
Price	d	-.346	1.441	.879
		(.000)	(.000)	(.000)
Inter-visit Durations	τ	.002	-.0112	
		(.000)	(.038)	
Page View Durations	ν	-2621.080		
		(.000)		
	ω	8254.683		
		(.000)		
Repeated Visits	$I_{\{v=1\}}$	-.760	1.516	1.295
		(.000)	(.000)	(.000)
	$I_{\{v=2\}}$	-.786	2.519	2.159
		(.000)	(.000)	(.000)
	$I_{\{v=3\}}$	-.832	1.792	1.923
		(.000)	(.000)	(.000)
	$I_{\{v>3\}}$	-.816	2.108	3.168
		(.000)	(.000)	(.000)
Page Requests	$\log n$	-2.407	.124	.187
		(.000)	(.002)	(.000)
	constant	.874	-1.783	-1.396
		(.000)	(.000)	(.000)
	Observations	29802	664	7162
	Log Likelihood	-17063.912	-369.314	-3959.228

Table 4: Discrete choice model results for transition probabilities

As shown in Table 4, higher discount rates are associated with higher purchase and revisit probabilities and lower exit probabilities. Although the results clearly show that higher discounts result in higher search and purchase probabilities, we need to be concerned about potential endogeneity. That is, one might argue that customers with higher price sensitivities may be more inclined to search and more likely to find higher discounts. In turn, if a higher discount rate is the result of (and not the cause of) strong consumer engagement (as revealed through the number of repeat visits and page requests), then the price effect (especially on exit decisions) may not be obvious as it seems in the results. Using the procedure described in Wooldridge (2005), we tested for these potential endogenity effects. The results of this test (summarized in Appendix 2) suggest that consumer search activities do not have a significant effects on discount rates, which implies that discount rates do not change over repeated searches. The high correlation (0.74) of discount rates we observe between consecutive searches in our data also supports the argument that customer search activities do not have a significant effect on discount rates. Note that the high correlation in discount rates was observed both within visits as well as across visits.

Table 4 also shows that the log of the number of page requests influences search, purchase, and revisit probabilities. Customers with one or two page views are more likely to end search than customers with deeper searches. Conditional on exisiting, those customers who have viewed more pages are more likely to purchase. Conditional on not purchasing, those who have viewed more pages are also more likely to revisit.

The impact of page view durations on search probabilities is captured in two variables. ν represents page view durations and ω represents the total time spent during a visit; both are measured in days. Those individuals who have longer page view durations are less likely to end their searches compared to those individuals who have shorter page view times. However, this effect changes with time, and as the total time spent at the visit increases, individuals are more likely to end their search.

Finally, the number and duration of inter-visit times also influences behavior. Compared to first-time visitors, those who revisit are more engaged and less likely to end their searches. Importantly, purchasing probabilities are strongly influenced by the number of revisits, i.e., the results show that purchasing probabilities are higher for those who revisit the site versus first-time visitors. While the directional impact of the number of repeat visits on future search, purchase, and revisit probabilities is clear from the results, the magnitude of these impacts is less clear (i.e., the coefficients do not monotonically increase or decrease as the number of visits increases). Inter-visit durations (measured in days) reveal that consumers who revisit after relatively short durations are less likely to end their searches and more likely to purchase.

5.2 Timing Model Results

Our timing model allows us to investigate the presence of learning effects (as reflected through page view durations) as well as customer heterogeneity and customer segmentation (as reflected through different purchasing probabilities based on short and long times between visits). We can

examine the effects of search depth on page view durations by comparing a model that uses a fixed arrival rate with one that allows the inter-arrival rate to depend on the number of page views and the number of visits. Figure 6 shows shows that page view time tends to decrease, both within a session as well as across visits. This suggests the presence of learning effects, as customers become more adept at navigating the retailer's site.
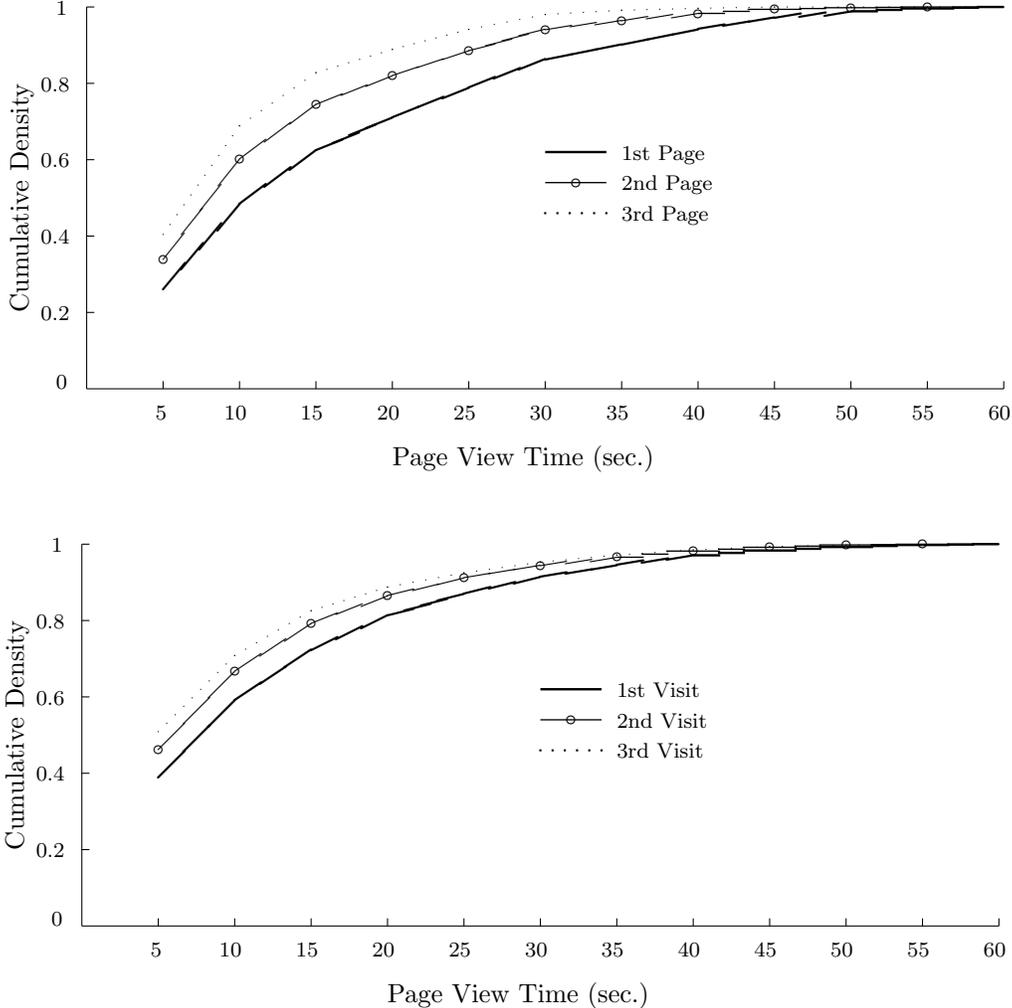

Figure 6: Cumulative distributions of page view times by page requests and visits

Estimation results, summarized in Table 5, also support the presence of learning effects, particularly within a visit. This is seen by comparing the log-likelihood values of a model in which we allow λ to depend on n, the number of page views, with a model in which we allow λ to depend on both n and v, the latter of which is the number of visits.

An investigation into how inter-visit times influence search and purchase behaviors uncovers the presence of two distinct customer types. We can see this by comparing three models that

Model	Log Likelihood
Constant λ	-16314.6
$\lambda = \lambda_n$	-15230.8
$\lambda = \lambda_{vn}$	-15219.6
Number of Observations	22000

Table 5: Model fit: Three different models for exponential rates of page view times

assume inter-arrival rates are: (1) constant; (2) a heterogeneous rate over one consumer population; and, (3) a mixture of two heterogeneous populations. The second model allows consumers to have different motivations for revisiting, which are revealed in heterogenous inter-visit rates with a Gamma distribution. The third model is based on evidence from a large body of marketing literature that finds two distinct categories of consumer behaviors: goal-oriented and experiential consumptions. We model these two customer segments in the third model by representing inter-visit rates as a mixture of two Gamma distributions.

Table 5 and Figure 7 compare these three models. The results clearly show that there is a large discrepancy between the constant rate and heterogeneous rate models. The results also clearly show that inter-arrival times differ across two distinct customer segments, i.e., the model that uses a mixture of two Gamma distributions provides evidence that there are two distinct customer segments: those who revisit with a goal towards purchasing and those who revisit with the goal of obtaining more product information.

Model	Log Likelihood
Constant μ	-26002.0
$\mu \sim Gamma(\theta, \delta)$	-19330.5
$\mu \sim \alpha Gamma(\theta_s, \delta_s) + (1 - \alpha)Gamma(\theta_l, \delta_l)$	-17936.5
Number of Observations	5117

Table 6: Model fit: Three different models for inter-visit rates

6 Validation

We use unbalanced in-sample and out-of-sample datasets to validate the prediction accuracy of our model and to verify that our first-order Markov chain modeling assumption is accurate. Under this unbalanced sampling, a purely random hit rate for each binary choice probability is 50%. Table 7 reports the fit statistics for three transition probabilities. Notice that the hit rates are more than 60% for all transition probabilities, which is a notable improvement in forecasting accuracy over the baseline of 50%.
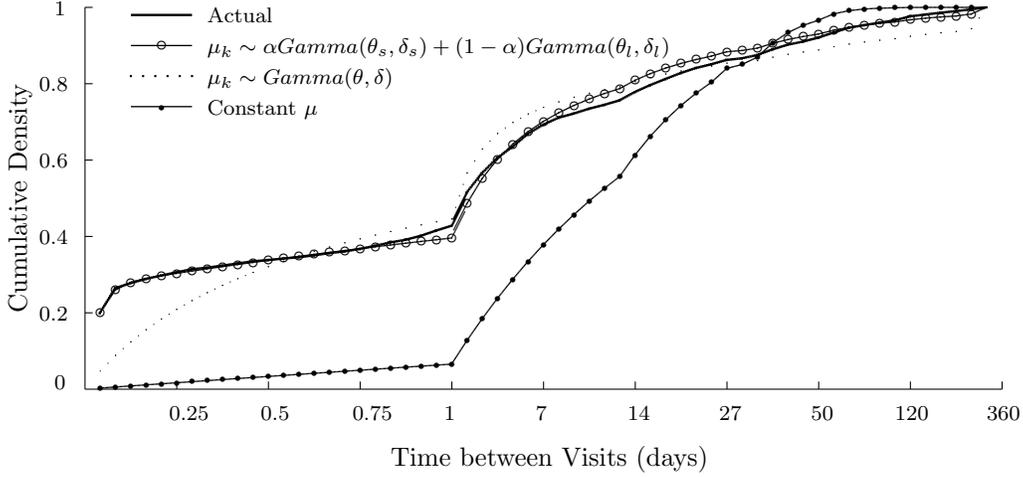
Figure 7: Actual and predicted cumulative distributions of inter-visit times

Prob.		First-order M.C.		Second-order M.C.		
		Hit Rate	Log Likelihood	Hit Rate	Log Likelihood	Observations
f	In-sample	60.76%	-17063.912	60.91%	-16949.631	29802
	Out-of-sample	60.49%	-7353.521	60.76%	-7355.446	12752
p	In-sample	62.62%	-369.314	62.86%	-367.341	664
	Out-of-sample	64.11%	-146.970	64.15%	-151.023	284
b	In-sample	62.68%	-3959.228	62.90%	-3940.398	7162
	Out-of-sample	63.03%	-1670.850	63.28%	-1662.035	3072

Table 7: Prediction accuracy for discrete choice probabilities

A further comparison of in-sample and out-of-sample predictions shows no discernable differences in hit rates. For purchase and revisit probabilities (p, b), out-of-sample hit rates are slightly higher than the in-sample hit rates. This is likely due to the large number of non-purchasing and non-revisiting searches in our sample.

In our models, we made the assumption that utilities are determined only by present search characteristics, that is, we used a first-order Markov chain model. In order to verify this assumption, we can model a second-order Markov chain in which consumer utilities are dependent on the present and previous searches. Table 7 indicates no significant differences in model fit (log likelihoods) and prediction performance (out-of-sample hit rates), which validates our assumption that consumer choices are primarily influenced by current search characteristics.

7 Application to Online Promotion Evaluation

We have observed that individuals' search and purchase behaviors are closely related with price and prior search activities, and that accounting for the number and timing of repeat visits is particularly important in understanding these behaviors. Figure 8 can be used to visualize why it is important to account for revisiting behavior in the context of promotion evaluation. The figure shows "cumulative" purchasing probabilities given relative discount rates (i.e., a price response curve). The cumulative purchase probabilities from the full dataset are shown in the figure, i.e., the 10% discount rate represents the percentage of individuals in the full dataset who were offered a discount less than 10% and purchased, the 20% discount represents the percentage of individuals who were offered a discount less than 20% and purchased, etc. These cumulative probabilities are differentiated by first-time and repeat visitors at the the retailer's website. Due to the small number of discounts that were offered above 60% in the data, we focus the discussion on interpretation of this chart for those individuals who have discounts below 60%.

These are several key points of interest. First, assume that customers were not tracked, and that searches and purchases that occurred after their first visit were not associated with the promotional link. In this case, The "First Visitor" line would represent the analyst's interpretation of how discount rates influence conversions. The very flat slop of this line suggests that it would be difficult in this case to discern differences in purchase probabilities as discount rates rise, subsequently leading to the conclusion that a promotion with a high discount rate level was not successful. Next, assume that all customers did enter through the promotional link, but that their visits were not linked. This case, represented by the "All" line, shows some evidence that purchase probabilities increase as discount levels increase. Importantly, the slope of the "All" line in this case is steeper than that of the "First Visitor" line, suggesting that an analyst would be more likely to conclude that a promotion with a higher discount rate was successful (in terms of not losing money) but not highly successful (in terms of generating high conversion rates and large incremental revenues).

What is most interesting, however, is a third case in which we assume that all customers who revisit the retailers' site do so through a non-promotional link. A large percentage of the purchases associated with the promotion would be classified as "latent demand" under this scenario – indeed, Figure 8 illustrates that the "Revisit" purchasing probabilities are nearly triple those of first-time visitors. Further, note that the slope of the "Revisit" line is much steeper than that of first-time visitors, which implies that revisit customers are more price-sensitive than first-time visitors. In turn, this means that the incorporation of revisiting behavior – and specifically understanding the dynamics of search, revisit, and purchasing probabilities on revenue generation – may lead to different conclusions regarding what the optimal discount level is.

We can further clarify the findings shown in Figure 8 and show how our model can be used to guide future promotion designs through a simple experiment. Using our full dataset, we use our dynamic discrete choice model to predict revisiting probabilities for each individual. Our model predicts that 2,913 of these individuals have purchasing probabilities greater than 50%. Table 8 reports the actual number of revisits per each predicted customer segment (high vs. low revisiting
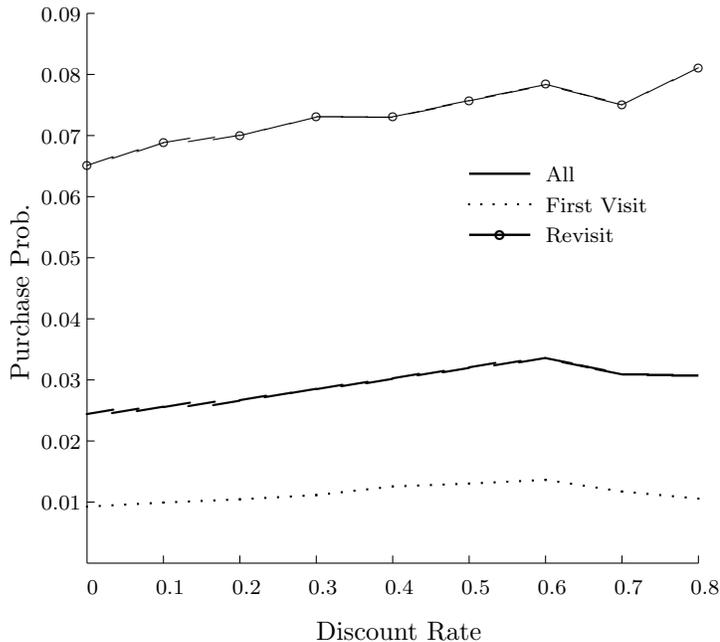
Figure 8: Price response by revisit behaviors

probabilities). Note that our hit rate is 67.7%, i.e., we correctly predict revisiting probabilities in the majority of cases.

Next, we select a random sample of 2,913 "anonymous" customers (who would be represented by the "All" line in Figure 8). This comparison, reported in the first row of Table 9, shows that both the conversion rate as well as revenues are almost three times as high for those individuals we predicted to have high probabilities of revisiting (which is consistent with the patterns observed in the full dataset that is shown in Figure 8).

We can build on this example by conducted a series of similar simulations in order to gain a better understanding of how discount rate levels impact promotional revenues. The results of these simulations are shown in Table 9. Each of the rows denoted by a specific discount rate shows the purchasing probability and associated revenues for each of the two samples. For example, our model would predict that if we showed a discount rate of 10% to all visitors in our dataset, that there would be a total of 2,067 individuals who would have revisiting probabilities greater than 50% and that 4.5% would eventually purchase, generating a revenue of $63,333. If we showed the same discount of 10% to a random sample of 2,067 visitors, we see that 1.4% purchase, generating revenues of $20,147. The subsequent rows in the table, representing discount rates of 20% to 60% were generated using the same logic. Note in all cases, that the number of revisitors is predicted using our model, and that as discount rates increase, the number of individuals with revisiting probabilities greater than 50% also increases.

By comparing the revenue columns for "All" demand and "Revisit" demand, we can see that the

optimal discount level is 50%, i.e., as discounts rise above 50% the additional revenues generated by those individuals who are "stimulated" to revisit and eventually purchase does not offset the loss in revenue due to giving high discount offers to customers who would have been willing to accept rates below 50%.

What is most interesting is a comparison of the purchasing probabilities and revenue generation associated with the current pricing data shown on the first row with those predicted by our simulation models. Remember that the current pricing reflects a distribution of discount offers presented to customers; the mean and standard deviation of the discount rate are 47.8% and 29.3%, respectively. However, the results of our simulations show that we can increase promotional revenues by 26.6% ($64,349 \rightarrow 81,469$) if we were to simply show all customers a discount rate of 50%. Under this scenario, the average purchase probabilities change slightly (7.8% $\rightarrow$ 7.1%), yet the expected revenues are large. This implies that current pricing policy may not be properly aligned with individuals' willingness-to-pay and that targeted marketing towards price-sensitive customers with high revisiting probabilities can generate additional revisitors ($2,913 \rightarrow 3,039$) and subsequently more purchases and revenues.

Actual	Predicted	
	High Prob. of Revisit	Low Prob. of Revisit
Revisit	1,973 (67.7%)	3,144 (18.0%)
No Revisit	940 (32.3%)	14,297 (82.0%)
Total	2913	17441

Table 8: Actual number of revisits by predicted revisit probabilities

	Anonymous ("All" Demand)		Target ("Revisit" Demand)		# Revisitors
	Purchase	Revenue	Purchase	Revenue	
Current Pricing	0.023	22,433	0.078	64,349	2,913
10%	0.014	20,147	0.045	63,333	2,067
20%	0.016	20,787	0.051	65,500	2,148
30%	0.018	22,852	0.057	72,203	2,406
40%	0.020	23,570	0.063	75,257	2,629
50%	0.023	<u>25,685</u>	0.071	<u>81,469</u>	3,039
60%	0.025	24,158	0.078	75,080	3,207

Table 9: Average purchase probabilities and total revenue for target and anonymous customers

8 Conclusions

We began our paper noting that, at a fundamental level, an investigation into whether it is important to incorporate latent demand into the design and evaluation of promotions involves understanding the linkages between individuals' search and purchase behaviors. We return to our key research questions and summarize possible implications on the design and evaluation of promotions depending on whether the relationships reflected in our research questions are true or false. The key implications of the findings from our study are summarized in Table 2 under the "True" column.

It is important to note that the findings from our study reflect one of a few studies that has explicitly examined the linkage between search and purchase behaviors. This was due to the very unique nature of our dataset which allowed us to: (1) clearly separate and capture incremental leisure travel demand from business travel demand; (2) simultaneously investigate the effects of within-visit and inter-visit characteristics on individuals' search and purchase behaviors; and (3) make causal inferences about how discount levels result in higher search, purchase and revisit activities.

The latter question, in particular, is one which to the best of our knowledge has not been previously examined in the literature (due the fact that researchers have not had access to datasets in which discount levels were randomly shown to consumers). Our finding, that search and purchase probabilities increase as discount levels increase, has important consequences, as it implies that the design of promotions influences both search patterns as well as purchases, potentially affecting conversion rates in non-linear ways. Failure to account for these interactions using simple attribution rules to evaluate promotions may, in turn, lead to incorrect conclusions on the effectiveness of a promotion.

A second key finding from our study relates to the fact that the majority of purchases did not occur among first-time visitors. This has two important implications. First, if customers revisit a site through a non-promotion link, they will be captured as "latent demand" and their purchases will not be associated with the promotion. Second, if customers make multiple visits to the retailers' website, then the purchase-decision cycle needs to be accounted for when determining the length of a promotion. In the context of air travel, where ticket prices are "high" relative to other consumer products, it is particularly interesting to note that majority of leisure customers who purchased visited the retailer's site multiple times. This findings may likely hold for other industries that have high-priced goods.

In the context of air travel demand, the fact that majority of customers who purchased searched during one or more visits preceding the purchase event may have important implications on the design of next-generation revenue management systems. These systems seek to integrate a discrete choice model (representing the alternatives an individual considers at the time of purchase) into capacity-based allocation algorithms. However, to date, all current and next-generation revenue management models assume customers' search and purchase decisions occur during the same visit. Consideration of choice sets that span multiple days and/or weeks may be needed to accurately model consumers' decisions, particularly in leisure markets that have customers who exhibit a high

degree of travel flexibility.

It is important to note that in different business contexts, other researchers have found results that contradict those from our study. This lack of current consensus in the literature highlights the need for additional research examining the linkages between individuals' search and purchase behaviors and their impacts on promotions, especially across a variety of different business contexts. Extensions of our work to business contexts in which visits to competitor retail sites are considered would also be particularly interesting.

Appendix 1

We define $q_{kij}(t)$ as the probability that decision maker k moves into state j after spending less than time t in state i. This probability is derived by relating it to W_{ij}, the time until the process in state i will transit to j. Formally,

$$
\begin{aligned}
q_{ij}(t) &= Pr(W_{ij} \leq t, W_{ij} < W_{ik}, \forall k \in A_i, k \neq j) \\
&= \int_0^t Pr(W_{ik} > s, k \neq j) \lambda_{ij}(s) e^{-\int_0^s \lambda_{ij}(\tau)d\tau} ds \\
&= \int_0^t \left(\prod_{k \neq j} e^{-\int_0^s \lambda_{ik}(\tau)d\tau} \right) \lambda_{ij}(s) e^{-\int_0^s \lambda_{ij}(\tau)d\tau} ds \\
&= \int_0^t \lambda_{ij}(s) e^{-\int_0^s \sum_{k \in A_i} \lambda_{ik}(\tau)d\tau} ds \\
&= \int_0^t \lambda_{ij}(s) e^{-\lambda_i s} ds \\
&= \int_0^t \lambda_i P_{ij}(s) e^{-\lambda_i s} ds \\
&= \int_0^t \frac{e^{\Gamma_i X_{ij} + \alpha_{ij} s}}{\sum_{k \in A_i} e^{\Gamma_i X_{ik} + \alpha_{ik} s}} \lambda_i e^{-\lambda_i s} ds.
\end{aligned}
$$

Define $k = 1, ... K$ as the set of individuals and t_{kij} as the time when individual k in state i makes a decision into state j. Given one-step transition probabilities:

$$
q_{kij}(t) = \int_0^t \frac{e^{\Gamma_i X_{ij} + \alpha_{ij} s}}{\sum_{m \in A_i} e^{\Gamma_i X_{im} + \alpha_{im} s}} \lambda_i e^{-\lambda_i s} ds.
$$

the corresponding likelihood function is expressed as:

$$
\begin{aligned}
L(\Gamma, \alpha | \lambda) &= \prod_{k=1}^K \prod_{i=1}^I \prod_{j \in A_i} \left(\frac{\partial q_{kij}(t_{kij})}{\partial t} \right)^{Y_{kij}} \\
&= \prod_{k=1}^K \prod_{i=1}^I \prod_{j \in A_i} \left[\frac{e^{\Gamma_i X_{kij} + \alpha_{ij} t_{kij}}}{\sum_{m \in A_i} e^{\Gamma_i X_{kim} + \alpha_{im} t_{kij}}} \lambda_i e^{-\lambda_i t_{kij}} \right]^{Y_{kij}}
\end{aligned}
$$

where

$$
Y_{kij} = \begin{cases} 1 & \text{if individual } k \text{ in state } i \text{ transit to state } j \\ 0 & \text{otherwise} \end{cases}
$$

Appendix 2

We investigate the endogeneity of discount rates with other factors by assuming that discount rates are correlated with other search factors as follows:

$$d = \pi_0 + \pi_1 \tau + \pi_2 \nu + \pi_3 1_{\{v=1\}} + \pi_4 1_{\{v=2\}} + \pi_5 1_{\{v=3\}} + \pi_6 1_{\{v>3\}} + \pi_7 log(n) + \xi \qquad (4)$$

Effects	Variables	d: discount rate	$p(t)$: Pr(purchase)
Inter-visit Durations	τ	-.0000454	-.0114
		(.535)	(0.033)
page view Durations	ν	49.7505	
		(.000)	
Repeated Visits	$1_{\{v=1\}}$	.0485	2.0998
		(.000)	(0.011)
	$1_{\{v=2\}}$	.0605	3.2440
		(.000)	(0.002)
	$1_{\{v=3\}}$	.0850	2.8223
		(.000)	(0.052)
	$1_{\{v>3\}}$	.1164	3.4945
		(.000)	(0.065)
Page Requests	$\log n$	.00294	0.1645
		(.000)	(0.015)
	Constant	.3971	2.9692
		(.000)	(0.643)
	d		-10.4617
			(0.514)
	Residual $\widehat{\xi}$		11.880
			(0.458)
Observations		42554	664
R^2		0.0223	0.1982

Table 10: Testing for endogeneity of price

Table 10 reports the regression results. Although repeated visits and page requests have a significant effect on discount rates, the very low R^2 value implies that there is no significant endogeneity effects between discount rates and search factors. This argument can be further strengthened by conducting an endogeneity following the procedure described in Wooldridge (2005). If we assume that the discount rate d is an endogenous variable, then in the following regression of purchasing probability p, the error term ξ for discount rate d in Equation (4) will be correlated with ϵ in Equation (5). The easiest way to test this is to include ξ as an additional regressor in Equation

(5). Since ξ is not observed, we may use residuals $\widehat{\xi}$ estimated from Equation (4).

$$p \;=\; \gamma_0 + \sigma_1 \boldsymbol{d} + \gamma_1 \tau + \gamma_2 \nu + \gamma_3 1_{\{v=1\}} + \gamma_4 1_{\{v=2\}} + \gamma_5 1_{\{v=3\}} + \gamma_6 1_{\{v>3\}} + \gamma_7 n + \epsilon \;(+\sigma_2 \widehat{\xi}) \;\;(5)$$

The last column in Table 10 represents the regression of purchase probability p on the residuals $\widehat{\xi}$ for discount rates. The p-value (0.458) for $\widehat{\xi}$ supports the argument there is no significant endogeneity effects between discount rates and search factors. We thus conclude that consumer search activities do not have significant effects on discount rates.

References

[1] Bucklin, Randolf E. and Catarina Sismeiro. 2003. A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, 40(3): 249-267.

[2] Chatterjee, Patrali, Donna L. Hoffman, and Thomas P. Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science*, 22(4): 520-541.

[3] Geraghty, Kevin. 2010. Personal communication with authors.

[4] Harteveldt, H. H., Wilson, C. P. and et al. 2004. Why leisure travelers book at their favorite sites, *Forrester Research: Trends*.

[5] Hoffman, Donna L., Thomas P. Novak, and Adam Duhachek. 2003. The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology*, 13(1&2): 3-16.

[6] Hoffman, Donna L. and Thomas P. Novak. 1996. Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing*, 60(July) 50-68.

[7] Johnson, Eric J., Steven Bellman, and Jerry Lohse. 2003. Cognitive lock-in and the power law of practice, *Journal of Marketing*, 67(April): 62-75.

[8] 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): 299-308.

[9] Lerman, Steven R. and Charles F. Manski. 1979. Sample design for discrete choice analysis of travel behavior: The state of the art. *Transportation Research*, 13A: 29-44.

[10] Manski, Charles F. and Steven R. Lerman. 1977. The estimation of choice probabilities from choice based samples. *Econometrica*, 45: 1977-88.

[11] McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior, in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press, 105-42.

[12] Moe, Wendy W. 2003. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream, *Journal of Consumer Psychology*, 13(1&2): 29-39.

[13] Moe, Wendy W. and Peter S. Fader. 2004. Dynamic conversion behavior at e-commerce sites. *Management Science*, 50(3): 326-335.

[14] Montgomery, Alan L., Shibo Li, Kannan Srinivasan and John C. Liechty. 2004. Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4): 579-595.

[15] Park, Young-Hoon and Peter S. Fader. 2004. Modeling browsing behavior at multiple web sites. Marketing Science, 23(Summer): 280-303.

[16] PhoCusWright. 2004. *The PhoCusWright Consumer Travel Trends Survey*, 6. Downloaded from http://store.phocuswright.com/phcotrtrsusi.html.

[17] Smith, Barry. 2010. Personal communication with authors.

[18] Smith, Barry, Ross Darrow, John Elieson, Dirk Guenther and B. Venkateshwara Rao. 2007. Travelocity becomes a travel retailer. *Interfaces*, 37(1): 68-81.

[19] Wooldridge, Jeffrey. 2005. *Introductory Econometrics: A Modern Approach*, 3rd edition. South-Western College Publishing.

[20] Wolfinbarger, Mary and Mary Gilly. 2001. Shopping online for freedom, control, and fun. *California Management Review*, 43(2): 3455.