

To Buy or Not to Buy? A Two Stage Model of within Site Search

Ammara Mahmood

Catarina Sismeiro*

June 2012

Preliminary and Incomplete[†]

Abstract

We investigate the *within* site purchase and search behavior of online customers visiting one of the largest European online travel agencies. Using a dynamic two stage model of purchase incidence and carrier choice we study how price uncertainty affects consumer purchase behavior. We find that current and future category value along with visitors' browsing experience and search effort are important predictors of purchase incidence. We find that consumers are forward looking and learn about the distribution of prices as they search. Spatial price variation results in greater search, while the number of flight options positively influences purchase incidence. Customers who search actively have a lower purchase threshold, however, customers who exit without making a purchase are less likely to purchase in subsequent visits to the website. Furthermore, expected expenditure and flight characteristics are key determinants of airline choice. Both observed and unobserved heterogeneity in search behavior are also found to influence purchase decisions. Tests of predictive ability also validate the proposed search model compared to a model without pre-purchase behavior. We also discuss important implications for website managers.

*Ammara Mahmood is Doctoral student at Said Business School, Oxford University. Email: am-mara.mahmood@sbs.ox.ac.uk. Dr. Catarina Sismeiro is Associate Professor at Imperial College Business School, UK. Email: catarina.sismeiro@imperial.ox.ac.uk

[†]Do not cite without authors' consent.

1 Introduction

Revenue management (or yield management) systems are today the standard pricing mechanism in many markets characterized by perishability and capacity constraints (Desiraju and Shugan,1999). These include markets for air travel, hotel bookings, and car rentals, all of significant economic impact not only offline but especially online. In recent years the internet has become a preferred source of information and transaction in such markets. Recent studies reveal, that the internet is the most frequently used medium for travel research (“The Traveler’s Road to Decision”, 2009) and that it is expected that one third of the world’s travel sales will be made online by 2012 (PhocusWright, 2011).

The increase in importance of online search and online purchases in markets where pricing is set by revenue management systems is likely due to the ease in searching for alternatives and in obtaining updated information on prices and availability online. The internet has indeed lowered search costs and allowed customers to search while prices, and availability, change continuously, a result of the algorithms at the core of these pricing mechanisms. In addition, the use of the internet has also opened new research opportunities as rich datasets on detailed consumer behavior become available that track both consumer search and their purchases.

With this study we contribute to the literature on revenue management systems by investigating the impact of such complex pricing strategies on consumer search and purchase behavior. It is our goal to understand how consumers cope with the significant uncertainty caused by such systems as prices change dynamically during the day, how consumers form their expectations, and what is the impact of their search effort on purchase. To the best of our knowledge, there is a lack of research that investigates such issues and that takes into account consumer search and purchase behavior when facing revenue management systems (Desiraju and Shugan,1999).

Hence, the primary aim of this study is to understand the impact of the significant spatio-temporal price dispersion introduced by revenue management systems on the purchase behavior of customers. We do so by jointly modeling search and purchase behavior using a novel and detailed panel data of online search and purchase from a large online travel agent. We use a flexible modeling

approach that does not impose a-priori restrictions on consumer search and purchase behavior. Instead we exploit our rich data to draw inferences regarding preferences, and estimate a two stage dynamic model to study the within site purchase and search behavior of online customers.

In this two-stage dynamic model consumers decide whether to make a purchase now or to continue searching and, conditional on purchase incidence, consumers select the airline carrier for their travel. We assume consumers have dynamic price expectations regarding ticket prices, whereby sophisticated consumers update their expectations based on observed prices. Expected future prices influence the expected value of future travel options, and as a result, expectations are allowed to impact the decision to continue searching or not. We test for alternative models of expectation formation, including temporally rising prices and rational price expectations. In addition, by including covariates capturing what a visitor is exposed to during browsing and the actions taken at the site, we measure how information gathered and search effort impact their purchase incidence decisions. Finally, and very importantly, we exploit the richness of our panel data and account for observed and unobserved heterogeneity, and correct for endogeneity, the latter often disregarded from models of consumer search.

Our empirical findings suggest that pre-purchase behavior is a key determinant of purchase outcomes and that ignoring search behavior can lead to misleading inferences. In addition, ignoring consumer pre-purchase behavior from all site visitors (including casual browsers who never purchase at the site) compromises the predictive power of the model. Hence, in the context of online search for travel related products, ours is one of the first papers that highlights the need to include all browsers in a model of search, whether or not they are also purchasers.

We also find evidence that consumer decisions are dependent on future category value and elements of the search environment significantly impact their behavior. In line with existing theories of consumer search (Lanzetta, 1963, Stigler, 1961) our dynamic two stage model indicates that consumers search in order to resolve uncertainty; we observe that when the available choice set is large, consumers appear more confident and search less, however, when there is greater price variation in the available options they search more. In addition, we find that search effort also impacts a consumers decisions to search, site visitors are more likely to make a purchase the more

actively they search within a short span of time as these customers appear to have higher search costs. However, customers that exit the website have a lower chance of making a purchase on subsequent visits. Furthermore, consumers use current prices to form expectations of future prices and dynamically update their price expectations.

The remainder of the paper is organized as follows. Section 2 summarizes the relevant literature; Section 3 provides an overview of the air travel industry and the its spatiotemporal price variation and Section 4 presents the rich dataset used. We present the model and key modeling assumptions in Section 5, and Section 6 summarizes the estimation approach. Finally, our main findings are presented in section 7 and in Section 8 we provide our conclusions and propose areas for future research.

2 Literature

There are three streams of literature relevant to this study: (1) literature on revenue management systems, (2) literature on search behavior, and (3) literature on online search and purchase behavior.

Revenue Management

Pricing under revenue management regimes is a complex phenomenon constrained by two factors “perishability” and “capacity” constraints (Desiraju and Shugan, 1999). Considering the case of air-travel, where yield management first started, perishability stems from the fact that once a flight departs the seats can no longer be sold. Capacity constraints on the other hand, arise from the physical limitations on the number of people who can be accommodated on a single aircraft. The combination of perishability and capacity constraints has driven airlines to adopt complex revenue management pricing systems to profitably fill each aircraft to capacity (Wardell, 1989). These revenue management strategies introduce significant temporal price variation. Research in operations management finds that in a normal day, fares can be updated up to 200,000 times in a travel agent’s computerized reservation system (Hopper, 1990). For a specific flight prices can change as often as seven times during a single day (Etzioni et. al., 2002).

Revenue management systems are today adopted in a variety of markets and industries, all of which are characterized by these same issues of perishability and capacity constraints including hotel bookings and car rentals. The basic idea of these systems is to continuously monitor demand (through for example centralized booking systems, which are at the center of the technology development) and adjust pricing to maximize the yield of each seat, each room or even each car (Boyd and Bilegan, 2003). For example, in the case of air travel, if the likelihood of selling a ticket at full price increases the number of seats available at lower fares decrease, and hence prices increase. Previous evidence shows that these systems have lead to significant increases in profitability that far outweigh their cost (Davis, 1994).

Because the algorithms behind these systems are so vital, existing research in the field of revenue management has thus far focused on the optimal pricing strategy of firms and their demand forecast (e.g., Dana, 1998, and Perakis and Sood, 2006). However, most of this work has assumed that consumers arrive as a stochastic process and do not endogenise consumer behavior based on firm pricing. For instance, Ben-Akiva (1987) and Sa (1987) forecast demand for flights using regression and time series models based on advanced and historical bookings data. Recently, efforts have been made to develop choice-based revenue management models, whereby discrete choice models are used to forecast consumer demand (e.g. Talluri and van Ryzin, 2002). For example, Ferguson et. al., (2011) use a two step approach to propose a choice-based revenue management system and estimate their model using bookings data. They do not consider search and pre-purchase behavior. Similarly Vulcano et. al. (2010) use data for bookings made by customers at an airline but their data lack information on consumer search and arrival processes which they simulate. These models still assume myopic customers and do not consider search. As a result, strategic waiting by customers has largely been ignored.

Boyd and Bilegan (2003) note that the main challenge for revenue management systems is to effectively use the information contained in consumer purchase requests, as eventually airlines would like to charge each customer their willingness to pay. Therefore, it is important to study consumer pre-purchase and purchase behavior to develop holistic insights into consumer willingness to pay. Elmaghraby and Keskinock (2003, p. 1,298) stress the importance of understanding of consumer

behavior in revenue management “An important element that is largely missing, both in most of the academic literature and price optimization software, is the consideration of strategic customer behavior.” This need is made even more significant if we consider the extreme price variability arising from revenue management, that creates added incentives for search, there is the need to better understand the implications of complex pricing strategies on consumer search behavior (Desiraju and Shugan, 1999).

Our study aims to address this gap in the literature by formally modeling consumer search and purchase behavior by predicting demand based on consumer search effort and reaction to the dynamic search environment.

Consumer Search Behavior

Our work also draws on the literature on consumer search models. According to the theoretical literature on consumer search, consumers continue to search as long as the benefits accruing from search exceed its costs (Stigler, 1961). The benefits from search have been cited as a better product or lower prices, while the costs comprise the time and effort involved in the search process. There are two types of search models (Baye et al. 2007): (1) simultaneous search (Stigler, 1961) and (2) sequential search models (Weitzman, 1979).

According to the theory of sequential search, consumers decide to stop or continue search if the marginal benefits of search outweigh the marginal costs of additional search. According to Morgan and Manning (1985) sequential search models are optimal if customers have perfect recall and have no time preference, this makes traditional sequential models problematic in markets with high price volatility and limited recall. The theory of simultaneous search or fixed sample search, on the other hand, assumes that consumers determine the size of the consideration set based on their search costs prior to actual search.

Empirical work in marketing has studied consumer search behavior and tested alternative search theories (e.g., Fox and Hoch, 2005; Gauri et. al., 2007). A large body of work investigates consumer response to promotions and price changes through purchase acceleration or delays and stockpiling (e.g., Neslin et. al, 1985; Mela et. al., 1998). Another stream of research studies spatial search

across stores and suggests that price sensitive consumers often cherry pick across stores to find the best deals (e.g. Putrevu and Ratchford, 1997; Fox and Hoch, 2005). Talukdar and Sudhir (2007) jointly study spatial and temporal search in the context of grocery purchase and measure search effectiveness in terms of the resulting monetary savings. Similarly, Fox and Hoch (2005) and Ratchford and Srinivasan (1993) also estimate savings emanating from search. Finally, a significant stream of research has studied consideration set formation and tested alternative models (e.g. Chiang et al., 1999; Mehta et al., 2003; Van Nierop et. al., 2010).

A limitation of the extant literature is the breadth of industries studied: most authors research search in the context of grocery shopping (Gauri et al., 2007; Fox and Hoch, 2005; Urbany et al., 1996) and durable goods markets (e.g., Ratchford, 1993; Conslik et. al., 1984; Sobel, 1984). Though temporal price variation is an important market force in these industries (e.g., promotional activity or price decreases due to the sequential introduction of new product generations), the level of price changes and price uncertainty is not as extreme as in the case of industries and markets subject to revenue management systems. Very few studies look at services and other more complex products subject to extreme price variability. Honka (2012) is a notable exception. This author studies search and switching behavior of consumers in the market for auto insurance by jointly estimating search costs, consideration sets, and switching costs. A key limitation of the study is that consumers decide on the policy to buy prior to search, a more realistic approach would involve consumers selecting policy characteristics while simultaneously searching for policies. In addition, Honka (2012) assumes that consumers have rational expectations about prices for all companies in the market, as we show later such an assumption would not be feasible for a market characterized by high levels of price variability.

Online Search Behavior

Compared to offline markets, consumers can search online stores with little time and effort: the internet has reduced the cost of searching. In addition, online commerce can track website visitors and observe their search behavior, something that had been either difficult or too expensive to engage in offline. An inherent advantage of the internet is the availability of detailed data on

consumer browsing and purchase patterns.

There is a well established tradition in marketing that explores consumer browsing behavior in online markets. There is evidence that purchase conversion is influenced by page design (Mandel and Johnson, 2002), number pages requested and time spent at the site (Sismeiro and Bucklin, 2004), and frequency of site visit (Moe and Fader, 2004). Studies in this tradition model browsing and purchase decisions as independent events or use browsing and search as covariates of the venal purchase probabilities. For instance, Moe and Fader (2004) use a stochastic approach to model purchase conversion and are unable to capture how purchase outcomes could be influenced by user actions while searching. In addition, previous studies of online search do not incorporate the impact of prices on a consumer's decision to purchase or continue searching.

Recently, empirical research on online consumer search behavior has received renewed attention as a means of better understanding consumer preferences and measuring search costs and their impact on consumer behavior. Researchers in this tradition have imposed structural assumptions regarding the search process to identify search costs given data limitations. For example, Kim et al. (2010), using aggregate view rank data from Amazon.com, model consumer search for camcorders as a sequential process. The authors assume that consumers are aware a priori of the distribution of prices and that the aggregate viewer rankings are based on individual-level optimal search sequences and compute reservation values to determine the optimal stopping rule. These assumptions are reasonable in the durable goods industry with limited uncertainty, though not so in a context with extreme uncertainty. In addition, the absence of actual search data makes some of the identification assumptions restrictive.

In a related study Santos et.al. (2012) empirically test sequential and simultaneous search models using individual level browsing and purchase data for online books. Their study rejects the sequential search model with a priori known price distributions. They use the parameters of their search model to estimate demand elasticities for online books and find no evidence that observed prices influence future decisions to search. However, Santos et. al. (2012) base their finding on transaction prices as their dataset does not contain prices observed at each search occasion. This is a significant limitation and could explain why the authors do not observe any effect of observed

price on information search. The authors note that their choice of books does not have general application as they study a homogenous product category with fairly limited price variation across the few dominant sellers and even over time. In contrast, the industry we study is characterized by high levels of price volatility.

In light of the growing importance of online travel purchase, academics are beginning to take interest in the pre-purchase behavior of online travel purchasers. In a recent study on consumer search for air travel, Nair et. al. (2010a) analyze the determinants of consumer's choice of travel website, browsing time and purchase. Nair et. al. (2010a) only observe data for the final transaction prices but not prices from other competing websites. Due to this data limitation the authors cannot determine the impact of prices observed during search on purchase outcomes, nor how consumers search for price information online.

In a related study Koulayev (2010) estimates demand for hotels by estimating a structural model of sequential online search on a data set from an aggregator site that does not record bookings but records clicks on hotel links. Instead of modeling a booking, the authors model hotel clicks by consumers. The authors assumed that by clicking on a hotel, consumers reveal a preference for that hotel. However, this can be a misleading assumption as consumers may click in order to gather more information and does not indicate necessarily a preference and much less an actual booking. Koulayev (2010) estimates the search cost distribution though the data includes one single observation per individual.

One of the significant limitations of recent work is that data is often at a different level at which theory has been developed or aggregated in some way. This poses specific problems as authors need to develop a link between the different level of analysis and impose strong assumptions in order for their structural models to hold at all levels, including the one at which data is collected.

3 Spatial and Temporal Price Variation

Travel related online businesses have historically accounted for 40% of the revenue from e-commerce (Combes and Patel, 1997). In addition, to being a leading e-business, the market for travel

in general and air travel in particular is characterized by price uncertainty arising from complex revenue management systems. The data we will use to investigate consumer search behavior in a context of revenue management systems is a dataset from an Online Travel Agent (OTA) that contains the searches and bookings of consumers in the market for airline tickets over several months. We believe this is an ideal setting for this work. In this section we highlight the nature of price uncertainty across time and across carriers stemming from the revenue management systems in the context of air travel.

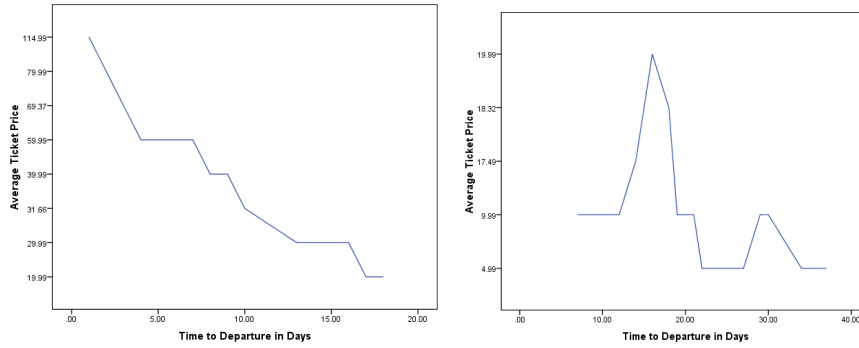
Temporal Price Variation

Conventional wisdom says that consumers should buy airline tickets early (the earlier one buys the tickets, the cheaper they will be) or perhaps just before the airplane departs (last minute deals).¹ In reality, revenue management systems cause pricing patterns that are far more complex than what conventional wisdom leads us to believe. For example, many firms now limit last minute deals (sometimes offering these through very specialized channels) just because last minute travels tend also to be business travels and hence less price sensitive. The algorithms behind the systems will allow price changes depending on demand conditions and on how many bookings or reservations are made.

To demonstrate this point, we present in Figure 3.1 the average posted prices for flights operating on two domestic European routes with a set departure date. Figure 3.1 dispels the traditional view that prices always increase as the departure date approaches. In fact, ticket prices do not follow a deterministic trend, making it difficult for consumers to make precise predictions about future prices. Figure 3.1 also highlights that prices may be more volatile for some flights compared to others depending on the particular supply and demand conditions, but that prices do change significantly over time till the departure date. It is this significant price variability over time that could lead consumers in the market for air travel to become more strategic and react differently to prices and change their search behavior.

¹We conducted experiments with 30 MBA students at Yale SOM, and asked them to plot the relation between price and days till departure for flights from New York to L.A. The vast majority of respondents plotted upward sloping graphs as they expected prices to rise closer to the departure date. A few respondents also indicated a drop in prices due to last minute deals.

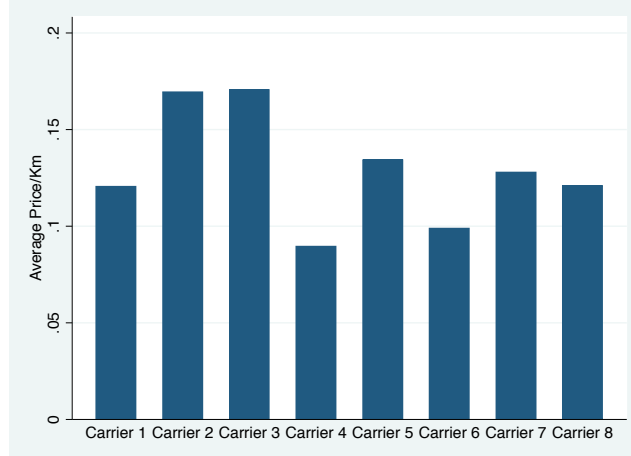
Figure 3.1: Temporal Price Variation



Spatial Price Variation

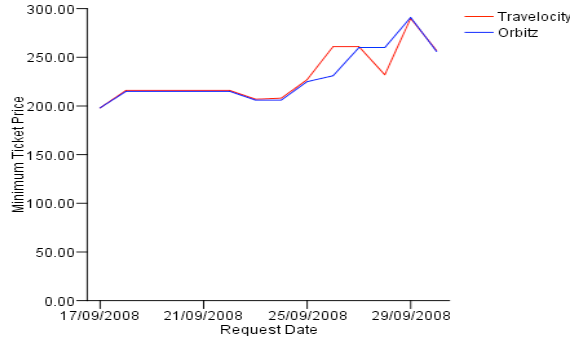
In addition to the temporal variation in prices, there is also evidence of spatial price variation stemming from price dispersion across airlines operating on a particular route. Within a given website customers need to resolve the additional uncertainty associated with different carriers offering different prices for the same travel itinerary. In Figure 3.2 we show the variation in average price (adjusted for distance) across airlines 15 days prior to departure. Despite controls for time to departure and distance we observe that the average price for some carriers is lower when compared to others (we present further details of price variability across carriers in detail in the data section). Hence, customers who search for air travel online face both spatial and temporal price variation.

Figure 3.2: Average Price/Km for Flights with 15 Days to Departure



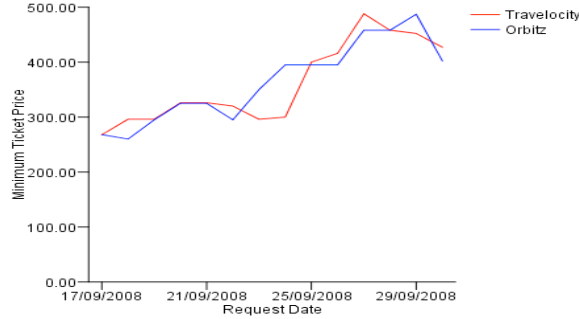
Another possible dimension of spatial price uncertainty is price variability across alternative OTAs. While prices vary over time, there is limited price dispersion across competing OTAs. To empirically test the degree of variability in prices across competing OTAs, we compared average ticket prices at two leading OTAs, Orbitz and Travelocity. Once a day from 17th September, 2008 until 30th September, 2008, we requested quotes for flights with exactly the same attributes (i.e., the same destination and departure and arrival date) from both websites. The flights included in this experiment were from New York to Las Vegas and New York to Washington D.C. departing 1st October, 2008 and returning on the 6th October, 2008. Our online searches took place at the same time of the day. Figure 3.3 and 3.4, exhibit the prices observed at the two OTAs during our experiment.

Figure 3.3: Comparison Of Minimum Ticket Price For Flight From New York To Washington D.C Departing On 1st October 2008 And Returning On 6th October 2008



We find that prices do not vary substantially across agents and both OTA websites provided very similar prices. This exploratory finding could explain the limited consumer search across websites that previous research has previously reported (e.g., Johnson et al., 2004). Indeed, previous studies emphasize this somewhat puzzling fact: that the level of search across websites tends to be very limited, despite the apparently low search costs in the online world. One possible explanation at least in the context of air travel is that the uncertainty in prices across websites is also very limited. Consumers seem to have to solve mostly two uncertainty problems: price variation over time and price variation across carriers.

Figure 3.4: Comparison Of Minimum Ticket Price For Flight From New York To Los Angeles Departing On 1st October 2008 And Returning On 6th October 2008



4 Data

For the purpose of this study we analyze site centric data from one of the largest European travel operators in the world.² The OTA has chosen to remain anonymous. In addition to the main travel website, the OTA operates price comparison sites which direct online traffic to the main ticketing website. The data set is novel and unique as it includes a complete record of pre-purchase behavior of consumers. We study the browsing and purchase behavior of users registered with the OTA and its subsidiaries. This enables us to identify repeated search by the same individual over our period of analysis. The data includes customers³ logging into the main OTA website and customers directed from price comparison sites, shop bots and search engines. Lack of across OTA data is not a limitation as there is evidence of limited across site search. Smith and Brynjolfson (2001) in their study of online shop bots show that 70% of consumers repeatedly visit a single site. More recently, Santos et. al. (2012) in their study of consumer online purchase and browsing behavior for books, find that consumers visited the same store 76% of the time within a week and up to 90% of the time the same store was visited within the same day. Santos (2012) also report based on comscore data that of customers who visited more than one store was 27% in 2002 and 33% in 2004, therefore, across site search activity is limited.

²<http://www.eurodestination.com/Flights%20&%20Airlines/european-low-cost-airlines-guide.html>

³We use customers, consumer and visitor interchangeably.

Consumers searching the website or any of the affiliated sites are quoted prices from pre contracted airline carriers. Customers can search for departure cities, arrival cities and travel dates. We define the combination of route⁴, dates and number of travelers as a single search request. After consumers request a flight the search engine displays the multiple options available. Consumers then have the option to select a flight and checkout, redefine the search criteria or exit the website. Every time consumers change the trip specification a new search request is generated.

The size of the data set required considerable effort in synchronizing the consumer specific data with the extensive flight information from the search engine. Careful analysis was required to filter out information not pertinent to the search model. Site visitors comprised both individual consumers and travel agents. The purpose of this study is to analyze the search patterns of consumers who search tickets for their personal travel, therefore, we excluded the booking activity of travel agents to avoid biases arising from their bulk purchasing activity (1.1% of all bookers were travel agents). To ensure that we observe all the search activity related to a specific booking, we considered flights searched during the month of March 2006, reserving the initial three months of online activity for variable initialization and search behavior in April for predictive analysis. The data reveals that bookings were made within 31 days of departure, and 50% of all bookings were made between 1-13 days to departure.

We removed all searches with incoherent search fields (e.g., departure dates after arrival dates) and instances where consecutively requested destinations were more than 400 miles apart.⁵ For such requests we could not know if mistakes had been made or if a consumer simply changed his travel plan. For example, if a consumer initially requested flights from New York to Boston and thereafter switches to Las Vegas to Boston or even switches to Las Vegas to Chicago we cannot be entirely certain of the consumer’s intent. In addition, we select twelve domestic routes⁶ which generate 90% of all domestic flight requests⁷ (domestic routes are those for which departure and

⁴For round trips, a route is a combination of arrival and departure cities.

⁵In the European context 400 miles is a significant distance.

⁶A route is defined as a combination of departure and arrival city pairs.

⁷Domestic bookings provide several advantages; we avoid currency conversions and we do not need to include the information of connecting flights, which could influence substantially the quality of the product. Furthermore, we minimize country specific effects because domestic flights are predominantly booked by residents of a single country

arrival city are within the primary country of activity of the OTA website under analysis).⁸

We included 8 airlines in our final estimation sample.⁹ Not all carriers operated on each route and the number of carriers for a given itinerary changed across time depending on seat availability. On average consumers had a choice between 2.6 airlines, with a minimum of 2 and a maximum of 4 carriers operating on a particular route. Air travel is a complex product with flights operating several times a day, including each flight option displayed to a customer in a choice model is not trivial. In the interest of tractability we combine the flights operated by an airline into one option. On average 2.2 different flights were displayed for each carrier.¹⁰ The final price per carrier was computed as the average price across all flights operated by the carrier.¹¹ We find price variation amongst airlines for flights with similar characteristics. Table 1 presents the average price across carriers and their market share. Carriers 5 and 6 have the highest market share and most frequently enter a consumer’s consideration set. At a given search occasion customers on average observe a standard deviation of 28 Euros for a flight on a particular date on a given route.

Table 1: Prices and Market Share Across Airlines

Carrier	Frequency of Availability	Average Price	Market Share %
Carrier 1	4,080	123.49	15.74
Carrier 2	734	100.42	4.05
Carrier 3	1,618	128.73	5.93
Carrier 4	707	57.17	5.04
Carrier 5	9,903	134.67	24.09
Carrier 6	5,196	117.69	25.55
Carrier 7	876	100.46	4.24
Carrier 8	6,560	131.25	15.35

The final data set comprises of 18,136 search requests generated by 5,087 site visitors. 2,776

⁸The OTA we study sells more than 40 different routes. However, the bulk of the business is concentrated around the top 12 routes for which information was collected. By focusing on domestic flight we avoid currency conversions and we do not need to include the information of connecting flights, which could influence substantially the quality of the product. Furthermore, we minimize country specific effects because domestic flights are predominantly booked by residents of a single country.

⁹We removed carriers that were never purchased during the period under analysis. Since the excluded carriers were part of the consumers’ consideration set, these carriers were used to compute variables measuring price variability and options for each search request.

¹⁰A maximum of 15 flights were displayed for a carrier while 20% of the time a single flight was operated by a given carrier.

¹¹For round-trips, we first computed the average price for each leg of the journey and then computed the sum for the two legs to arrive at the final price per carrier.

site visitors made at least one purchase during the period under analysis.

4.1 Search Behavior

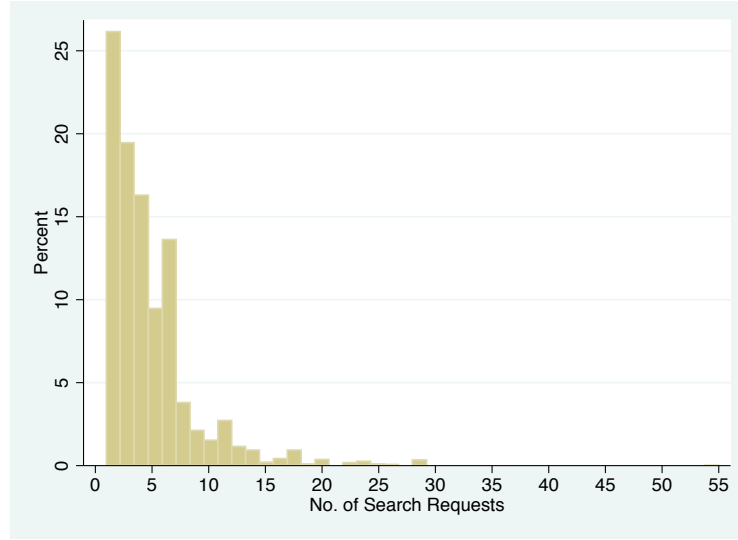
The data reveals that most site visitors exited or made a purchase after 4.9 search requests. Visitors who made no purchase exited after 3.9 search requests on average while customers who purchased at least once made search 5.3 requests on average (see Table 2).

Table 2: Summary of Search Behavior

	Visitors		Purchasers	
	Mean	Std. Dev	Mean	Std. Dev
No. Searches per visitor	4.90	3.81	5.22	4.01
No. Purchases per visitor	2.76	3.30	3.76	3.33
No. Search Sessions	3.04	2.32	3.58	2.44
No. Requests per Session	3.87	2.96	4.14	3.07
No. Date Changes	2.67	3.32	3.64	3.39
No. Route Changes	0.48	1.23	0.57	1.37

Since we cannot observe the actual time site visitors spent searching we decompose search into search sessions to get a better understanding of how actively customers searched. In line with previous literature on consumer browsing behavior, a new search session begins if a request is made after an idle period of 30 minutes or more (Sismeiro and Bucklin, 2004; Cateledge and Pitkow, 1995). Table 2 summarizes the search behavior of all visitors and purchasers at the website. Table 2 highlights the fact that on average purchasers searched more actively within a session compared to all visitors. There is also evidence of consumer heterogeneity in the amount of search, while approximately 25% of the sample made 2 search requests, a few customers made more than 10 search requests (see Figure 4.1).

Figure 4.1: Distribution of Number of Search Requests

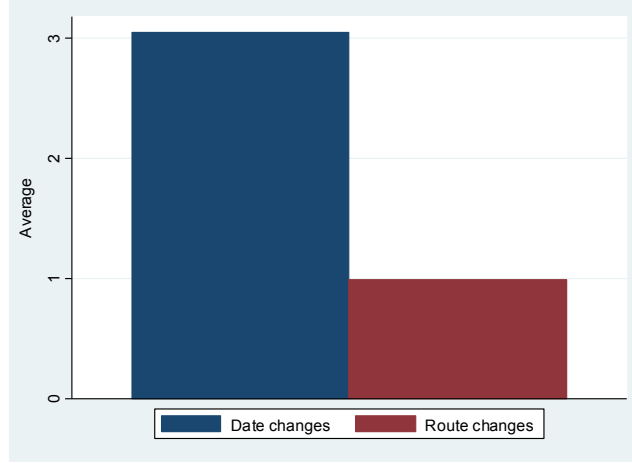


Visitors in our sample on average started 3 new search sessions, which means that on average customers searched for over 1.5 hours. Within each session customers made 3.87 requests on average. We also find that 50% of repeat search takes place on the same day. Our detailed data set allows us to observe what costumers requested at each occasion. The data reveals that customers tried to find better deals by changing their travel dates as opposed to their route. 70% of the visitors did not change route, while only 7% of visitors did not change dates. Hence, the data indicates that customers were aware of the variability in prices and changed dates to find better deals.

As is the case with online search data our data is limited in terms of demographic information about customers Brynjolfsson et al. (2010). We can only distinguish customers on the basis of their observed search behavior, for the purpose of our study this is not a serious limitation as our focus is on identifying consumer search preferences.

The data reveals that customers do not always purchase the lowest price option available. Approximately 56% of the purchase occasions consumers purchased at the lowest session price. This is inline with observed industry behavior, according to PhoCus Wright (2004) 60% of airline customers

Figure 4.2: Degree of Flexibility



purchase at the lowest price. This pattern highlights the need to focus on pre-purchase behavior to understand how in addition to price sensitivity, the search environment, search effort and flight characteristics influence consumer preferences.

5 Modeling Approach

Our modeling approach is premised on the fact that a visitor's decision to purchase or search is a function of underlying preferences and the search environment. Our primary purpose is to predict and understand online purchase behavior within a website, for a product characterized by high levels of price volatility.¹² We use our individual level data on both observed choice sets and search behavior to inform our model of pre-purchase and purchase behavior.¹³

¹²

We do not model the decision to select a website as this has already been explored by extant literature (Nair et. al., 2010a). Since there is evidence that across site search is limited, we focus on the dynamics of search within a site and its implication for OTAs.

¹³

In our context the traditional sequential search model is not directly applicable. Given the large number of flight options displayed to customers it is unrealistic to assume consumers have unlimited recall, similarly, given the as the options available to customers could be unavailable in subsequent search requests.

At a given search occasion we assume that site visitors are looking for a flight which is a combination of a specific route and travel date. Site visitors are aware of the most suitable flight in terms of flight characteristics and may have carrier preferences but are uncertain about the price. We therefore, assume that when searching consumers are looking for the best price and trying to resolve price uncertainty. Unlike traditional sequential search models (e.g. Kim et. al., 2010) site visitors do not decide whether they will search for an additional carrier, rather they search for a combination of date and route and are unaware of the number of flight options that will be displayed as a response of their request. Site visitors vary in terms of their flexibility for date and route preferences and the amount of search they are willing to undertake. Our random coefficients approach allows us to capture this heterogeneity.

We model a site visitor's decision to make a purchase, as a two stage process. In the first stage, the visitor has the option to (i) make a purchase, (ii) make another search request at the website or another website. Visitors decide based on the information gathered, future price expectations, search effort and flight characteristics. In the second stage, given the decision to make a purchase, visitors decide which airline to choose. The two stage decision process can be summarized in Figure 5.1.¹⁴ The two stage approach separates the decision to purchase and carrier choice, which as we show below are affected by different factors. By jointly estimating the choice and incidence decisions we avoid the problem of endogenous choice sets, as we estimate carrier choice conditional on the decision to purchase. Since, prices in the airline industry change frequently, options searched in one search may no longer be available in the next search,¹⁵ therefore, we do not use a consideration set approach to model the size of a consumers choice set.

The purchase probability of carrier j at occasion t is given by

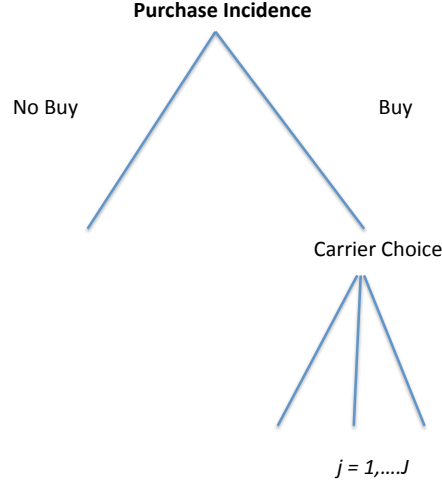
$$P_t^h(j) = P_t^h(j|purchase) \cdot P_t^h(purchase) \quad (5.1)$$

Whereby, the probability that visitor h chooses carrier j at search occasion t is the product of

¹⁴The hierarchy of the decision tree is purely analytical, the consumer could make the brand choice decision before deciding to purchase and vice versa.

¹⁵Koulayev (2010) in his study of consumer search for hotels notes that considering a single search decision is advantageous as in markets with rapidly changing prices, consumers may not be able to record prices in their previous search requests.

Figure 5.1: Two Stage Decision Process



the probability of purchase incidence and the conditional choice probability. We describe the two stages in greater detail in the following sections.

5.1 Purchase Incidence

At a given occasion t visitor h 's indirect utility of making a purchase is defined as:

$$V_{ht} = U_{ht} + \varepsilon_{ht} \quad (5.2)$$

where U_{ht} has the following specification:

$$U_{ht} = \lambda_{h1}IV_{ht} + \lambda_{h2}EIV_{ht+1} + s_{ht} + \delta_{ht} + \gamma_{ht} \quad (5.3)$$

We assume the outside option of no purchase to have a utility of 0. The visitor continues searching if utility from buying now is less than the utility from postponing purchase (i.e. $U_t^h < 0$).

The indirect utility of buying now, U_{ht} is based on the category value of purchase IV_{ht} , expected future utility of buying later EIV_{ht+1} , search effort s_{ht} , information gathered during search δ_{ht} and observed heterogeneity γ_{ht} .

We assume the error term ε_{ht} to have an extreme value i.i.d distribution which gives us the following closed for expression for the probability of purchase:

$$P_t^h(purchase) = \frac{\exp(V_{ht})}{1 + \exp(V_{ht})} \quad (5.4)$$

Since consumers are looking for the best available flight, the decision to purchase now vs. later depends on the current and future category value. IV_{ht} is the inclusive value parameter which captures the attractiveness of making a purchase based on carrier specific characteristics and price. Formally, $IV_{ht} = \ln \sum_j e^{u_{hjt}}$, where u_{hjt} is the deterministic component of the utility of visitor h 's indirect utility of carrier j at occasion t . In addition, we assume a visitor's decision to buy now vs. later depends on his expected future utility. The term EIV_{ht+1} denotes the expected future value of purchasing at occasion $t+1$. Analogous to IV_{ht} , EIV_{ht+1} is similar to the inclusive value term, except the utility is based on future price expectations.¹⁶ Formally, $EIV_{ht+1} = \ln \sum_j e^{E(u_{hjt+1})}$ we describe the future price expectations in detail in section 5.2. We expect current category value to have a positive impact on purchase incidence while higher expected future value is likely to result in a delay in purchase.

A visitor's decision to stop search may depend on the strong preference for the viewed products or due to underlying search costs. It is therefore, difficult to disentangle the two effects (Koulayev, 2010). We use our rich data set on consumer search behavior to inform our estimates of the heterogeneity amongst customers regarding search costs. We assume that individual search decisions are reflective of their underlying search cost distributions. Therefore, we include the following search actions taken by individuals as determinants of purchase incidence.

¹⁶

We assume that flights are fairly homogenous and the best option consumers are looking for is the cheapest flight. Therefore, we abstract away from expectations regarding the future draw of airlines that the consumer will see in the search request.

$$s_{ht} = \alpha_1^h Sess + \alpha_2^h Req_Sess + \alpha_3^h \Delta Date_{ht} + \alpha_4^h \Delta Route_{ht} \quad (5.5)$$

In the context of online search the main search costs include the time spent browsing and the effort involved in changing the search criteria, i.e. changing route and date. We include the number of search sessions (*Sess*) as determinant of purchase incidence. This allows us to better understand the temporal element of search. Site visitors who allow considerable time to pass between their searches experience temporal price variation. We expect most purchases to result within a single search session as visitor's are likely to be more targeted on gathering information. When customer's return after 30 minutes the variability in prices may increase the degree of uncertainty and may deter purchase. We use the number of search requests made within a session (*Req_Sess*), as a measure of a visitors involvement in the search process¹⁷. The number of requests is a better measure of the time spent searching as it is not affected by noise associated with consumers attending to other activities while their browser is left open. The more searches made without idle time also shows that the visitor places a high value on time and therefore, might have higher costs of search compared to someone who devotes more time to search. We also include variables capturing whether site visitors changed route or travel dates while searching ($\Delta Date_{ht}$ and $\Delta Route_{ht}$). We expect casual browsers with low search costs to change their route more frequently as they do not have concrete travel plans, while changing travel dates suggests consumers are looking for cheaper alternatives.

Traditionally it has been stressed that consumers search to reduce uncertainty regarding the item to be purchased (e.g. Hansen, 1972) and the greater the uncertainty the higher will be the amount of search (Lanzetta 1963). To capture the impact of uncertainty arising from price volatility, on consumer search we employ a measure of price variance; we include the standard deviation in ticket prices for each search request as a covariate in our model. For each leg of the journey we compute the standard deviation in response prices displayed to customers. In line with behavioral

¹⁷ Sismeiro and Bucklin (2004) also divide consumer search into similar session in their study of online browsing

theorists we consider the following specification for δ_{ht} :

$$\delta_{ht} = \mu_1^h \sigma_{ht} + \mu_2^h n_{ht} \quad (5.6)$$

We include σ_{ht} to capture the impact of spatial price variation on the amount of search. Parameter n_{ht} is the average number of flight options for each carrier displayed to customers after each request.¹⁸ Punj and Staelin (1983) find that the amount of information influences consumer search decision. We posit that consumers search in order to expand their choice set, therefore, the more options visitors are displayed the less likely they are to continue searching.

Several psychologists and behavioral theorists study the impact of individual characteristics, previous experience, environmental variables, time availability and size of the consideration set on the patterns of consumer search (e.g., Urbany et al. 1989; Beatty and Smith, 1987; Lanzetta, 1963). γ_{hjt} denotes individual characteristics, we include the following variables to capture observed heterogeneity in consumer search behavior:

$$\gamma_{hjt} = \varrho_1^h OTA_{ht} + \varrho_2^h Trip_{ht} + \varrho_3^h Experience_{ht} + \varrho_4^h Day_Req_{ht} + \sum_{i=1}^{11} \varrho_{5i}^h Route_{hti} \quad (5.7)$$

where:

$OTA_{ht} = 1$ if customer was directed from the main travel agency website, 0 otherwise,

$Trip_{ht} = 1$ if customer is searching for a round trip, 0 otherwise,

$Experience_{ht} = 1$ if prior booking experience within the past 1 year, 0 otherwise,

$Day_Req_{ht} = 1$ if customer searched between 8 a.m and 6 p.m, 0 otherwise,

$Route_{hti}$ = dummies indicating route requested

$\varrho_1 - \varrho_{5i}$ = parameters to be estimated

Demographic variables have often been used to control for consumer heterogeneity, however we had

¹⁸We also included the total number of flight options as a covariate, however the average options provides better fit.

very sparse data on demographics.¹⁹ Instead we include variables capturing observed heterogeneity. Observed differences in consumer behavior is a more accurate control for consumer heterogeneity than demographics, as it is likely that customers within a household may exhibit considerable variation in search behavior. Another advantage of our observed heterogeneity variables is the fact that they change over time. For instance, if at one occasion the consumer searches during the day but on the next occasion he logs on during the night, we are able to account for this difference in behavior.

5.2 Carrier Choice

At each search occasion the visitor has the option to select between several differentiated airline carriers. We assume that consumers have certain brand preferences for carriers operating on their selected routes. Since multiple flights are operated by a single carrier on a given route and date combination, we aggregate the flights to the carrier level. We consider the following specification for flight characteristics which influence consumer's carrier choice:

$$u_{hjt} = \xi_j^h + \beta_1^h P_{hjt} + \beta_2^h Flight_Duration_{hjt} + \beta_3^h Flex_Time_{hjt} \quad (5.8)$$

Carrier choice depends on consumers' inherent preference for carriers measured by carrier specific fixed effects ξ_j^h , the average price of the carrier P_{hjt} , and carrier characteristics. Following principles of utility maximization we expect the flight with the highest utility to be chosen. The total utility from a particular carrier is the sum of the deterministic component and an unobserved component such that:

$$v_{ht} = u_{ht} + \epsilon_{hjt} \quad (5.9)$$

The unobserved component of utility denoted by i.i.d error term ϵ_{hjt} . which gives us the following

¹⁹The lack of demographic data is not specific to our data set. This is a characteristic of most online data sets as noted by Bucklin and Sismeiro (2009).as noted

conditional choice probability of carrier j being selected at occasion t .

$$P_t^h(j|purchase) = \frac{\exp(v_{jt}^h)}{\sum_j \exp(v_{jt}^h)} \quad (5.10)$$

We include the average ticket price for each carrier operating on the searched route and date as a measure of expected expenditure in the carrier choice utility. In addition, consumers may select different carriers based on the availability of non stop flights. Therefore, we include a measure of the average flight duration for each carrier $Flight_Duration_{hjt}$.²⁰ The variable is computed as the total time taken for the journey, for round trips this variable is the sum of the travel time for both legs of the journey. A priori we expect customer's would prefer carriers with shorter journey times. We also include a flight time dummy $Flex_Time_{hjt}$ to capture the convenience of the flight, for instance consumer's might prefer flights during the day as it is easier to commute to the airport, as opposed to flights in early in the morning or late at night. Hence, we expect flights with convenient times will be preferred by customers. Table 3 provides summary statistics for covariates used in the model

Table 3: Summary Statistics for Covariates

Variable	Mean	Std. Dev.	Min	Max
Price	125.45	56.47	18	975
Avg. Session Price	136.99	54.97	18	739
Standard dev. In Prices	24.20	30.41	0	866
Days to Departure	13.36	8.67	1	31
No. Flight Options	2.36	1.13	1	15
No. Search Sessions	3.04	2.32	1	17
No. Requests per Session	3.87	2.96	1	26
Flight Duration in hours	2.39	0.83	1	5
Search in Day Dummy	0.58	0.49	0	1
Previous Booking Experience	0.08	0.50	0	15
Customer directed from OTA	0.15	0.36	0	1

We further define the expected future utility at time $t + 1$ as follows:

²⁰Carriers with more non-stop flights on the route would have lower average flight duration.

$$E(u_{hjt+1}) = \xi_j^h + \beta_1^h E(P_{hjt+1}) + \beta_2^h E(Flight_Duration_{hjt+1}) + \beta_3^h E(Flex_Time_{hjt+1}) \quad (5.11)$$

We assume that consumers only form expectations regarding prices for the flights they have observed. Therefore $E(Flight_Duration_{jt+1}) = Flight_Duration_{jt}$ and $E(Flex_Time_{jt+1}) = Flex_Time_{jt}$, i.e. carrier characteristics do not change across time.²¹ Thus, at a given search occasion consumers decide whether they would purchase the available options at P_{hjt} or whether they would continue searching to consider $E(P_{hjt+1})$ in the future, given that flight characteristics remain the same. We outline alternative models of expectation formation in the following section.

5.3 Price Expectations

Following Zhang (2012) we model expected future price as a reference price that influences the purchase incidence decision. Unlike traditional reference price models we do not formulate expectations for current purchase decisions, instead we build expectations for future prices. We assume that visitors form expectations of future prices informed by past experience and information gathered during search. We compare three alternative methods of expectation formation; expectations with learning, rising price expectations and rational price expectations.

5.3.1 Price Expectations with Learning

We assume that consumers search in order to learn about the price process and they update their expected price after every search request t , where $t = 1, \dots, T_h$. T_h denotes the number search requests for each individual.²² As search progresses consumers update their price expectations such that:

$$E(P_{hjt+1}^{learn}) = \alpha E(P_{hjt-1}^{learn}) + (1 - \alpha) P_{hjt} \quad (5.12)$$

²¹Zhang et. al. (2012) make similar simplifying assumptions regarding feature and display for packaged goods, and only allow consumers to form expectations regarding future prices.

²²It should be noted that price expectations are made for each trip, when a consumer searches a new trip after making a booking t is set to 0. Hence, T_h is the number of search requests made for a particular trip by visitor h .

For each search request $E(P_{jt+1}^{learn})$ is computed as the weighted average of the price expectation in the last request and the current price where α is the weight assigned to prior price expectations.²³ Customer recall is considered to be a first order Markov process, this assumption is in line with (Montgomery et. al., 2004) who define the browsing of customers to be dependent on only the last decision and not the entire history. At the initial search request we assume consumers have some beliefs about the price of a ticket based on their past booking experience. Since we cannot trace consumer behavior for more than a period of 2 months (as cookies are deleted) we use the extensive bookings data base to estimate the relation between price, time till departure, seasonality, weekend, routes and carrier specific effects

$$P_{jt'}^{book} = \varpi_0 + \omega_1 \text{Departure}_{jt'} + \omega_2 \text{Weekend}_{jt'} + \sum_{i=1}^{11} \omega_{2+i} \text{Month}_{ijt'} + \omega_{14} \text{Carrier}_{jt'} + \sum_{k=1}^{11} \omega_{14+k} \text{Route}_{kjt'} \quad (5.13)$$

where $t' = 1 \dots T'$ is the occasion at which a booking for carrier j was made. We estimate the coefficient vector based on information on transaction prices for flights booked in July 2004 till April 2006. 145,829 bookings were used to estimate the coefficients of the price equation. Based on the estimated parameters we predict the initial price estimate for each visitor's first search request for a particular trip. Hence, the reference price for each user at the first search occasion is the predicted value $\left(E(P_{hj1}^{learn}) = \widehat{P_{j1}^{book}}\right)$. The initial price expectation allows customers to have a prior belief about the prices before they begin the search process.²⁴

5.3.2 Rising Price Expectations

It is a common belief that airlines charge higher prices for tickets purchased only a few days

²³ We use grid search to estimate the optimal value of α the weight placed on previous prices vs. current prices. search provide better fit.

²⁴ Predicted prices were also used to define the initial price expectation for the first time a carrier appeared in search results. For instance, if carrier 2 appeared for the first time on search occasion 3, the initial price expectation is defined by the predicted value.

prior to departure as the demand for these customers is relatively inelastic (Carlton and Perloff, 2000) and the cheapest seats are the ones to be sold first (Pender and Baum, 2000). Therefore, in line with these beliefs we assume that waiting and not buying can result in price increase from one search occasion to the next. To incorporate these rising price expectations we assume that consumer expectations about rising prices are drawn from a truncated normal distribution, where the truncation point is set as the current price for each carrier observed by the searcher.

$$E(P_{hjt+1}^{rise}|P_{hjt}) = f(\overline{P_{hj}}, \sigma_j, P_{hjt}) \quad (5.14)$$

where $f(\overline{P_{hj}}, \sigma_j, P_{hjt}) = \frac{\phi(\overline{P_{hj}}, \sigma_j^2)}{1 - \Phi(P_{hjt}, \sigma_j^2)}$. Hence, at every occasion consumers expect prices to increase in future. Therefore, $E(P_{hjt+1}^{rise}) > P_{hjt}$.²⁵ We assume that consumers form an expectation about the price on the next search occasion based on the prices they observe in the current search. It should be noted that while consumers expect prices to increase in future, they are sophisticated enough to adjust their price expectations downwards if they see a decline in the price. For instance, if consumers saw a price of \$50 at occasion $t = 1$, he would expect that at $t = 2$ the expected price would be greater than \$50, i.e. $E(P_{hj2}^{rise}) > \$50$. However, if at $t = 2$ the observed price was \$30, the consumer will adjust his expectation such that $E(P_{hj3}^{rise}) > \$30$. For each carrier the moments of the distribution were based on the mean of all searched prices and the standard deviation in these prices.²⁶

5.3.3 Rational Price Expectations

Under the rational price expectations specification, we assume that consumers know the distribution of prices. This assumption has been frequently made in models of consumer search (e.g.

²⁵

Koulayev (2010), makes a similar assumption regarding price expectations for ordered search results for hotels. In Koulayev's (2010) model consumers cannot observe the prices on the second page, therefore, they make an assumption regarding the prices on the next page of results, before deciding whether to click or not.

²⁶ We also used the extensive bookings database to compute the moments but did not find any statistical difference between the two measures.

Kim et. al., 2010). We assume that visitors have prior knowledge about the relation between price and flight attributes. We further assume that consumer expectations are based on their past booking experience. We establish the relation between prices time till departure, seasonality, weekend, routes and carrier specific effects.

$$E(P_{hjt+1}^{prational}) = \widehat{P_{jt}^{book}}(Departure, Weekend, Month, Carrier, Route) \quad (5.15)$$

We use the observed booking prices to estimate the parameters defining the relation between prices and flight characteristics. We use Eq. 5.13 to establish the link between prices and flight characteristics, based on this relation consumers can determine the expected future price of the flights they have observed in the current search request.

6 Estimation

We use a hierarchical bayesian approach to simultaneously estimate the incidence and choice models. We use the Markov Chain Monte Carlo (MCMC) sampling to generate draws from the posterior densities of model parameters. For the random coefficients distributions, we use the normal distribution as the prior and the inverse Wishart distribution for the variance. Our choice of hyper parameters is based on weak priors allowing the data to drive the results. The simultaneous estimation approach ensures that covariance is allowed among the incidence and brand choice parameters. We use 10,000 iterations for burn in and an additional 1,000 iterations to determine the posterior distribution of parameters.

6.1 Endogeneity

The error term ε_{hjt} in the carrier choice equation (Eq. 5.9) may contain unobserved factors that influence prices and consumer choice. The presence of endogeneity can seriously bias estimates of discrete choice models (see Andrew and Curim (2010), for a discussion of the importance of accounting for endogeneity in disaggregate multi stage models of demand). In the case of airlines, factors like seasonal demand or fuel price hikes might affect the price, while these factors would have

been observed by air carriers when setting price, the researcher needs to account for the impact of these unobservables on price. Another source of endogeneity could be that the error term includes flight characteristics such as the choice of the airport, which may be positively correlated with the price variable due to airport taxes. Therefore, by accounting for endogeneity we take care of these factors. Our paper is one of the few papers that accounts for endogeneity in multi-stage decision models. We use a two stage instrumental variable approach in which the first stage we regress P_{jt} on a set of instruments Z_{jt} and flight characteristics X_{jt} i.e.,

$$P_{jt} = \phi_{0j} + \phi_{1j}Z_{jt} + \phi_{3j}X'_{jt} + v_{jt} \quad (6.1)$$

The instrument Z_{jt} is the mean price of all other available carriers as instruments (\bar{P}_{jt}). In addition we use the flight characteristics X'_{jt} which have not been included in the final choice model to account for any omitted variable bias. X'_{jt} includes weekend dummy, days till departure, journey distance month and route dummies. In the second stage the predicted price \hat{P}_{hjt} is inserted in equation (5.6) such that the carrier choice utility is defined as :

$$u_{hjt} = \xi_j + \beta_1^h \hat{P}_{hjt} + \beta_2^h Flight_Duration_{hjt} + \beta_3^h Flex_Time_{hjt} \quad (6.2)$$

The predicted price is free from any endogeneity bias arising from the correlation between unobserved factors and the error term. To the best of our knowledge ours is the first paper which accounts for endogeneity in a model of consumer pre-purchase behavior. In models with more complex sequential search models where error terms enter the model in a non-linear way accounting for price endogeneity is not straight forward and it has been assumed that the unobservable component of utility is uncorrelated with the error term (e.g. Koulayev, 2010 and Kim et. al., 2010).

7 Empirical Results

In this section we report the main empirical findings and compare the predictive ability of the proposed model. We estimate different specifications of the expected price; constant reference prices,

expectations with learning, rising price expectations, and rational expectations a base model with no search, as well as a base model with no search . In addition, we calibrate the model on the entire sample of site visitors and a purchaser only data set. By estimating the alternative benchmark models we are able to empirically measure the gains from our proposed modeling approach. Table 4 presents a comparison across the three specifications for price expectations.

Comparison of in sample fit based on Bayesian Information Criterion (BIC) across the three expected price specifications suggests that the model with consumer learning best explains the observed search behavior. The weight attached to current session prices (α) was estimated as 0.7, indicating consumer give more weight to current prices when forming expectations. It is not surprising that the model with rational price expectations is the worst performing in terms of fit, due to the uncertain prices in the airline industry, prices seldom conform to straight forward price rules, therefore, expectations that link future prices to flight characteristics is the least accurate model.

To check the robustness of our results we compare the full search model with a base model without search and the search model calibrated on a a subset of purchasers. Table 5 presents the base model without search, the final model calibrated with all site visitors (full search model with learning) and a model calibrated on the sub set of purchasers. Model comparison based on the BIC suggests that our proposed model of search better explains consumer behavior compared to the benchmark model without search. Comparison of the base model and the full search models highlights that ignoring consumer pre-purchase behavior results in poor in sample fit and an under-estimation of the impact of price. In addition, the model estimated on a subset of purchasers has an insignificant price coefficient. Due to the differences in sample size we cannot directly compare the purchaser only and visitor model. We conduct tests of predictive ability of the two models in hold out samples and present the results in section 7.3. In the following sub sections we present an overview of the main results for purchase incidence and carrier choice based on the full search

Table 4: Comparison of Model of Price Expectations

	Rational Expectations			Rising Price Expectations			Expectations with Learning		
	Estimates	97.5% Confidence Interval		Estimates	97.5% Confidence Interval		Estimates	97.5% Confidence Interval	
<i>Incidence Parameters</i>									
Inclusive Value	0.724	0.7197	0.7285	0.891	0.8863	0.8951	1.043	1.0377	1.0469
Expected Future Value	0.407	0.4025	0.4120	0.169	0.1641	0.1735	-0.097	-0.1014	-0.0924
No. of Searches in Session	0.084	0.0783	0.0886	0.086	0.0803	0.0908	0.082	0.0778	0.0876
No. of Sessions	-0.165	-0.1699	-0.1604	-0.164	-0.1685	-0.1589	-0.166	-0.1700	-0.1605
Flex Date	0.218	0.2135	0.2232	0.218	0.2132	0.2231	0.217	0.2127	0.2211
Flex Route	-0.058	-0.0627	-0.0539	-0.059	-0.0637	-0.0550	-0.060	-0.0640	-0.0552
Standard Deviation	-0.046	-0.0507	-0.0398	-0.046	-0.0510	-0.0400	-0.045	-0.0500	-0.0401
Flight Options	0.609	0.6041	0.6138	0.609	0.6043	0.6142	0.609	0.6049	0.6140
Previous Experience	0.066	0.0610	0.0703	0.068	0.0634	0.0728	0.069	0.0645	0.0730
Round Trip	-0.267	-0.2718	-0.2626	-0.315	-0.3198	-0.3104	-0.326	-0.3308	-0.3215
Day Request	0.100	0.0952	0.1042	0.103	0.0985	0.1074	0.105	0.1007	0.1094
Intercept	-2.710	-2.7152	-2.7063	-2.609	-2.6143	-2.6052	-2.575	-2.5790	-2.5706
<i>Choice Parameters</i>									
Intercept Carrier 1	0.370	0.3641	0.3749	0.369	0.3635	0.3746	0.371	0.3649	0.3769
Intercept Carrier 2	-0.229	-0.2347	-0.2239	-0.274	-0.2798	-0.2690	-0.323	-0.3280	-0.3175
Intercept Carrier 3	0.247	0.2421	0.2536	0.289	0.2839	0.2953	0.334	0.3283	0.3401
Intercept Carrier 4	-0.183	-0.1886	-0.1771	-0.186	-0.1915	-0.1802	-0.207	-0.2125	-0.2016
Intercept Carrier 5	-0.032	-0.0368	-0.0262	-0.088	-0.0930	-0.0826	-0.119	-0.1232	-0.1128
Intercept Carrier 6	0.711	0.7058	0.7168	0.842	0.8366	0.8476	0.952	0.9471	0.9567
Intercept Carrier 7	0.872	0.8662	0.8768	0.782	0.7770	0.7876	0.758	0.7524	0.7624
Price	-0.009	-0.0167	-0.0014	-0.009	-0.0166	-0.0003	-0.011	-0.0177	-0.0040
Flight Duration	-0.277	-0.2822	-0.2722	-0.292	-0.2969	-0.2871	-0.300	-0.3053	-0.2954
Day Flight	-0.092	-0.0972	-0.0863	-0.108	-0.1132	-0.1023	-0.115	-0.1202	-0.1097
Log-Likelihood	-44,122.00			-44079.000			-44,045.00		
BIC	44,288.70			44,245.70			44,211.70		
N	18,136			18,136			18,136		

Table 5: Comparison of Full Search Model and Benchmark Models

	Base Model			Full Search Model			Purchaser Search Model		
	Estimates	97.5% Confidence Interval		Estimates	97.5% Confidence Interval		Estimates	97.5% Confidence Interval	
<i>Incidence Parameters</i>									
Inclusive Value	0.726	0.7253	0.7266	1.043	1.0377	1.0469	0.958	0.9512	0.9653
Expected Future Value				-0.097	-0.1014	-0.0924	-0.155	-0.1608	-0.1478
No. of Searches in Session				0.082	0.0778	0.0876	0.136	0.1271	0.1444
No. of Sessions				-0.166	-0.1700	-0.1605	-0.163	-0.1719	-0.1555
Flex Date				0.217	0.2127	0.2211	0.059	0.0518	0.0669
Flex Route				-0.060	-0.0640	-0.0552	-0.039	-0.0464	-0.0315
Standard Deviation				-0.045	-0.0500	-0.0401	-0.059	-0.0670	-0.0505
Flight Options				0.609	0.6049	0.6140	0.626	0.6187	0.6343
Previous Experience	0.867	0.8657	0.8673	0.069	0.0645	0.0730	0.101	0.0937	0.1084
Round Trip	0.011	0.0098	0.0114	-0.326	-0.3308	-0.3215	-0.051	-0.0589	-0.0431
Day Request	-0.655	-0.6553	-0.6539	0.105	0.1007	0.1094	0.105	0.0974	0.1120
Intercept	1.283	1.2825	1.2836	-2.575	-2.5790	-2.5706	-1.448	-1.4540	-1.4406
<i>Choice Parameters</i>									
Intercept Carrier 1	0.116	0.1149	0.1170	0.371	0.3649	0.3769	0.461	0.4525	0.4709
Intercept Carrier 2	-0.109	-0.1102	-0.1084	-0.323	-0.3280	-0.3175	-0.332	-0.3408	-0.3227
Intercept Carrier 3	0.202	0.2008	0.2027	0.334	0.3283	0.3401	0.273	0.2632	0.2819
Intercept Carrier 4	-0.563	-0.5637	-0.5619	-0.207	-0.2125	-0.2016	-0.352	-0.3610	-0.3432
Intercept Carrier 5	-0.025	-0.0262	-0.0243	-0.119	-0.1232	-0.1128	-0.063	-0.0709	-0.0550
Intercept Carrier 6	0.645	0.6439	0.6457	0.952	0.9471	0.9567	0.915	0.9065	0.9228
Intercept Carrier 7	-0.217	-0.2180	-0.2163	0.758	0.7524	0.7624	0.756	0.7465	0.7642
Price	-0.001	-0.0020	0.0004	-0.011	-0.0177	-0.0040	0.004	-0.0095	0.0170
Flight Duration	0.022	0.0209	0.0224	-0.300	-0.3053	-0.2954	-0.450	-0.4578	-0.4416
Day Flight	-0.084	-0.0845	-0.0828	-0.115	-0.1202	-0.1097	-0.106	-0.1140	-0.0973
Log-Likelihood	-47,355.00			-44,045.00			-43,378.00		
BIC	47,433.45			44,211.70			43,538.93		
N	18,136			18,136			12,917		

model with learning.

7.1 Purchase Incidence

The final estimates and confidence intervals for the preferred model are presented in Table 5 column 2. Overall the estimated parameters have the expected signs. Visitor actions are significantly impacted by current and future category value. When consumers expect higher future utility they are likely to forgo purchase on the current search occasion. On average we find that consumers current category attractiveness measured by the inclusive value parameter has an estimated coefficient of 1.043 while future category attractive has a coefficient of -0.097. This result suggests that a decline in current prices have a greater impact on current purchase incidence compared to an equally large discount in the future. This behavior is consistent with theories of discounted utility, as consumers value a gain at present more than a gain in the future.

The results suggest that search effort is an important determinant of purchase incidence. We find that the number of 30 minute search sessions have a negative impact on purchase incidence. This is an interesting finding which suggests that customers who return to the website repeatedly over time may have low search costs and spend more time searching. We regard this as proclivity for temporal search. However, the number of searches within a session has a positive impact on incidence, this indicates that consumers are more likely to make a purchase the more they search. When customers search repeatedly without delay they are actively involved in the search task and are hence, more likely to make a purchase. Moe and Fader (2004) find similar evidence that purchase incidence is higher the more time customers spend browsing the site. Concentrated search within a short span of time captures consumer behavior directed towards increasing the number of travel options, we regard this behavior as spatial search. We also find evidence that changes in the requested trip influences purchase incidence. Customers who change dates appear more likely to purchase as customers with serious purchase intent often change their dates to find better prices, such behavior is therefore indicative of spatial search. On the other hand, visitors who frequently change their destination appear to be casual browsers without concrete travel plans and are less likely to purchase. Comparing the magnitude of the effect of route and date changes we find that

date changes have a larger impact on purchase incidence than route changes.

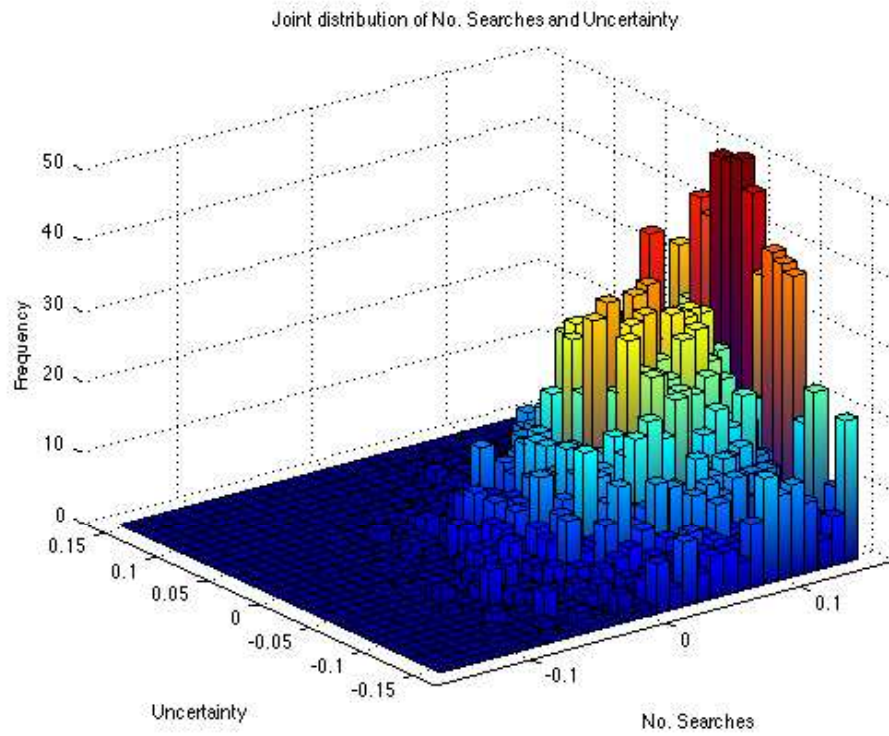
The above results regarding the impact of browsing behavior on search costs suggests that greater spatial search is reflective of high search costs and increases purchase incidence, while temporal search is indicative of low search costs and hence reduces the likelihood of purchase conversion. This is a powerful result for OTA's who can improve purchase conversion by targeting visitors engaged in spatial search.²⁷

In line with the widely accepted view that consumers search more in the presence of uncertainty (e.g. Lanzetta, 1963; Urbany et. al., 1989), we find evidence that consumers search in order to resolve uncertainty. The negative coefficient for standard deviation in observed prices indicates that consumers tend to search more when there is greater spatial variation in prices. When consumers observe volatility in product value, consumers would like to be more confident before deciding to purchase, hence the purchase threshold is higher. Despite the control for the number of options seen by customers, variance in price reduces the incidence of purchase. For site visitors flight options have a positive impact on purchase. While customers are averse to the variation in prices, greater variety reduces the need to invest time in search as they feel more confident regarding their purchase decision. Figure 7.1 shows the joint distribution of the coefficient on uncertainty and the parameters measuring the impact of search effort. Visitors who do not purchase due to uncertainty also have a positive coefficient for the number of searches within a session.

We also find that search behavior is affected by observed consumer heterogeneity. Customers searching for round trips tend to search more as they spend time finding the best flight option for both legs of the journey. Some customers search through price comparison sites while other customers directly log on to OTAs. Our results indicate that customers directly visiting the OTA are less likely to book a flight, perhaps customers are more confident about purchase when they are directed from price comparison websites. The dummy for search during the day is positive and significant, this implies that customers with a serious purchase intent log on during the day (between

²⁷ Ellison and Ellison (2009) find that obfuscation strategies can be used by firms to increase the search costs of consumers to reduce the price sensitivity and the amount of search.

Figure 7.1: Joint Distribution of No. Searches in Session and Uncertainty



8 am and 6 pm). This is valuable information for OTA's, by introducing price variation across times of the day OTA's could take advantage of the difference in purchase incidence by timing of search. In line with Nair et. al. (2010a) we find that prior purchase behavior at a site is a determinant of current purchase. Consumers who have purchased at the website before are more likely to purchase again, hence site loyalty is an important determinant of purchase incidence.²⁸ We also tested the impact of time till departure on search behavior, but found no evidence that customers are affected by the time constraint. This suggests that customers normally start search when they are certain about their travel plans and there is no evidence that customers who start searching in advance will search more than customers who begin search closer to the date of departure.

7.2 Carrier Choice

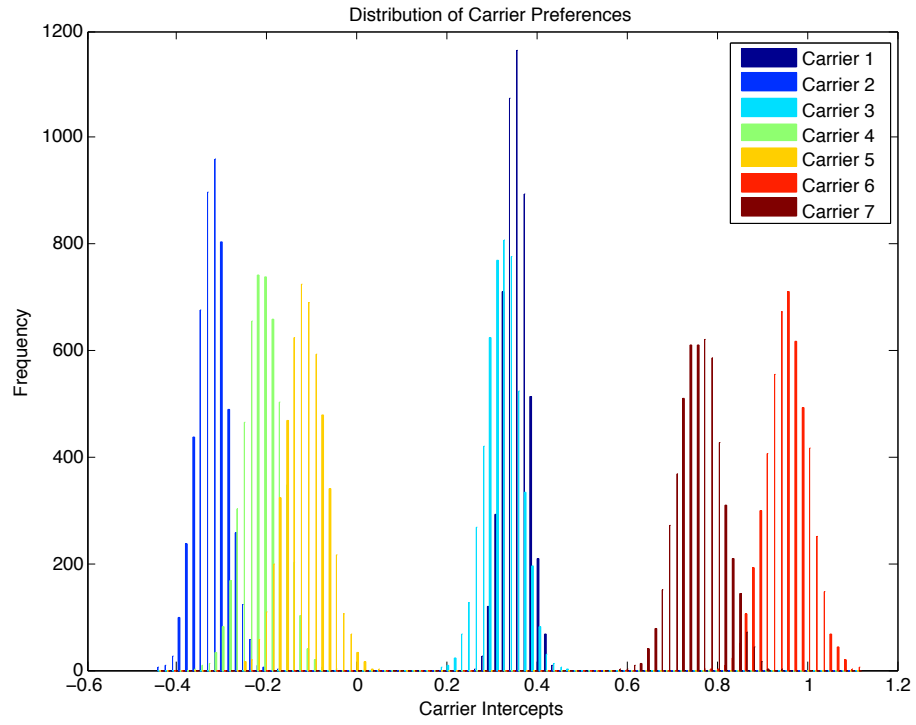
Table 4 indicates that there is considerable variation in the estimates for the carrier dummies, this indicates that some airlines are preferred over others. Figure 7.2 shows the distribution of consumer preferences for the various carriers. Carriers 2, 4 and 5 were generally quite unpopular amongst site visitors while Carrier 6 and 7 are normally preferred. This suggests consumers place great importance to carrier quality in addition to price and other observed flight characteristics.

In accordance with our expectation, when prices are high there is a greater financial risk associated with purchase, hence customers are less likely to purchase when prices are high (Punj and Staelin, 1983). Comparison of the price coefficient across various models in Table 5 reveals that the consumer price sensitivity is underestimated when search is not modeled. While the coefficient on price is positive but insignificant in the purchaser only model highlighting the fact that including all site visitors in the estimation sample improves model reliability.

We also find that customers prefer flights with short duration as indicated by the negative coefficient on the flight duration parameter. However, carriers with arriving and departing flights

²⁸In line with earlier studies of revenue management, which suggest that airline customers normally fall into two categories, business and leisure (e.g., Dana, 1998), we tested for the impact of trip type on purchase incidence, but found no difference in search patterns across the two groups. We specifically tested whether the behavior of leisure customers (i.e. customers searching for flights on weekends and customers traveling with children), behaved differently from business travelers. However, we found these variables to be insignificant and were dropped from the final model. Number of passengers also did not influence search behavior.

Figure 7.2: Distribution of Carrier Preferences



operating during different timings of the day are not preferred.

7.3 Model Validation

Out of Sample fit

We test the predictive ability of our proposed model using out of sample tests. We use the data on consumer search from April 2006, to test the predictive power of our model. The hold out sample comprised of 2,840 search requests and 757 purchases generated by 1,126 site visitors.

Table 6 presents a summary of the predictive accuracy for the hold out sample. According to Table 6, the full search model calibrated on all site visitors correctly predicts purchase incidence approximately 78% of the time, while the model calibrated on purchasers has a hit rate of 58% and the model without search is the worst performing with a hit rate of 28%. Similarly, the visitor model is more accurate than the purchaser model in predicting purchase incidence compared to the base model and the model calibrated on a subset of purchasers.

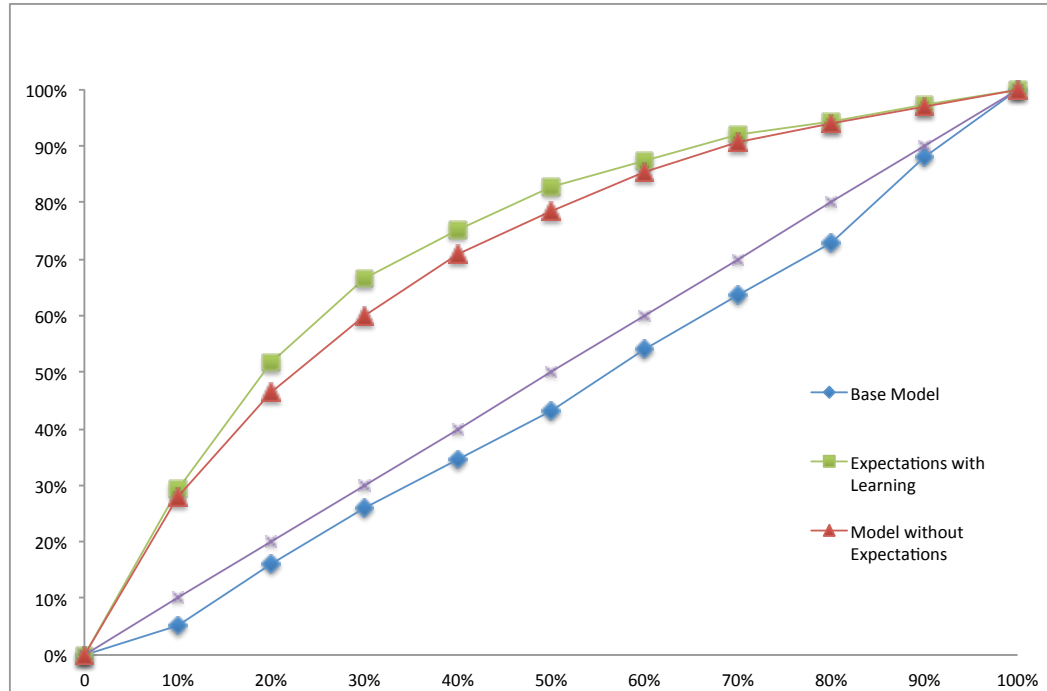
Table 6: Out of Sample Hit Rate

	Base Model	Full Model	Purchaser Model
Purchase prediction (%)	26.76%	80.30%	37.30%
Hit rate (%)	27.46%	77.68%	57.99%
Note: We use a 0.5 probability cutoff, i.e. a purchase is predicted when the probability is at least 0.5			

Based on out of sample hit rates we conclude that the full model (including both purchasers and visitors who do not make a purchase), has greater predictive power. The superior predictive ability of our preferred model highlights the need to include site visitors who do not purchase in models of consumer behavior. The behavior of site visitors who do not purchase contains valuable information which can enable firms to better predict purchase incidence. Several recent studies of consumer search behavior have been limited by the availability of data on non-purchasers and have performed a *conditional analysis* of search as they have studied the search behavior conditional on the customer eventually making a purchase (e.g. Honka, 2010; Kim et. al., 2010; Nair et. al., 2010a).

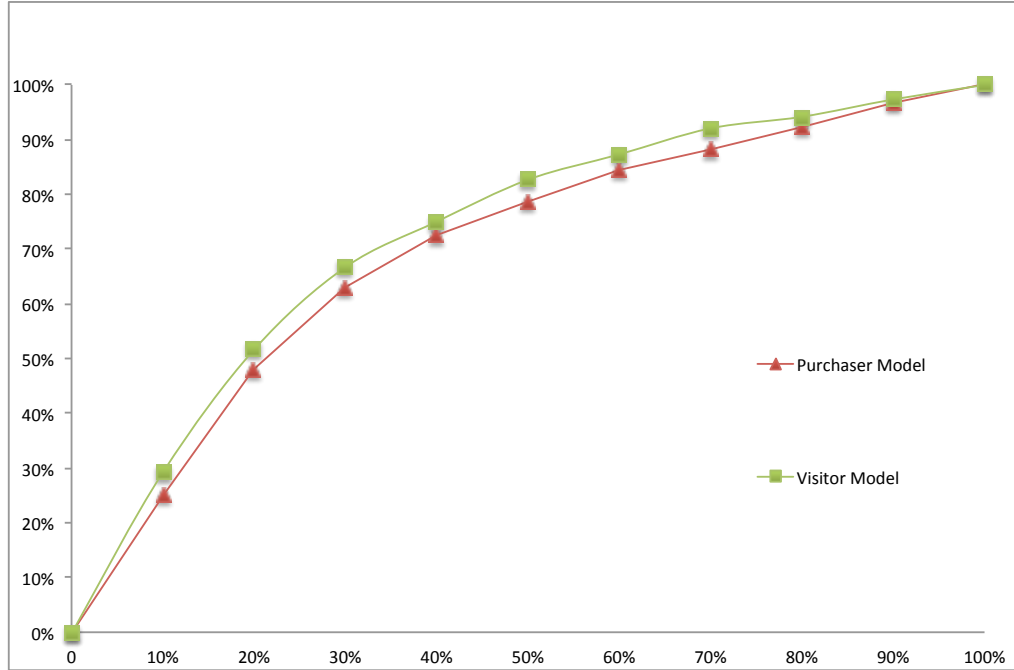
We further test the ability of our proposed model to accurately target customers. Figure 7.4 presents lift charts for the full search model, model without future price expectations and the base model. To create the charts we sorted the purchase probabilities for all holdout visitors, as predicted by the models. We then took the 10% of all (holdout) visitors with the highest predicted probability and predicted how many would make a purchase. This procedure was then repeated for 20% of the visitors, 30%, and so on. We then plotted the fraction of online purchases that each model would have been able to capture at different targeting percentages. Our proposed modeling approach, the full search model, outperformed both the base model and the model without future price expectations in terms of lift. The lift lines corresponding to the full search model are always above all others, Figure 7.4 shows that by targeting the best 30% of all holdout web site visitors we are able to capture about 67% of online buyers if we use the full search model. The base model perform poorly and only captures 26% of buyers. This suggests that including search in the model is essential to accurately predict consumer behavior. We further find evidence that consumers form future price expectations as the model without future category value only captures 57% of online buyers.

Figure 7.3: Lift Charts for Purchase Prediction (Comparison of Model with Search and Without Search)



Similarly, we find that casual site visitors contain valuable information that can help inform web site managers to better target customers. Figure 7.5 compares the performance of the proposed model estimated on all site visitors and a purchaser only sample. Again the model calibrated on all site visitors performs better than the conditional purchaser only model.

Figure 7.4: Lift Charts for Purchase Prediction (Comparison of Model calibrated on Purchasers and Visitor)



8 Conclusion

We present a joint analysis of consumer search and purchase behavior for a product categorized by high levels of price uncertainty. Complex revenue management pricing algorithms introduce uncertainty in prices across time and across airlines, as a result consumer search behavior in such dynamic environments is likely to differ from behavior in more stable industries. However, little is known about the impact of revenue management pricing on consumer behavior. Ours is one of the few studies which attempts to understand the impact of this spatio-temporal price uncertainty on consumer purchase behavior. We apply a flexible modeling approach to a rich data set on the browsing and purchase behavior of a large panel of customers visiting a leading European OTA. Our two stage model of incidence and choice does not impose restrictions on the search behavior of visitors, instead we use covariates based on information gathered and consumer actions at the

website to answer how consumers cope with the significant price uncertainty, how consumers form their expectations, and how search effort impacts purchase.

Our empirical results highlight the importance of pre-purchase behavior as a predictor of purchase incidence. We find that ignoring pre-purchase behavior results in misleading estimates and poor fit. In the context of online search for travel related products, ours is one of the first papers which highlights the need to incorporate visitors who do not purchase in a model of search. Tests of out of sample predictive power conclude that the full search model calibrated on all site visitors has greater predictive power compared to models estimated on a sample of purchasers. This suggests that the behavior of non purchasers includes important information that can help improve purchase conversion.

Our results suggest that consumers are forward looking and use observed prices to dynamically update their price expectations. These price expectations in turn determine the anticipated future utility of travel options. When expected future value is high consumers are more likely to wait and continue searching, however when expected future value is low consumers are more likely to make a purchase. However, in line with theories of discounted utility visitors place greater value on current utility compared to future utility. Our dynamic two stage model confirms that uncertainty results in greater search, we find that spatial price variation makes visitors less confident about the purchase decision resulting in greater search, while the more options available to customers the more confident they are about the decision and less time is spent searching. Consumer search costs as reflected by the investment in search effort are also important determinants of purchase incidence. In line with existing studies (e.g. Moe and Fader, 2004), consumers are more likely to make a purchase the more actively they search, however, once customers exit the website there is a lower chance of making a purchase on subsequent visits. Customers who change dates frequently are more likely to purchase at the website while customers who change routes do not exhibit serious purchase intent. Our detailed data coupled with a flexible modeling approach allows us to account for heterogeneity in customer behavior as well as possible endogeneity.

Our within site analysis of consumer information search has managerial relevance for OTAs in particular and online businesses in general. While researchers are often limited by the availability

of data, OTAs have access to detailed browsing and purchase data. Hence, OTAs can use their extensive database to incorporate the search behavior of non-purchasers to better predict purchase incidence as proposed by our research. In addition to improvements in prediction, managers can use our findings to identify the determinants of a consumer’s decision to continue or abandon search. We find that consumers are less likely to purchase in the presence of price variation across carriers, OTAs could alter the flights displayed to customers to reduce the price variation across carriers. From a website managers perspective improvements in forecasting and even small increments in purchase conversion can result in considerable growth in sales revenues.

Furthermore, our findings regarding frequency of search have important implications for OTAs. Our results suggest that consumers who actively search within a short span of time are more likely to purchase while purchase likelihood declines when customers resume search after a 30 minute interval. Currently, OTAs do not target customers while they are actively browsing, instead follow up emails and weekly newsletters with special offers are sent to encourage repeat visit. According to our findings conversion rates can be improved if websites take measures to increase customer involvement during the time they are actively searching. For instance, OTA could display special offers or recommend flights to customers who frequently change their travel dates. Since travel is not an impulse purchase, customers start active search once they are certain of their plans, therefore targeting active customers at could be more profitable for OTAs than sending weekly email alerts to all customers.

Our study has certain limitations. Our existing analysis focuses on a single product category, however OTAs sell several complimentary product categories. It would be insightful to explore how consumer search influences basket choice decisions. While Nair et. al. (2010b) study consumer basket choice across travel portals, their analysis is limited as they do not observe the impact of prices observed during search. Future research could extend our methodology to a incorporate basket choice in a multi-stage model of within site search. Another limitation of the present study is the lack of information regarding consumer behavior at competitor sites. By augmenting the existing data set with details on consumer behavior at other sites, a more holistic model accounting for both within and across site search could be calibrated.

References

- Andrew, Rick L. and Imran S. Curim, (2009) "Multi-Stage Purchase Decision Models: Accommodating Response Heterogeneity, Common Demand Shocks, And Endogeneity Using Disaggregate Data", *International Journal of Research in Marketing*, 26, 197-206.
- Andrew, Rick L. and Peter Ebbes, (2009), "Properties of Instrumental Variables Estimation in Logit-Based Demand Models: Finite Sample Results", Working Paper, University of Delaware.
- Baye, M.R., J. Morgan, and P. Scholten, (2007), "Information, Search, and Price Dispersion", *Handbook of Economics and Information Systems*, 1.
- Bell, David R. and Randolph E. Bucklin, (1999), "The Role of Internal Reference Points in the Category Purchase Decision", *Journal of Consumer Research*, (September), 128-43.
- Ben-Akiva, Moshe, (1987), "Improving Airline Passenger Forecasts Using Reservation Data", *Fall ORSA/TIMS Conference*, St. Louis, MO.
- Boyd, Andrew E. and Loana C. Bilegan, (2003), "Revenue Management and E-Commerce", *Management Science*, 49(10), 1363-1386.
- Bucklin, Randolph E. and Catarina Sismeiro, (2009), "Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing", *Journal of Interactive Marketing*, 23, 35-48.
- Bucklin, Randolph E. and Catarina Sismeiro, (2003), "A Model of Web Site Browsing Behavior Estimated on Clickstream Data", *Journal of Marketing Research*, 40(February), 249-67.
- Beatty, Sharon E. and Scott M. Smith (1987), "External Search Effort: An Investigation Across Several Product Categories", *Journal of Consumer Research*, 14 (June), 83-95.
- Briesch, Richard A., Lakshman Krishnamurthi, Tridib Mazumdar and S. P. Raj, (1997), "A Comparative Analysis of Reference Price Models", *Journal of Consumer Research*, 24(2), 202-214.
- Brynjolfsson, E., A. Dick, and M. Smith (2004), "Search and Product Differentiation at an Internet Shopbot", *MIT Working Paper*.
- Carlton, Dennis W. and Jeffrey M. Perloff, (2000), *Modern Industrial Organization*. 3rd Edn. Addison Wesley Longman: Reading.
- Catledge, Lara D. and James E. Pitkow (1995), "Characterizing Browsing Behaviors on the World Wide Web", *Computer Networks and ISDN Systems*, 27(6), 1065-73.
- Chiang, Jeongwen, Siddhartha Chib, Chakravarthi Narasimhan, (1999), "Markov Chain Monte Carlo And Models Of Consideration Set And Parameter Heterogeneity", *Journal of Econometrics*, 89, 223-248.
- Chintagunta, Pradeep, Jean-Pierre Dubé, and Khim Yong Goh (2005), "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household Level Brand Choice Models", *Management Science*, 51 (5), 832-49.

- Dana, James D. Jr., (1999). "Using Yield Management to Shift Demand When the Peak Time is Unknown", *The RAND Journal of Economics*, 30(3), 456-474.
- Davis, Paul (1994), "Airline Ties Profitability Yield To Management", *SIAM News*, 27(5).
- Degeratu, Alexandru M., Arvind Ranagswamy and Jianan Wu, (2000). "Consumer Choice Behavior in Online and Traditional Supermarkets: The Effects of Brand Name, Price, and other search attributes", *Journal of Research in Marketing*, 17, 55-78.
- Desiraju, Ramarao and Steve M. Shugan, (1999). "Strategic Service Pricing and Yield Management", *Journal of Marketing*, 63(1), 44-56.
- Fox, Edward J. and Stephen J. Hoch, (2005), "Cherry-Picking", *Journal of Marketing*, 69 (1), 46-62.
- Elmaghraby, Wedad J. and Pinar Keskinok, (2003), "Dynamic Pricing In The Presence Of Inventory Considerations; Research Overview, Current Practices, And Future Directions", *Management Science*, 49 (10), 1287-13
- Eric, J. Johnson, 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), pp. 299-308.
- Etzioni, Oren, Craig A. Knoblock, Rattapoom Tuchinda, and Alexander Yates (2003). "To Buy or not to Buy: Mining Airfare Data to Minimize Ticket Purchase Price", ACM, Conference on Knowledge Discovery in Data, Washington, D.C. New York, NY: ACM, 119-28.
- Gauri, Dinesh, K. Sudhir, and Debabrata Talukdar, (2008), "The Temporal and Spatial Dimensions of Price Search: Insights from Matching Household Survey and Purchase Data", *Journal of Marketing Research*, 45 (2), 226-240.
- Goldman, Arieh, and Johny K. Johansson, (1978), "Determinants of Search for Lower Prices: An Empirical Assessment of the Economics of Information Theory", *Journal of Consumer Research*, 5, 176-86
- Guevara, Cristian A. and Moshe Ben-Akiva, (2006), "Endogeneity in Residential Location Choice Models", *Transportation Research Record*, 1977, 60-66.
- Hansen, Fleming, (1972), *Consumer Choice Behavior: A Cognitive Theory*, New York: Free Press
- Hausman, Jerry, (1997), "Valuation of New Goods Under Perfect and Imperfect Competition", in *The Economics of New Goods*, R. Gordon and T. Bresnahan, eds. Chicago: University of Chicago Press.
- Hausman, Jerry, (1978), "Specification Tests in Econometrics", *Econometrica*, 46 (6), 1251-72.
- Heckman, James J., (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46 (4), 931-59.
- Kim, J., B. Bronnenberg, and P. Albuquerque (2010), "Consumer Search and Online Demand for Durable Goods", *Marketing Science*, 29(6), 1001-1023 .

- Lattin, James M. and Randolph E. Bucklin (1989), "Reference Effects of Price and Promotion on Brand Choice Behavior," *Journal of Marketing Research*, 26(3), 299-310.
- Lanzetta, John T., (1963), "Information Acquisition in Decision-Making" in *Motivation and Social Interaction Cognitive Determinants*, ed. O.J. Harvey, New York: Ronald Press, 239-265.
- Lewis, Matthew S., (2011), "Asymmetric Price Adjustment and Consumer Search: An Examination of the Retail Gasoline Market", *Journal of Economics and Management Strategy*, 20(2), 409-449.
- Loewenstein, George F. (1988), "Frames of Mind in Intertemporal Choice", *Management Science*, 34 (2), 200-214.
- Louviere, Jordan, et al., (2005), "Recent Progress on Endogeneity in Choice Modeling", *Marketing Letters*, 16 (3-4), 255-65.
- Mandel, Naomi and Eric J. Johnson, (2002), "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices", *Journal of Consumer Research* 29(2), 235-45.
- Mehta, N., R. Surendra, and K. Srinivasan (2003), "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation", *Marketing Science*, 22(1), 58 - 84.
- Moe, Wendy W. and Peter S. Fader, (2004), "Capturing Evolving Visit Behavior in Clickstream Data", *Journal of Interactive Marketing*, 18(1), 5-19.
- Montgomery, Alan, Shibo Li, Kannan Srinivasan and JohnC. Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data", *Marketing Science*, 23(4), 579-595.
- Nair, Ciju, Tat Y. Chan and Amar Cheema, (2010a), "Modeling Online Browsing and Purchase of Airline Tickets", *University of Washington, Working Paper*.
- Nair, Ciju, (2010), "Essays on Online Browsing and Purchase" . Electronic Theses and Dissertations. Paper 255
- Pender, Lesley and Tom Baum, (2000), "Have The Frills Really Left The European Airline Industry", *Journal of Tourism Research*, 2(6), 423-436.
- Perakis, Georgia and Ashul Sood, (2006), "Competitive Multi-period Pricing for Perishable Products: A Robust Optimization Approach", *Mathematical Programming*, 107(1-2), 295-335.
- Petrin, Amil and Kenneth Train, (2010), "A Control Function Approach to Endogeneity in Consumer Choice Models", *Journal of Marketing Research*, 47(1), 2-13.
- PhoCusWright, (2011), "Global Online Travel Overview Second Edition Report", PhoCusWright Inc. and Vividence Corporation.
- Punj, Girish N. and Richard Staelin, (1983), "A Model of Consumer Information Search Behavior for New Automobiles", *The Journal of Consumer Research*, 9(4), 366- 380.
- Rajendran, K.N. and Gerard J. Tellis, (1994), "Contextual and Temporal Components of Reference Price", *Journal of Marketing*, 58(1), 22-34.

- Ratchford, Brian T., and Narasimhan Srinivasan (1993), "An Empirical Investigation of Returns to Search", *Marketing Science*, 12(1), 73 - 87.
- Ratchford, Brian T, Myung-Soo. Lee, and Debabrata Talukdar (2003), "The Impact of the Internet on Information Search for Automobiles", *Journal of Marketing Research*, 40(2), 193 - 209.
- Santos, De los B., A. Hortacsu, and M. R. Wildenbeest (2010), "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior", *Working Paper*.
- Sa, Joao, (1987), "Reservations Forecasting in Airline Yield Management", MIT Flight Transportation Laboratory Report R87-1, Cambridge, MA.
- Sismeiro, Catarina and Randolph E. Bucklin, (2004), "Modeling Purchase Behavior at an E-Commerce Website: A Task-Completion Approach", *Journal of Marketing Research*, 41(3), 306-323.
- Train, Kenneth E., (2007), *Discrete Choice Models with Simulation*, New York: Cambridge University Press.
- Talluri, K., G. van Ryzin. 2004. The Theory and Practice of Revenue Management. Kluwer Academic Press, Boston.
- Urbany, E. Joel, Peter R. Dickson and William L. Wilkie, (1989), "Buyer Uncertainty and Information Search", *Journal of Consumer Research*, 16(2), 208-215.
- Van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, Philip Hans Franses, (2010), "Retrieving Unobserved Consideration Sets From Household Panel Data", *Journal of Marketing Research*, 47(1), 63-74.
- Villas-Boas, Miguel J. and and Russell S. Winer, (1999), "Endogeneity in Brand Choice Models", *Management Science*, 45(10),1324-1338.
- Vulcano, Gustavo, Garret van Ryzin and Wassim Chaar, (2010), "Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization", *Manufacturing & Service Operations Management*, 12(3), 371-392.
- Weitzman, Mark, (1979), "Optimal Search for the Best Alternative", *Econometrica*, 47, May, 641-54.
- Zhang, Qin, P.B Seetharaman and Chakravarthi Narasimhan, (2012), "The Indirect Impact of Price Deals on Households' Purchase Decisions Through the Formation of Expected Future Prices", *Journal of Retailing*, 18(1), 88-101.