Transportation Research Part A 82 (2015) 1–16

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

The impact of advance purchase deadlines on airline consumers' search and purchase behaviors

Susan L. Hotle [a], Marco Castillo [b], Laurie A. Garrow [a,*], Matthew J. Higgins [c,d]

[a] Georgia Institute of Technology, School of Civil and Environmental Engineering, 790 Atlantic Drive, Atlanta, GA 30332-0355, United States
[b] George Mason University, Department of Economics, 4400 University Drive, MSN 1B2, Fairfax, VA 22030, United States
[c] Georgia Institute of Technology, Ernest Scheller Jr. College of Business, 800 West Peachtree NW, Atlanta, GA 30308, United States
[d] National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 26 April 2014
Received in revised form 27 August 2015
Accepted 2 September 2015
Available online 20 September 2015

Keywords:
Air travel demand
Price endogeneity
Online search and purchase
Deadline effects

ABSTRACT

Airlines frequently use advance purchase ticket deadlines to segment consumers. Few empirical studies have investigated how individuals respond to advance purchase deadlines and price uncertainties induced by these deadlines. We model the number of searches (and purchases) for specific search and departure dates using an instrumental variable approach that corrects for price endogeneity. Results show that search and purchase behaviors vary by search day of week, days from departure, lowest offered fares, variation in lowest offered fares across competitors, and market distance. After controlling for the presence of web bots, we find that the number of consumer searches increases just prior to an advance purchase deadline. This increase can be explained by consumers switching their desired departure dates by one or two days to avoid higher fares that occur immediately after an advance purchase deadline has passed. This reallocation of demand has significant practical implications for the airline industry because the majority of revenue management and scheduling decision support systems currently do not incorporate these behaviors.

1. Introduction

Classic theories of consumer search for perishable goods predict that prices should fall as a deadline approaches. For example, the value of bakery goods and newspapers decreases over time, *i.e.*, these products are more valuable at the start of the business day than at the end of the business day. In contrast, products (or seats) in the airline industry are unique in that their value <u>*increases*</u> over time. Consequently, whereas the baker may cut prices as the business day comes to a close, consumer dynamics in the airline industry lead to the opposite effect. That is, prices tend to increase as the flight departure date approaches.

Airlines are able to induce this type of pricing behavior through the use of advanced purchase deadlines. By offering a discount fare that must be purchased by a certain deadline (*i.e.*, a minimum number of days in advance of flight departure), airlines can induce price-sensitive consumers to make their purchases further in advance of flight departure. This leaves less price-sensitive consumers in the market, which allows airlines to charge higher prices for tickets closer to departure.

* Corresponding author. Tel.: +1 (404) 385 6634.
E-mail addresses: shotle3@gatech.edu (S.L. Hotle), mcastil8@gmu.edu (M. Castillo), laurie.garrow@ce.gatech.edu (L.A. Garrow), matt.higgins@scheller.gatech.edu (M.J. Higgins).

http://dx.doi.org/10.1016/j.tra.2015.09.001
0965-8564/

In general, airlines typically sell multiple discounted products with different advance purchase deadlines. A study by Puller and Taylor (2012) found, for example, that discounted fare products represented 66% of their sample of U.S. bookings. Among these discounted fare products, 93.3% were associated with just four advance purchase deadlines: 21 days (3%), 14 days (47%), 7 days (32%), and 3 days (12%).

Even though advance purchase deadlines lead to systematic fare increases, their exact timing is uncertain. For example, the presence of the seven-day deadline does not necessarily mean that prices will increase on a flight for tickets purchased six (versus seven) days in advance of departure. This is because revenue management systems determine how many tickets of a particular product should be offered for sale. For flights in which it is expected that a large number of consumers will arrive in the last week prior to departure, the revenue management system will recommend selling a limited number of discounted tickets. From the consumer's perspective, this means that the discounted product with a seven-day advance purchase deadline will sell out more than seven days in advance of departure. As this example shows, the presence of advance purchase deadlines combined with demand fluctuations induces price uncertainty in markets. Further, variation in prices can be particularly high in markets served by both low cost and legacy carriers due to misalignment in product offerings. This misalignment is caused by low cost carriers selling (only) one-way fares and legacy carriers offering a mix of one-way and round-trip fares.

In this paper, we examine how consumers respond to these advance purchase deadlines and associated price uncertainties induced by these deadlines using multiple datasets from an online travel agency (OTA), QL2 Software (a firm that many travel and retail firms use to collect and analyze competitors' pricing information), a major U.S. airline, and the Airlines Reporting Corporation (a clearinghouse that processes all tickets purchased through travel agencies in the U.S., including OTAs). The OTA data provide information on the number of searches and purchases that occur in a market for specific search and departure dates. The QL2 Software data provide information on the fares available to consumers at the time they searched. Online search data from a major U.S. airline are used to validate results and a sample of tickets from the Airlines Reporting Corporation (ARC) is used to validate length of stay assumptions. To model the number of searches (and purchases), we use an instrumental variable (IV) approach to correct for price endogeneity and predict the number of searches (and purchases) in a market for specific search and departure dates. Our results provide insights into the impact of advance purchase deadlines on airline consumers' search and purchase behaviors.

The remaining sections are organized as follows. Section 2 reviews relevant literature to motivate why airlines offer discounted products with associated advance purchase deadlines. Section 3 describes the data. Methodology and empirical results are presented in Sections 4 and 5, respectively. Section 6 uses clickstream data from a major U.S. carrier's website to validate the key findings of the study, namely that consumer search increases immediately prior to advance purchase deadlines and new consumers enter the market over time. Section 7 discusses implications for aviation practice and Section 8 concludes by summarizing the key findings and providing direction for future research.

2. Literature review

Several studies have developed theories to explain why airline prices increase as the departure time nears. The interest is motivated, in part, by the fact that the airline industry does not fit with traditional theories of search theory that predict prices fall in markets with the arrival of homogeneous consumers. McAfee and te Velde (2006) propose a theory to explain why prices rise in the airline and other markets that: (1) face uncertain and high demand; (2) have fixed capacity that can be augmented only at a relatively high marginal cost; (3) sell perishable goods; and, (4) commit to a price schedule (and capacity) at the beginning of the selling period. The last point is applicable to the airline industry, as airlines first set their price schedules by determining what products to sell and at what set of prices. They then use revenue management systems to determine how many products to sell at each price point (Li, 2001). Airline schedules are also published at the beginning of the selling period. McAfee and te Velde (2006) show that in markets that exhibit these four characteristics, prices will rise as the purchase deadline approaches. The increase in prices over time is due to underlying consumer dynamics, and specifically the arrival of new, less price-sensitive consumers.

Many authors model aggregate demand uncertainty by assuming there are multiple consumer types with different arrival processes. In the context of the airline industry, this assumption means that price-sensitive leisure consumers tend to search and purchase fares further in advance of flight departure than price-insensitive business travelers. Li (2001) and Dana (1998, 1999a, 1999b) use an aggregate demand uncertainty framework to show that it is optimal for airlines to offer multiple products distinguished by price and advance purchase deadlines. In this case, the advance purchase deadlines serve to segment the market and can even contribute to efficient allocation of demand across flights (Dana, 1998, 1999a, 1999b; Gale and Holmes, 1992, 1993).

Airlines and researchers have also explored the use of opaque products to stimulate leisure travelers that exhibit a high degree of travel flexibility without cannibalizing revenue from business travelers. Many of these opaque products target "last minute" travelers that can purchase close to departure date and are likely to be price sensitive, but insensitive with respect to travel date and/or destination. See Fay (2008), Gallego and Phillips (2004), Lee et al. (2012), Granados et al. (2008), Jerath et al. (2010), Jiang (2007), and Post (2010) for representative articles in this area. Examining last minute opaque product sales is outside the scope of this study, as these last minute purchases are not present in our analysis database.[1]

Within the economics literature, peak-load pricing models are used to explain the efficient allocation of demand across different periods. Consistent with peak-load pricing models, advance purchase deadlines may also result in multiple price levels on flights. This can occur when products with advance purchase deadlines sell out on popular, peak-period flights but are still available for sale on less popular, off-peak flights. This effectively shifts price-sensitive consumers from peak to off-peak periods (Gale and Holmes, 1992).

In summary, the extant literature has developed several theories to explain why airline prices increase as the departure dates approach and why it is beneficial for airlines to offer discount fares with advance purchase deadlines. These theories require the presence of at least two consumer segments: one that arrives early in the booking process and is price-sensitive and one that arrives later in the booking process and is less price-sensitive. With the exception of Hotle and Garrow (2014), few studies have been able to empirically test the validity of these theories and none have been able to verify that consumers searching online close to flight departure represent newly arriving (and not returning) consumers. The presence of automated search tools and different pricing policies used by airlines further complicates the search process, and we are not aware of any studies that have examined how these factors may influence search and purchase behaviors. Our study contributes to the literature by examining these questions and providing empirical evidence that supports existing theory.

3. Data

To understand how individuals respond to advance purchase deadlines and price uncertainties induced by these deadlines, data are needed on individuals' search and purchase behaviors. Using clickstream data, researchers have developed ways to identify individual consumers and track their online search and purchase behaviors across one or more websites (*e.g.*, Bucklin and Sismerio, 2009).

In an ideal world, researchers would be able to use online clickstream data to identify all of the individual itineraries consumers viewed across multiple travel sites, along with their ultimate purchase decisions. Unfortunately, most companies do not have the resources required to extract and store this type of detailed, page-level information. As a consequence, initial studies of online search and purchase behaviors predominately focused on predicting metrics that did not require extracting detailed page content. For example, Johnson et al. (2004) and Zhang et al. (2007) developed models to predict the number of online air travel stores consumers visited over a 30-day time period. A notable exception is Brynjolfsson et al. (2010), who extracted page-level content from a major shop bot for books to show that consumers who search multiple screens are motivated by non-price factors, such as seller reputation.

Our data, which was provided by an OTA, contain information on the number of searches and purchases for a particular product. Unfortunately, detailed information on the actual set of products (or itineraries) viewed by consumers and their corresponding prices was not available from the OTA. To obtain this information a second dataset provided by QL2 Software was used. Variable definitions and descriptions are presented in Table 1 and correlations are provided in Appendix Tables A1 and A2.

3.1. OTA clickstream data

Clickstream data were collected from a single OTA's website. Although data for this study are from a single OTA, we expect our results to be applicable across major OTAs, specifically Travelocity, Expedia, and Orbitz. That is, we do not expect OTAs to return substantially different choice sets with respect to the number of airlines and number (and type) of prices associated with nonstop flights. The majority of differences across OTAs are more likely related to which connecting itineraries are included in the display and the order in which itineraries are shown to consumers (Smith et al., 2007). From a policy context, assessing potential differences across these three major OTAs is now less relevant, as Expedia acquired Travelocity in January of 2015 and Expedia made an offer to acquire Orbitz in February of 2015.

The data provide information on the number of searches and purchases for a particular product. A product is defined by a set of search parameters entered by the consumer, specifically the market (defined by a specific origin and destination airport pair), trip type (*i.e.*, one-way or round-trip), and outbound departure date. A consumer can enter more than one set of search parameters, which is represented in the database as multiple independent searches. Observations corresponding to round-trip itineraries that had an *outbound* departure date between November 15, 2007 and December 15, 2007 are included in this analysis.[2] A booking horizon of 30 days is associated with each departure date. For example, for round-trip itineraries with an outbound departure date of November 15, a panel of the number of searches and purchases occurring each day between October 16 (30 days in advance) and November 14 (1 day in advance) is created.

[1] Economic theories that seek to explain why airline prices increases as a deadline approaches typically assume two customer segments. These models assume that leisure customers will fall out of the market as we move closer to a departure date, thereby resulting in less price sensitive customers as the deadline approaches. We acknowledge that there is likely a last-minute, price sensitive segment that may include non-business travelers. Analyzing the composition of last-minute travelers, within seven days of departure, would be an interesting research problem.

[2] A Thanksgiving indicator variable is included in all models to control for any additional holiday demand.

Table 1
Variable definitions and descriptions.

Independent variables	
Searches	Number of searches on the OTA's website for a specific origin airport, destination airport, search date, and (outbound) departure date. Only round trips for a specific outbound departure date are included in the number of searches; however, multiple return dates are included
Purchases	Number of purchases on the OTA's website; note the same qualifiers used for searches also apply to purchases
Dependent and instrumental variables	
Price	Lowest nonstop round-trip fare available across all competitors selling nonstop fares in a market (in dollars). The price applies for a specific origin airport, destination airport, search date and outbound departure date. We use the nonstop fare corresponding to a one-day trip length to calculate price; the exact departure and return dates searched by the consumer (and the corresponding fares) are not known
Distance	Market distance, defined as the distance between a specific origin airport and destination airport (in miles)
Major	Number of major competitors that provide nonstop service in the market. Major airlines include American, Continental, Delta, Northwest, United, and US Airways
LCC	Number of low-cost carrier competitors that provide nonstop service in the market. Low cost carriers include American Trans Air (ATA), AirTran, JetBlue, Southwest, and Spirit
Weekend	Indicator variable equal to 1 if the search date occurred on a Saturday or Sunday and 0 otherwise
DFD	Days from departure, defined as the (outbound) departure date – search date
DFD1	Indicator variable equal to 1 if DFD equals 1, 0 otherwise
...	...
DFD 30	Indicator variable equal to 1 if DFD equals 30, 0 otherwise (DFD 30 is reference category)
Thanksgiving	Indicator variable equal to 1 if the searched departure date occurred from the Saturday before Thanksgiving to the Sunday after Thanksgiving (*i.e.*, 11/17/2007–11/25/2007), zero otherwise
Leisure	Indicator variable equal to 1 if the market is extensively leisure and 0 if the market is extensively business. This classification was based on the Borenstein Business Index, which gives the percent of business passengers arriving and departing from each Metropolitan Statistical Area (Borenstein, 2010). If either the percent business passengers arriving or departing at an airport was less than 33%, we classified the market as extensively leisure. There are 46 business markets and 14 leisure markets, for a total of 60 markets, included in the analysis
BusDes	Portion of consumers arriving to a destination metropolitan area considered to be business (in decimal format). This was defined by the Borenstein Business Index. If, for example, 70% of arriving consumers were considered business, then BusDes = 0.70
Seat	The number of seats flown in the market for departures occurring in November and December of 2007. This information is from the T-100 Domestic Segment form the Bureau of Transportation Statistics (Bureau of Transportation Statistics, 2011)
CV	Coefficient of variation (standard deviation divided by the mean) of the lowest offered one-day round-trip nonstop fares across all competitors for a specific itinerary search (defined by a particular origin airport, destination airport, search date and outbound departure date). That is, the lowest one-day round trip nonstop fares offered by competitors may differ when a consumer performs a search. This is the CV of the lowest fares offered across competitors
PopOrig	The population of the origin city as reported by the U.S. Census Bureau (2010)

This unique 30-day booking horizon provides the opportunity to analyze search and purchase behaviors as a function of advance purchase deadlines. Although we would expect the distribution of tickets associated with each advance purchase deadline to differ across markets, we would not expect the advance purchase deadline periods themselves (of 3, 7, 14, and 21 days) to change.

The distribution for the lengths-of-stay contained in the analysis database could not be calculated, as the *return* (or *inbound*) dates were not available in the OTA database. However, among those consumers who purchase a round-trip ticket, the percentage of tickets with lengths-of-stay greater than 14 days is expected to be small. To verify this assumption, we obtained a supplemental dataset from ARC of all round-trip tickets purchased through OTAs for travel in the U.S. in the fourth quarter of 2009.[3] Fig. 1 shows the length of stay distribution for markets included in our analysis for simple round-trip tickets with outbound departure dates of November 15 to December 15; 97.2% of these tickets have lengths of stay between 0 and 14 days.[4] The distribution of lengths of stay shown in Fig. 1 is similar to that reported by Brunger (2010) based on June 2006 ticketing data from Continental Airlines that found that the average length of stay was 3.36 days and 7.91 days for business and leisure passengers, respectively, with an overall average of 5.44 days. Notwithstanding this limitation, we do know that round trips included in the database have lengths-of-stay that are bounded between 0 and 331 days (the maximum number of days in advance of departure that a consumer can search and purchase a ticket).[5]

3.2. QL2 Software and Southwest pricing datasets

The OTA data provide information on the number of searches and purchases for a particular search date and outbound departure date, but does not provide information on the actual itineraries and prices viewed by consumers. To gather this missing price information, we used data compiled by QL2 Software, a company that many travel and retail firms use to collect and analyze competitors' pricing information. Within the airline industry, QL2 Software and related companies

[3] Data were not available prior to 2009.
[4] Simple round-trip tickets do not include stop-overs. Tickets with up to one outbound connection and one inbound connection were included in the analysis.
[5] Only 0 to 30 days of stay are shown in Fig. 1.

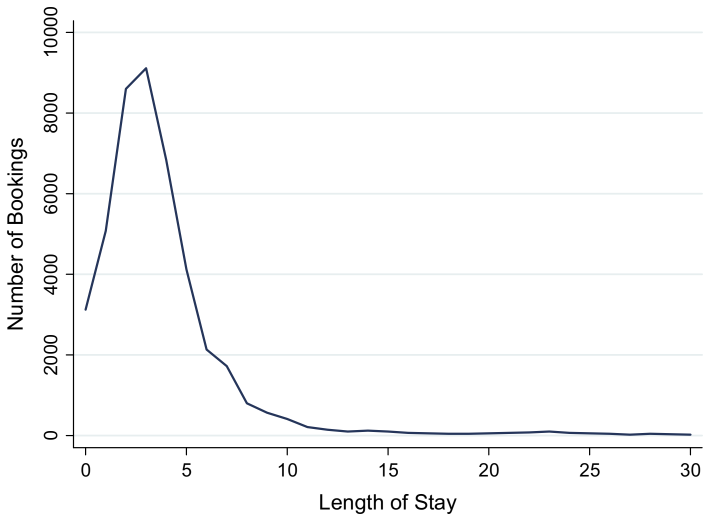

Fig. 1. Length of stay using ARC information.

can legally collect and sell pricing information for all airlines in the U.S. except for Southwest Airlines. Information about Southwest Airlines was collected directly by researchers at the Georgia Institute of Technology; see Pope et al. (2009) for additional details on this data collection effort.

The QL2 Software and Southwest Airlines pricing databases provide one- and seven-day stay round-trip prices for all nonstop itineraries in a market. Nonstop fares were obtained from each of the major airline's sites (*e.g.*, AA.com) as well as for at least one major online travel agency (*e.g.*, Orbitz). For our purposes, a pricing observation will be defined as the lowest nonstop fare that was offered by each airline flying nonstop in a specific market on the date that the website was queried and for each specific day of flight departure. The lowest fare offered was used given that "...approximately 60 per cent of online leisure travelers purchase the lowest fare they can find..." (PhoCusWright, 2004; Weinstein and Keller, 2012).

Although the lengths-of-stay vary in the OTA database, it was not feasible to collect fare information for every possible length-of-stay combination. In practice, this is a key challenge that airlines face in integrating competitive pricing data into their revenue management systems.[6] This underlying scalability issue is a major reason why airlines monitor a subset, but not all, of their competitors' prices. In practice, it is common for airlines to use automated web bots, such as those maintained by QL2 Software, to check fare availability for outbound departure dates that correspond to advance purchase deadlines. The presence of these web bots (representing firm, and not consumer behaviors) are represented as large peaks in the data corresponding to searches that are 3, 7, 14, and 21 days in advance of the outbound departure date.

Note that the models reported in this paper are based on the lowest one-day round-trip fare. As a robustness check we also tested different fare assumptions by using the lowest seven-day round-trip fare. Intuitively, we expect the results to be robust to underlying fare assumptions, as the lowest one-day and seven-day nonstop round-trip fares will be highly correlated (correlation of 0.77 in our database). Conceptually, this correlation is high because the one-day and seven-day products share the same outbound fare and only differ on their return fare. Extending this logic, we expect the lowest round-trip fares associated with any other length-of-stay to be highly correlated for a particular outbound departure date. Additional details related to the analysis of lowest fares from the QL2 Software database can be found in Mumbower and Garrow (2010).

Descriptive statistics for the lowest available one-day round-trip fares weighted by the number of searches and purchases are shown in Tables 2 and 3, respectively. A total of 46 business markets and 14 leisure markets are included in the analysis; a list of these markets is included in Appendix Table A3; corresponding airport codes are reported in Appendix Table A4. Since business consumers (*i.e.*, price-insensitive and time-sensitive) and leisure consumers (*i.e.*, price-sensitive and time-insensitive) are expected to have different behaviors, these two consumers segments are analyzed separately.

We acknowledge that we were not able to directly differentiate between an individual business and leisure consumer in the clickstream data. We were, however, able to segment the markets as "extensively business" and "extensively leisure" using the Borenstein Business Index (Borenstein, 2010). The Borenstein Business Index is derived from the 1995 American Travel Survey that provides information on the trip purpose of arriving and departing passengers.[7] If the percent of business passengers arriving or departing was less than 33%, we classified the market as "extensively leisure." All other markets were

[6] To put this in context, Delta Air Lines operates more than 5400 daily flights (Delta Air Lines, 2014). If we were to collect round-trip price information for each length-of-stay combination for each of these nonstop flights for a single departure date, we would need to collect more than 1.7 million fares. If we were to do this for all flights across Delta's entire booking horizon (that includes flights departing 0–331 days in advance), the total rises to almost 600 million fares, which includes just nonstop (not connecting) flights.

[7] Although the index publishes information at an airport level, all airports in a metropolitan area have the same index. For example, 49.5% of consumers arriving into the New York City metropolitan area are considered business passengers, and the percent of arriving business consumers is assigned to be 49.5% for New York City's three main airports: EWR, JFK, and LGA.

Table 2
Descriptive statistics for lowest available one-day round-trip fare weighted by number of searches.

	Obs	Min	Mean	Median	Max	Std. Dev.	CV
Market							
Leisure	64,602	98.00	252.86	228.00	868.00	88.07	0.3483
Business	119,002	42.00	327.17	268.80	1584.80	203.14	0.6209
Distance (in miles)							
0–250	20,131	98.00	277.34	236.00	1488.00	147.74	0.5327
251–500	38,603	42.00	273.29	228.00	1042.80	169.44	0.6200
501–750	25,126	148.00	267.07	248.80	1153.80	89.16	0.3338
751–1000	62,723	98.00	253.10	228.00	868.00	87.80	0.3469
1001–1250	18,090	130.00	361.32	326.80	1009.20	146.05	0.4042
1251–1500	18,931	216.80	528.98	392.80	1584.80	305.24	0.5770
Days from departure[a]							
1–2	2316	118.00	402.25	348.80	1584.80	245.86	0.6112
4–6	27,950	98.00	377.57	319.20	1564.80	243.96	0.6461
8–13	44,743	98.00	285.72	248.80	1153.80	149.00	0.5219
15–20	41,352	42.00	263.78	238.80	998.80	115.26	0.4369
22–30	47,755	42.00	256.58	236.00	978.80	109.31	0.4260

[a] Statistics for 3, 7, 14, and 21 days from departure are excluded as searches are dominated by automated web bot searches.

Table 3
Descriptive statistics for lowest one-day round-trip fare weighted by number of purchases.

	Obs	Min	Mean	Median	Max	Std. Dev.	CV
Market							
Leisure	2147	98.00	238.43	218.79	568.80	85.82	0.3599
Business	6956	42.00	299.03	253.80	1564.80	185.86	0.6215
Distance (in miles)							
0–250	1492	98.00	267.28	236.00	1488.00	141.87	0.5308
251–500	1746	42.00	243.11	198.40	1042.80	151.30	0.6224
501–750	1118	158.80	250.95	229.20	861.80	71.53	0.2851
751–1000	2116	98.00	238.35	218.79	568.80	85.62	0.3592
1001–1250	1172	130.00	353.14	306.00	950.80	145.86	0.4130
1251–1500	777	232.80	513.18	360.80	1564.80	333.80	0.6504
Days from departure							
1–3	1564	118.00	374.36	298.80	1564.80	244.25	0.6524
4–7	1605	108.00	331.32	263.80	1564.80	220.30	0.6649
8–14	2323	98.00	263.01	238.80	986.80	120.57	0.4584
15–21	1850	42.00	241.65	225.50	976.19	98.57	0.4079
22–30	1080	42.00	236.60	222.00	692.80	91.35	0.3861

classified as "extensively business." Further, a cutoff of 33% results in an intuitive categorization of business and leisure markets, *i.e.*, as seen in Appendix Table A3 the extensively leisure categorization includes all markets with an origin airport or destination airport in Florida, as well as Long Island MacArthur Airport (ISP) in Islip, New York. The majority of service at ISP is by Southwest Airlines. For these reasons, we use a cutoff of 33% to categorize markets as predominately leisure or predominately business.[8]

Tables 2 shows that, on average, the lowest fares searched were \$252.86 and \$327.17 in leisure and business markets, respectively. The difference is explained by business consumers searching closer to an outbound departure date, when fares are typically higher. This can be seen in the distribution of lowest offered fares by days from departure. Across all markets, the average lowest searched fare is \$256.58 for 22- to 30-days from departure and increases to \$402.25 for 1- to 2-days from departure. Also, the range and variation in fares seen by consumers in business markets is typically larger than that of leisure markets. The lowest offered fares are loosely correlated with distance, with a noticeable increase in the median and mean lowest offered fares for markets above 1000 miles. Similar relationships are seen in Table 3 when the lowest fares are weighted by purchases; the most notable difference (as expected) is that the mean and median prices are lower for purchases versus searches.

3.3. Representativeness of database

Our final dataset contains 381,607 searches (183,604 of which occurred on non-deadline dates) and 9103 purchases across 60 markets. This represents an overall conversion rate (*i.e.*, the ratio of the number of purchases to searches) of

[8] We conducted a sensitivity analysis and compared models that used cutoffs of 20%, 30%, 33%, 40%, and 50%. For models that used cutoffs lower than or equal to 33%, the results are very similar. For the 40% and 50% thresholds, the business models broke down because there were too few clusters labeled as predominantly business.

5.0% on the days not affected by web bots. The conversation rate is consistent with those reported in the literature; Moe and Fader (2004), for example, note that typical conversion rates for online retailers rarely exceed 5%.

The markets included in our analysis represent U.S. markets that are larger than average. Using the T-100 database, we ranked 4428 business markets and 2956 leisure markets that had an average demand of at least one passenger per day during November and December 2007 (Bureau of Transportation Statistics, 2011). Table 4 provides the rank for the 46 business markets and 14 leisure markets included in our analysis. Our leisure markets are drawn from the top 15% whereas our business markets were drawn from the top 51%; leisure markets had higher rankings than our business markets since the demand of leisure markets tends to be lower than that of business markets.

4. Methodology

Consistent with the extant literature, we use a linear model to predict air travel demand (*e.g.*, Bhadra, 2003; Granados et al., 2012; Mumbower et al., 2014). Specifically, we use linear regression methods to estimate the number of searches (or number of purchases) for market i with outbound departure date j that are made t days in advance of the outbound departure date. A key methodological challenge with this framework was finding a set of valid instruments to correct for price endogeneity.

Many prior studies of airline demand have failed to properly address price endogeneity and have assumed that prices are exogenous. However, in demand models, prices are endogenous because prices are influenced by demand and demand is, in turn, influenced by prices (this is often referred to as simultaneity of supply and demand). The presence of endogeneity results in a correlation between an explanatory variable and the error term (or unobserved factors) and effectively violates a main assumption required to ensure consistency (Greene, 2003).

Price endogeneity is well documented in the economics and management literatures; see for example, Guevara-Cue (2010), Train (2009), and Mumbower et al. (2014) for more comprehensive reviews of endogeneity in the air travel setting. Many empirical studies have shown that price coefficients are underestimated if endogeneity is not corrected. These include studies that estimate demand for high speed rail travel (Pekgün et al., 2013), household choice of television reception options (Goolsbee and Petrin, 2004; Petrin and Train, 2010), household choice of residential location (Guevara and Ben-Akiva, 2006; Guevara-Cue, 2010), choice of yogurt and ketchup brands (Villas-Boas and Winer, 1999), choice of a new vehicle (Berry et al., 1995, 2004; Train and Winston, 2007), and brand-level demand for hypertension drugs (Branstetter et al., 2011).

There are multiple methods that can be used to correct for price endogeneity, including the Generalized Method of Moments (GMM) that accounts for endogeneity using instruments. An Ordinary Least Squares (OLS) instrumental variable estimate of β (shown in Eq. (1)) can be used when errors are homoskedastic. However, the presence of heteroskedasticity in our data was found using a test proposed by Pagan and Hall (1983). Therefore, this study uses the Generalized Method of Moments (GMM) estimate (Eq. (2)), which includes weighting matrices to correct for heteroskedasticity:

$$\hat{\beta}_{IV} = \{X'Z(Z'Z)^{-1}Z'X\}^{-1}X'Z(Z'Z)^{-1}Z'y \tag{1}$$

$$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'y \tag{2}$$

where

W = weighting matrices,
X_{1i} = endogenous variable,
$W_1, \ldots, W_r$ = exogenous explanatory variables,
$Z_1, \ldots, Z_m$ = instruments.

Instruments must satisfy two conditions. First, the instruments must be uncorrelated with the error term. Second, they need to be correlated with the endogenous variable (Judge et al., 1985). In our context, this means we need to find instruments that are correlated with airfares (price) but not correlated with a consumer's purchase or choice of a flight.

Mumbower et al. (2014) review instruments that have been or could potentially be used in airline applications and classify these instruments into four main categories: (1) cost-shifting instruments; (2) Stern-type measures of competition and market power; (3) Hausman-type price instruments; and, (4) BLP-type measures of non-price characteristics of other products. Cost-shifting instruments help explain why costs differ across geographic areas and/or product characteristics. Stern-type measures of competition and market power focus on the number of products in the market and also the time since a product (and/or firm) was introduced into the market (Stern, 1996). Hausman-type price instruments are based on prices of the same airline in other geographic contexts (Hausman et al., 1994; Hausman, 1996). BLP instruments, introduced by Berry et al. (1995), are based on the average non-price characteristics of other products.

We use three cost-shifting instruments, two Stern-type instruments, and one Hausman-type instrument in our search and purchase models. Our cost-shifting instruments include: distance, an indicator for whether the destination is extensively business, and the population of the metropolitan area surrounding the origin airport. The first cost-shifting instrument has been used in prior studies (*e.g.*, both Hsiao (2008) and Granados et al. (2008) use distance). Intuitively, we expect costs to vary as a function of distance (or length of haul) due to the fact that costs are highly correlated with fuel and labor. Costs may also vary across airports. For example, airlines often provide additional services (most notably frequent flyer lounges and priority check-in lanes) at large airports and/or destinations that serve a large percentage of business travelers.

Table 4
Representativeness of OTA markets.

Ranking	Business		Leisure	
	Max passengers	Markets	Max passengers	Markets
1–100	222,323	5	229,525	2
101–200	91,033	9	82,071	1
201–300	68,009	7	54,842	6
301–400	54,567	5	37,167	3
401–500	45,587	2	27,280	2
501–600	38,156	3	22,078	0
601–700	32,703	0	17,914	0
701–800	29,403	0	14,485	0
801–900	25,271	1	12,344	0
901–1000	22,823	2	10,458	0
1001–1100	20,075	1	8438	0
1101–1200	18,068	2	7029	0
1201–1300	16,553	3	6207	0
1301–1400	15,002	1	5354	0
1401–1500	13,571	0	4464	0
1501–1600	12,556	0	3632	0
1601–1700	11,462	0	2933	0
1701–1800	10,385	0	2244	0
1801–1900	9567	0	1704	0
1901–2000	8727	1	1150	0
2001–2100	8012	0	681	0
2101–2200	7291	3	492	0
2201–2300	6774	1	359	0

Stern-type instruments use measures of market power by multiproduct firms and measures of competition as instruments. Levels of market power focus on the number of products in the market and also the time since a product (and/or firm) was introduced into the market. Our Stern-type instruments include the number of low cost carriers offering nonstop service in a market during the study time period and the number of nonstop seats offered in the markets for November and December of 2007 interacted with days from departure. These instruments are similar to those used in prior studies (*e.g.*, Berry and Jia (2010) use the number of all carriers offering service on a route and Mumbower et al. (2014) use the number of nonstop seats offered in a market).

Finally, we use one Hausman-type instrument. Hausman-type instruments are based on prices of similar brands, usually in different geographic contexts. In our data, we have prices for all brands (defined as nonstop flights offered across different competitors) and use the square of the coefficient of variation across the offered fares as our instrument. Note that because we are predicting the number of searches at a particular OTA website (that includes products from multiple competitors), we include fare information for all competitors in the instrument.

The instruments we use in our search and purchase models differ. In our search models, our instruments include the number of nonstop seats offered in the markets for November and December of 2007 interacted with days from departure, the number of low cost carriers offering nonstop service in a market during this time period, and the square of the coefficient of variation across the offered fares. In our purchase models, we include these three instruments and three additional ones for distance, the population of the metropolitan area surrounding the origin airport, and the percent of customers considered to be business arriving into the destination airport.

All of our instruments are valid. We used three tests to test for: (1) endogeneity, (2) the strength of instruments; and, (3) validity of instruments. First, we checked for the presence of endogeneity using the Durbin–Wu–Hausman test. Rejecting the null hypothesis indicates that the focal variable is endogenous.[9] Second, we determined the strength of the instruments using a first-stage estimation F-test. For this test, if the p-value is insignificant and/or the F-statistic is less than the critical value provided in Stock and Yogo (2005), then the set of instruments are considered to be weak.[10] Lastly, we use a Hansen's J statistic[11] to test the validity of our instruments.[12]

Finally, it is important to note that our actual prices are latent. We have assumed that the minimum price can be used as a proxy for actual prices paid. Other methods, such as that based on multiple imputations (*e.g.*, see Rubin, 1987) could be used.

[9] We find that the variable fare was indeed endogenous; *e.g.*, for the search (purchase) model the test for endogeneity returned a p-value of 0.0013 (0.0685), significantly rejecting the null hypothesis of exogeneity.

[10] For the strength of instruments test, the p-value for the search (purchase) model was 0.0006 (0.0009) thereby rejecting the null hypothesis of weak instruments. Also, using a critical value of 12.83 as outlined in Stock and Yogo (2005), we reject the null of weak instruments given a maximum size distortion of no more than 15% with an F-statistic of 17.435.

[11] Although commonly used, it should be noted that the J statistic and other tests of over-identification are inconsistent (Newey, 1985).

[12] The validity of instruments test for the search (purchase) model returned a p-value of 0.109 (0.191), which accepts that the instruments are indeed valid. In sum, our instruments are valid (and strong) across all search and purchase models. For more information on the estimation and testing of instrumental variable regressions, refer to Baum et al. (2003), Greene (2003), or and Stock and Yogo (2005).

With this method, the researcher draws prices to impute from the empirical distribution of prices available for each observation. Then for each imputation, the researcher estimates the model, correcting for endogeneity using instruments, and uses the average of the estimators obtained from the imputations. Given we are using linear regression models to predict demand, the scale problem should not occur for all the variables in the model at is occurs with the correction of endogeneity in discrete choice models (*e.g.*, see Cramer (2007), Daly (2008), Guevara and Ben-Akiva (2012), and Lee (1982)). Thus, if imputing the minimum or the mean price is indeed innocuous (and just impacts model fit), the other estimated coefficients should not change much. To determine if we need to use an imputation method, we performed two statistical tests to show that imputing the minimum or mean prices was innocuous, *i.e.*, that our assumption impacted model fit but not the consistency of estimators. These tests were based on comparing two models: the first using minimum price, the second using mean price. The model based on minimum price fit the data the best and, more importantly, the other estimated (non-price) coefficients did not change much. A Hausman test was used to compare the non-price coefficients between the two models and accepted the null hypothesis that there is no significant difference between the models based on minimum versus average prices. Next, we used a Hansen-*J* test to verify that our instruments were indeed exogenous; this test also passed. Collectively, results from these two tests suggest that our price imputation assumption is innocuous. Therefore, we argue that our assumption of using minimum price is valid. We do acknowledge, however, that caution should be used if using the results to perform elasticity calculations, as different elasticities will occur across models that use different price inputs.

5. Results

5.1. Descriptive statistics for lowest fares

Figs. 2–4 and Table 5 present information about the lowest fares, number of searches and number of purchases by days from the outbound flight departure. Combined, these figures and table help visualize the price uncertainties faced by individuals.

Fig. 2 shows how the average minimum offered nonstop fare evolves throughout the booking period in business and leisure markets. The average minimum offered nonstop business market fare was always greater than its corresponding leisure market fare. The number of days prior to departure when fares experience the largest day-to-day increases differs in business and leisure markets. In leisure markets, consumers generally see constant fares up until seven days from departure. In business markets, consumers generally see constant fares up until 21 days from departure; consumers are also more likely to see large fare increases at seven and 14 days from departure.

Increases in the average minimum nonstop fares are highly correlated with advance purchase deadlines (shown by the vertical lines in Fig. 2). An advance purchase deadline corresponds to the last day a fare would have been offered in the market. Consequently, given an advance purchase deadline at time t, we expect fares to increase at time $t-1$. Given that airlines use different pricing strategies, we also expect fares to increase at time $t-2$ for our analysis database. That is, the increase in fares two periods after a deadline can be attributed to the fact that the majority of U.S. legacy carriers use round-trip pricing whereas low cost carriers (LCC) use one-way pricing. Under round-trip pricing, a single price is quoted for the outbound and inbound itineraries, and the advance purchase deadline is associated with the outbound departure date. Under one-way pricing, separate prices are quoted for the outbound and inbound itineraries, and advance purchase deadlines can differ for the outbound and inbound itineraries.

As an example, consider an individual who purchases a one-day round-trip ticket. We assume for this example that discount product offerings have not been influenced by revenue management controls and are always available within the allowable selling period. At 14 days from the outbound departure date, the outbound and inbound fares offered by legacy and LCC carriers will have identical 14-day advance purchase restrictions. At 13 days from departure, product misalignment occurs because the legacy carrier *jointly* prices the outbound and inbound itineraries (using a 7-day advance purchase restriction) whereas the LCCs *separately* price the outbound and inbound itineraries. That is, at 13 days from departure, a consumer is able to purchase an outbound fare with a 7-day advance purchase fare and an inbound fare with a 14-day advance purchase fare from a LCC. At 12 days from departure, LCC and legacy carrier products are realigned as the seven-day advance purchase restriction applies consistently to both outbound and inbound fares. This explains why price increases associated with advance purchase deadlines occur over a two-day period in our analysis database.

Although *on average* the offered fares increase around the advance purchase deadline, this increase is uncertain and may be seen only by a small percentage of consumers. This uncertainty is mainly due to interactions among airlines' revenue management systems, pricing systems, and fluctuations in demand forecasts. Table 5 shows fare trends from the consumer perspective, specifically how often the lowest available nonstop fare available at DFD t changes on day $t-1$. For example, in going from three to two days from departure: 26.7% of business itineraries experienced an increase in fares, 13.2% experienced a decrease, and 60.1%, stayed the same.[13] However, certain periods were more likely to experience fare changes.

[13] We tested the sensitivity of results by using different thresholds to define an increase and/or decrease in fares. Specifically we defined a difference in Table 5 as "any" difference of fare (of one cent or more), but also generated results defining a difference as one in which the change was at least \$10 or at least \$15.

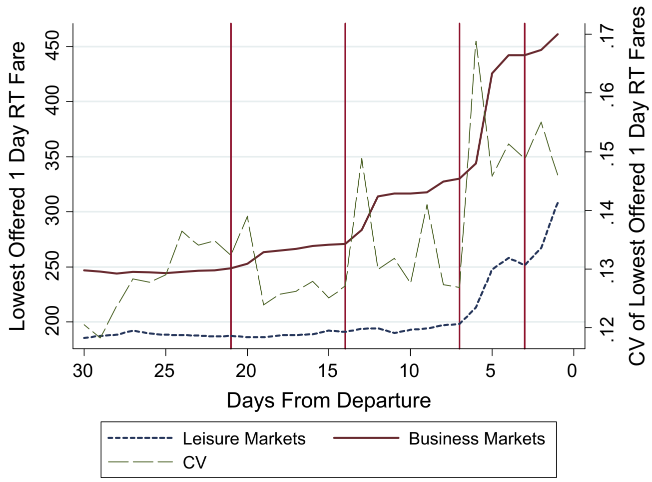

Fig. 2. How the lowest offered fare evolves in leisure and business markets.

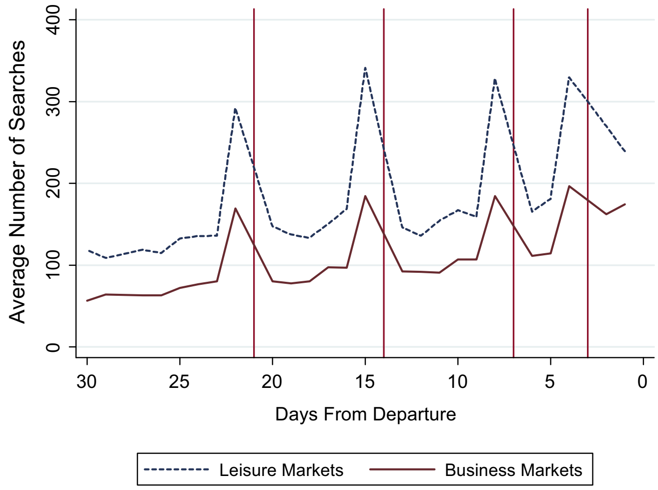

Fig. 3. Average number of searches per market.

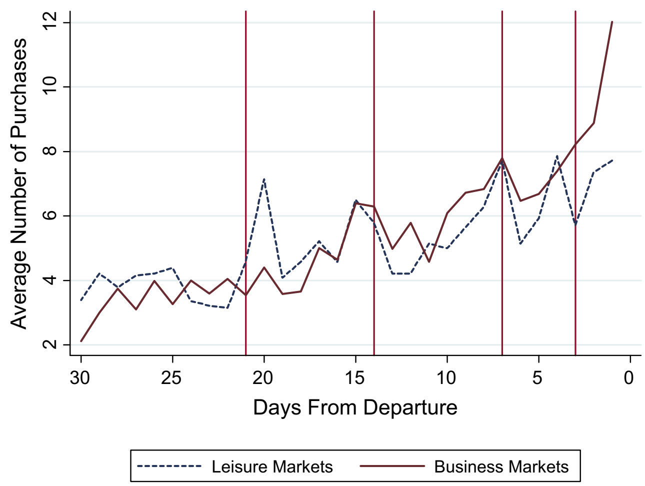

Fig. 4. Average number of purchases per market.

Table 5
How often the lowest offered nonstop one-day round trip fare changes.

DFD	How will the lowest fare available today change if I search tomorrow?					
	Business markets			Leisure markets		
	% Decrease	% Stay same	% Increase	% Decrease	% Stay same	% Increase
2	11.3	53.2	35.5	8.2	44.2	47.6
3	13.2	60.1	26.7	10.3	53.8	35.9
4	12.4	72.1	15.5	16.5	72.9	10.5
5	10.0	54.6	35.4	15.2	50.0	34.8
6	8.1	28.1	63.8	7.0	22.1	71.0
7	10.4	49.5	40.2	12.0	35.0	52.9
8	13.5	68.4	18.1	13.4	66.7	19.9
9	10.8	62.2	27.1	12.3	66.1	21.7
10	12.8	68.2	19.0	15.5	66.6	18.0
11	11.9	72.5	15.6	14.6	69.6	15.7
12	13.9	62.1	24.1	15.0	69.6	15.4
13	12.8	37.8	49.4	15.3	59.6	25.2
14	13.9	48.8	37.3	13.7	61.6	24.7
15	14.3	68.7	17.1	13.6	72.7	13.6
16	15.2	69.0	15.8	14.0	71.1	15.0
17	14.9	67.3	17.8	16.8	71.4	11.8
18	12.7	70.7	16.6	13.0	74.3	12.7
19	14.3	69.5	16.2	11.0	75.2	13.8
20	13.4	53.1	33.5	14.0	71.7	14.3
21	14.5	59.1	26.4	13.6	71.5	14.9
22	13.9	66.9	19.3	13.9	74.5	11.7
23	13.0	68.8	18.2	15.8	70.7	13.6
24	15.3	68.1	16.6	13.8	69.6	16.7
25	13.4	71.3	15.4	11.4	74.4	14.2
26	14.7	72.5	12.8	11.9	73.9	14.2
27	14.5	70.7	14.8	16.0	72.4	11.6
28	13.1	73.5	13.4	13.9	74.0	12.2
29	14.2	73.3	12.6	13.5	74.5	12.0
30	12.6	73.2	14.2	14.3	71.7	14.0

Note: Day-to-day increases in fares that occur more than 25% of the time are shaded.

The DFDs with probabilities greater than 25% of experiencing an increase are highlighted. The influence of advance purchase deadlines on inducing price uncertainties is clearly seen by the higher probabilities associated with DFDs occurring at $t-1$ and $t-2$ days after the purchase deadline. For a given deadline at time t, the probability of a fare increase is higher from ($t-1$ to $t-2$) than from (t to $t-1$) which can also be explained by the different pricing strategies of legacy carriers and LCCs.

By comparing business and leisure markets, we see that it is more likely the lowest offered fares will increase for the 21 and 14 advance purchase deadlines in business markets. This suggests airlines are aggressively using advance purchase deadlines to segment business and leisure consumers in business markets.

From a modeling perspective, these pricing uncertainties can be incorporated by including a measure of the coefficient of variation (CV) across the offered fares (see Fig. 2).[14] The CV (standard deviation divided by the mean) represents the range of prices a consumer would likely see on the OTA's website for a specific day from departure. For example, a consumer can log into the OTA on a specific search date and request an itinerary for a specific origin, destination, outbound and inbound departure dates. Typically an OTA website would return the offered fares by several airlines.

The CV represents the average distribution of these offered (non-stop) fares representing a one-day length of stay over time. We see that as the day of departure approaches, the CV increases as the offered fares become more variable across airlines. The CV appears to peak the day after an advance purchase deadline. This reflects the variation in prices caused by differences in round-trip and one-way pricing policies across carriers. The large drop in the CV near the deadline date is attributed to both the increase in the mean offered minimum fare and fewer competitors offering seats on non-stop itineraries (*i.e.*, flights sell out close to departure).

5.2. Descriptive statistics for number of searches

Although the typical airline consumer may not be aware of when the advance purchase deadlines occur and that they signal fare increases, flexible-date search tools can aid consumers in identifying these trends. Flexible-date search tools are available through both OTA and airline websites (although firms differ in how prominently they display their flexible search tools). These tools typically show fares using either: (1) a matrix format displaying the lowest roundtrip fares avail-

[14] The CV is combined for business and leisure markets, as the number of leisure market samples was small.

Table 6
Search and purchase model results.

	Search	Purchase
Price/1000	−89.679*** (23.4)	−0.976*** (0.337)
Major	6.561** (3.2)	0.137*** (0.030)
Ln(distance)	8.948** (4.5)	0.191*** (0.046)
Weekend	−2.158*** (0.49)	−0.120*** (0.014)
Thanksgiving	4.970*** (1.7)	0.021 (0.015)
Leisure	10.458 (7.7)	−0.076 (0.071)
DFD1	23.200*** (4.6)	0.550*** (0.109)
DFD2	20.747*** (4.1)	0.396*** (0.078)
DFD3	42.571*** (0.3)	0.369*** (0.072)
DFD4	21.566*** (4.4)	0.345*** (0.085)
DFD5	16.940*** (3.3)	0.334*** (0.069)
DFD6	10.140*** (1.8)	0.219*** (0.046)
DFD7	38.777*** (10.6)	0.253*** (0.044)
DFD8	11.618*** (2.2)	0.193*** (0.039)
DFD9	7.297*** (1.3)	0.212*** (0.041)
DFD10	7.508*** (1.2)	0.202*** (0.040)
DFD11	7.191*** (1.3)	0.150*** (0.033)
DFD12	6.602*** (1.2)	0.166*** (0.035)
DFD13	4.765*** (0.85)	0.111*** (0.036)
DFD14	31.82*** (9.1)	0.164*** (0.036)
DFD15	8.221*** (1.8)	0.155*** (0.040)
DFD16	3.410*** (0.83)	0.077** (0.036)
DFD17	4.014*** (0.74)	0.121*** (0.032)
DFD18	2.763*** (0.59)	0.071** (0.032)
DFD19	2.471*** (0.61)	0.046 (0.031)
DFD20	2.996*** (0.66)	0.090*** (0.032)
DFD21	31.564*** (8.9)	0.019 (0.019)
DFD22	5.653*** (1.4)	0.060** (0.025)
DFD23	1.565*** (0.51)	0.043 (0.026)
DFD24	1.243** (0.52)	0.051* (0.030)
DFD25	0.790* (0.46)	0.018 (0.032)
DFD26	0.144 (0.52)	0.068* (0.037)
DFD27	0.326 (0.47)	0.042 (0.026)
DFD28	1.183*** (0.46)	0.074** (0.034)
DFD29	0.194 (0.47)	0.042 (0.030)
Constant	−45.532* (25.4)	−0.980*** (0.235)
J-statistic *P*-value	0.109	0.191

Clustered standard errors in parenthesis.
Search model instruments for price include Seat/1000 × DFD, LCC, and CV^2.
Purchase model instruments for price include BusDes, PopOrig, Seat/1000 × DFD, Distance, LCC, and CV^2.
Shaded cells represent those used to interpret the influence of AP deadlines on search and purchase behaviors.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

able for outbound and inbound departure dates along with the three days before and after the preferred dates; or, (2) a calendar displaying the lowest one-way fares available for one month of possible departures. Legacy carriers and OTAs typically use the matrix format whereas LCCs typically use the calendar format. This is because the matrix format naturally lends itself to displaying round-trip fares whereas the calendar format naturally lends itself to displaying one-way fares.

The question of interest is how consumers' search and purchase behaviors are influenced by price uncertainties induced by advance purchase deadlines. Fig. 3 shows the average number of searches in business and leisure markets as a function of days from departure. The number of searches corresponding to 3, 7, 14, and 21 days from departure are excluded from the chart as each of these days contains approximately 30,000 or more searches. These unnaturally large spikes reflect the presence of web bots in the OTA data.

5.3. Descriptive statistics for number of purchases

To complete the descriptive analysis, Fig. 4 shows the average number of purchases in business and leisure markets as a function of days from departure. In contrast to Fig. 3, information for all days from departure is included since the number of purchases is not affected by the presence of web bots. Although the influence of deadline effects is less clear for purchase

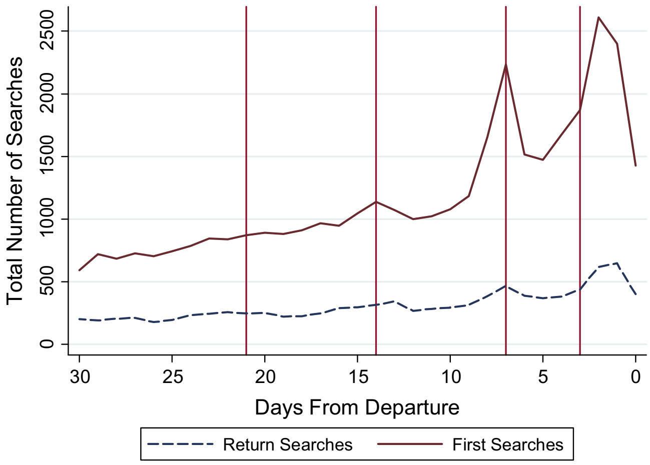

Fig. 5. Validation of search behavior using a major carrier's clickstream data.

(versus search) behavior, we do see some evidence of increased purchase activity on or just before advance purchase deadlines. This increase is most prevalent (in both leisure and business markets) for the 7-day advance purchase deadline (which typically sees a very large increase in fares). The peak at seven days followed immediately by a valley at six days suggests that consumers may be shifting their preferred departure date by one day in order to qualify for a fare that has a 7-day advance purchase requirement.

5.4. Model results

The descriptive analysis reveals many interesting patterns related to price uncertainties induced by advanced purchase deadlines and the influence of advance purchase deadlines on individuals' search and purchase behaviors. Additional insights can be gleaned from the regression models that predict the number of searches and the number of purchases, shown in Table 6. Both models include a control for whether the market was extensively leisure, account for endogeneity and control for intra-market correlations by using the clustering option to determine standard errors of the parameter estimates.

Results show that the search tends to increase as the day of departure approaches. The influence of deadlines on search behavior is seen in the shaded cells. The coefficients corresponding to the advanced purchase deadlines are large, as these dates include web bot searches. To examine the influence of deadlines on human search behavior, we need to examine the coefficients the day before and the day after a deadline. The coefficients associated with the number of searches are larger for the day immediately preceding a deadline than the day immediately after a deadline. The difference in coefficients is most pronounced for the 14-day advance purchase deadline (DFD13 = 4.765 < DFD15 = 8.221) and the 21-day advance purchase deadline (DFD20 = 2.996 < DFD22 = 5.653). This result suggests that individuals are willing to shift their outbound departure dates by a day or two in order to purchase a lower fare, particularly those customers who are purchasing further in advance of flight departure (who are more likely to be leisure customers).

Results from the regression model also indicate that search decreases on weekends but increases as the number of major competitors offering nonstop service in the market increases. Search also increases as the distance between the origin and destination airports increases, particularly in leisure markets. This is likely due to the fact that as distance increases, driving and other alternative modes of transportation become less attractive compared to air.

The results from the purchase specifications are similar to those seen for search models, *i.e.*, fewer purchases occur on weekends and more purchases occur in long-haul markets and markets in which there are more major competitors. Since purchases are made by humans (versus web bot), we can examine the DFD coefficients corresponding to an AP deadline and the day after an AP deadline. Results show an increase in purchases for the 7-day, 14-day, and 21-day AP deadlines (but not for the 3-day AP deadline). This provides additional evidence that individuals (particularly leisure travelers purchasing further in advance) are shifting their travel dates by a day to take advantage of lower fares prior to an AP deadline.

6. Validation

For validation of consumer search behavior, we use a sample of clickstream data representing consumers' search behaviors for three leisure and seven business markets from a major U.S. carrier. The departure dates represented this data overlap with those in the OTA data and the markets are similar.[15] In addition to validation, these new data enable us to track individual

[15] Due to non-disclosure agreements, we cannot reveal the markets represented in the data as they could be used to identify the carrier that provided the clickstream data.

consumers across multiple pages and multiple sessions, and identify new and returning consumers. This means we were able to identify and remove web bots from the clickstream data and we were also able to define searches as either: (1) the first set of search parameters entered by a consumer during a visit; or, (2) any set of search parameters entered by a consumer.[16]

Fig. 5 demonstrates the number of searches in the collected markets using the first set of search parameters entered by a consumer during a visit. Consistent with what we observe in the OTA data, we see spikes in the number of searches on and/or just prior to the advance purchase deadlines. The spike is most pronounced at seven days from departure. This is not surprising since more business markets are contained in the clickstream data.

Interestingly, Fig. 5 also provides supportive evidence of extant search theories that suggest prices should rise in the presence of deadlines due to the arrival of new (and less price-sensitive) consumers in the market (*e.g.*, see Stokey, 1979; McAfee and te Velde, 2006; Mantin and Koo, 2010). Further, by defining searches using just the first set of parameters versus all parameters entered by the consumer, we are able to determine that the pattern shown in Fig. 5 is not due to increased search intensities (or increases in the number of searches) as the results were similar for both search definitions.

7. Discussion

To the best of our knowledge, this is the first study that has empirically examined how advance purchase deadlines influence airline consumers' search and purchase behaviors. Several interesting findings emerge from our study, two of which represent market conditions that are not accounted for in existing theories describing consumer search under deadlines for perishable goods with fixed capacity and pre-determined pricing schedules. First, price uncertainties are induced by advanced purchase deadlines and high price dispersion is caused by misalignment of product offerings across carriers. This latter phenomenon, which occurs when a LCC offers one-way fares and a legacy carrier offers round-trip fares, is exacerbated right after an advance purchase deadline.[17] Second, the presence of flexible search tools facilitates the ability of consumers to search for fares across multiple departure dates. These search tools effectively allow consumers to "avoid" an advance purchase deadline by guiding them on how they need to switch their desired departure dates.

Differences in pricing policies across carriers combined with search tools make it easier for consumers to expand their choice sets across multiple departure dates. This results in increased search activity immediately prior to an advance purchase deadline and demand shifting to periods immediately prior to an advance purchase deadline. These results have significant implications on current aviation practice, as revenue management and scheduling models typically assume demand is independent across different days.

In reality, however, demand appears to be shifting to those days search tools are directing them to (or to the least full flights across multiple departure days). In this sense, the search tools can be viewed as an extension of peak load pricing problems, where the peak is determined across multiple days. This may benefit both leisure and business consumers by shifting price-sensitive leisure demand to the least time-desirable flights, saving capacity for late-arriving business travelers with stronger time preferences. However, airlines may not view this as a profitable strategy.

It is interesting to note that over the past five years, Delta has changed where it displays its "flexible search day" tools. This tool used to be predominately displayed on its home page, but can currently only be accessed through clicking on a (more opaque) advanced search tool option. In contrast, Southwest Airlines prominently displays a link to its low fare calendar when the first set of itinerary search results is returned. One possible reason for these different website designs is that the consumer mix for Delta is more heterogeneous than the consumer mix for Southwest, suggesting Delta benefits more from using advance purchase deadlines to segment their consumers (as was seen in our data by comparing business and leisure markets). From a practical perspective, many of the decision support tools used by airlines to support revenue, pricing, and scheduling decisions currently do not model consideration sets that span multiple days.

8. Conclusions and future research directions

In this study, we modeled airline travelers' online search and purchase behaviors using an analysis database from an online travel agency and QL2 Software. We model individuals' search and purchase behaviors using an instrumental variable approach that corrects for price endogeneity. Our study contributes to the literature by providing some of the first empirical insights into how individuals respond to advance purchase deadlines and price uncertainties induced by advance purchase deadlines.

Results show that the number of searches and purchases that occur in a market for specific search and departure dates are a function of search day of week, days from departure, lowest offered fares, variation in lowest fares offered across competitors, and market distance. Search activity peaks before a deadline and declines immediately after a deadline. This suggests that automated search tools help individuals learn about prices across multiple departure and/or return dates. Moreover,

[16] Additional details used to clean and process the carrier's clickstream data are provided in Hotle and Garrow (2014). In particular, we used a heuristic to identify webbots. Specifically, we used the IP address as a proxy for an individual and defined a search as pertaining to an "individual" versus "webbot" if, during the three-week period of departure dates, the IP address had at most three origin airports, at most three destination airports, and at most three frequent flyer numbers associated with the IP address.

[17] The primary motivation for carriers to use round-trip pricing is to segment business and leisure travelers as round-trip pricing enables segmentation by length of stay and/or days of travel (*e.g.*, pricing may differ for those trips that include a Saturday night stay).

individuals appear to be switching their desired departure dates by one or two days in order to avoid higher fares that occur immediately after an advance purchase deadline has passed. This is an important finding, as current revenue management systems do not take this behavior into account. Determining revenue impacts associated with failing to take this behavior into account is an important future research direction.

Looking ahead, it will be interesting to see how competitive pricing evolves, and whether LCCs will continue to use one-way pricing strategies. The primary motivation for carriers to use round-trip pricing is to segment business and leisure travelers as round-trip pricing enables segmentation by length of stay and/or days of travel (*e.g.*, pricing may differ for those trips that include a Saturday night stay). Currently, airlines face the same limitation we faced in our study – it is computationally not feasible for them to monitor all of their competitors' fares. However, by restricting the analysis to a smaller subset of lengths of stay and/or by leveraging the fact that fares with the same departure (or return) date will be highly correlated, carriers may be able to develop more efficient algorithms for monitoring competitor fares. Determining whether the ability of carriers to monitor their competitors' fares is beneficial or harmful to consumers is a second important future research direction. Finally, it is important to note that our study was based on online data from a specific OTA. It would be interesting to repeat our analysis using data from multiple distribution channels to see if our findings extend to other distribution channels. It is plausible that this limitation is making a sample selection bias. Future work may attempt to gather data for multiple distribution channels. While difficult, doing so would allow for the implementation of a selection model and provide a test of whether our results are generalizable across other distribution channels.

Acknowledgment

This research was partially funded by a National Science Foundation Graduate Research Fellowship.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tra.2015.09.001.

References

Baum, C.F., Schaffer, M.E., Stillman, S., 2003. Instrumental variables and GMM: estimation and testing. Stata J. 3 (1), 1–31, <http://www.stata-journal.com/sjpdf.html?articlenum=st0030> (accessed 08.02.14).

Berry, S., Jia, P., 2010. Tracing the woes: an empirical analysis of the airline industry. Am. Econ. J.: Microecon. 2 (3), 1–43.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.

Berry, S., Levinsohn, J., Pakes, A., 2004. Differentiated products demand systems from a combination of micro and macro data: the new car market. J. Polit. Econ. 112 (1), 68–105.

Bhadra, D., 2003. Demand for air travel in the United States: bottom-up econometric estimation and implications for forecasts by origin and destination pairs. J. Air Transp. 8 (2), 19–56.

Borenstein, S., 2010. An Index of Inter-city Business Travel for use in Domestic Airline Competition Analysis. UC Berkeley, Index available at <http://www.nber.org/data/bti.html> (accessed 08.02.14).

Branstetter, L., Chatterjee, C., Higgins, M.J., 2011. Regulation and Welfare: Evidence from Paragraph-IV Generic Entry in the Pharmaceutical Industry. NBER Working Paper 17188. <http://www.nber.org/papers/w17188>.

Brynjolfsson, E., Dick, A.A., Smith, M.D., 2010. A nearly perfect market? Differentiation vs. price in consumer choice. Quant. Mark. Econ. 8, 1–33.

Brunger, W.G., 2010. The impact of the Internet on airline fares: the "Internet price effect". J. Revenue Pricing Manage. 9 (1/2), 66–93.

Bucklin, R.E., Sismerio, C., 2009. Click here for internet insight: advances in clickstream data analysis in marketing. J. Interactive Mark. 23 (1), 35–48.

Bureau of Transportation Statistics, 2011. Air Carriers: T-100 Domestic Segment (All Carriers). <http://www.transtats.bts.gov/Fields.asp?Table_ID=311> (accessed 08.02.14).

Cramer, J., 2007. Robustness of logit analysis: unobserved heterogeneity and mis-specified disturbances. Oxford Bull. Econ. Stat. 69 (4), 545–555.

Daly, A., 2008. Elasticity, model scale and error. Paper presented at the European Transport Conference, Leeuwenhorst, The Netherlands.

Dana, J., 1998. Advance purchase discounts and price discrimination in competitive markets. J. Polit. Econ. 106 (2), 395–422.

Dana, J., 1999a. Using yield management to shift demand when the peak time is unknown. RAND J. Econ. 30 (3), 456–474.

Dana, J., 1999b. Equilibrium price dispersion under demand uncertainty: the roles of costly capacity and market structure. RAND J. Econ. 30 (4), 632–660.

Delta Air Lines, 2014. Stats & Facts: Updated June 2014. <http://news.delta.com/Stats-Facts> (accessed 16.10.14).

Fay, S., 2008. Selling an opaque product through an intermediary: the case of disguising one's product. J. Retail. 84 (1), 59–75.

Gale, I.L., Holmes, T.J., 1992. The efficiency of advance-purchase discounts in the presence of aggregate demand uncertainty. Int. J. Ind. Organ. 10 (3), 413–437.

Gale, I.L., Holmes, T.J., 1993. Advance-purchase discounts and monopoly allocation of capacity. Am. Econ. Rev. 83 (1), 135–146.

Gallego, G., Phillips, R., 2004. Revenue management of flexible products. Manuf. Serv. Oper. Manage. 6 (4), 321–337.

Goolsbee, A., Petrin, A., 2004. The consumer gains from direct broadcast satellites and the competition with cable TV. Econometrica 72 (2), 351–381.

Granados, N., Gupta, A., Kauffman, R.J., 2012. Online and offline demand and price elasticities: evidence from the air travel industry. Inform. Syst. Res. 23 (1), 164–181.

Granados, N., Gupta, A., Kauffman, R.J., 2008. Designing online selling mechanism: transparency levels and prices. Decis. Support Syst. 45 (4), 729–745.

Greene, W.H., 2003. Econometric Analysis, fifth ed. Prentice Hall, Upper Saddle River, New Jersey.

Guevara-Cue, C.A., 2010. Endogeneity and Sampling of Alternatives in Spatial Choice Models. Dissertation for Doctor of Philosophy, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Guevara, C.A., Ben-Akiva, M.E., 2012. Change of scale and forecasting with the control-function method in logit models. Transp. Sci. 46 (3), 425–437.

Guevara, C.A., Ben-Akiva, M., 2006. Endogeneity in residential location choice models. Transp. Res. Rec. 1977 (1), 60–66.

Hausman, J.A., 1996. Valuation of new goods under perfect and imperfect competition. In: Gordon, Robert J., Bresnahan, Timothy F. (Eds.), The Economics of New Goods. University of Chicago Press, Chicago, pp. 207–248.

Hausman, J., Leonard, G., Zona, J.D., 1994. Competitive analysis with differentiated products. Ann. Econ. Stat. 34, 159–180.

Hotle, S., Garrow, L.A., 2014. Competitor pricing and multiple airports: their role in customer choice. Transp. Res. Rec. 2400, 21–27.

Hsiao, C-.Y., 2008. Passenger Demand for Air Transportation in a Hub-and-spoke Network. Dissertation for Doctor of Philosophy, Civil and Environmental Engineering, University of California, Berkeley. <http://www.nextor.org/pubs/HsiaoDissertation2008.pdf> (accessed 05.18.11).
Jerath, K., Netessine, S., Veeraraghavan, S.K., 2010. Revenue management with strategic customers: last-minute selling and opaque selling. Manage. Sci. 56 (3), 430–448.
Jiang, Y., 2007. Price discrimination with opaque products. J. Revenue Pricing Manage. 6 (2), 118–134.
Johnson, E.J., Moe, W.W., Fader, P.S., Bellman, S., Lohse, G.L., 2004. On the depth and dynamics of online search behavior. Manage. Sci. 50 (3), 299–308.
Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., Lee, S.-C., 1985. The Theory and Practice of Econometrics, second ed. Wiley Series in Probability and Mathematical Sciences.
Lee, L., 1982. Specification error in multinomial logit models. J. Economet. 20 (2), 197–209.
Lee, M., Khelifa, A., Garrow, L., Bierlaire, M., Post, D., 2012. An analysis of destination choice for opaque airline products using multidimensional binary logit models. Transp. Res. Part A 46 (10), 1641–1653.
Li, M.Z.F., 2001. Pricing non-storable perishable goods by using a purchase restriction with an application to airline fare pricing. Eur. J. Oper. Res. 134 (3), 631–647.
Mantin, B., Koo, B., 2010. Weekend effect in airline pricing. J. Air Transp. Manage. 16, 48–50.
McAfee, R.P., te Velde, V., 2006. Dynamic pricing in the airline industry. In: Hendershott, T. (Ed.), Handbook of Economics and Information Systems. Elsevier, Amsterdam, pp. 527–570.
Moe, W.W., Fader, P.S., 2004. Dynamic conversion behavior at e-commerce sites. Manage. Sci. 50 (3), 326–335.
Mumbower, S., Garrow, L.A., 2010. Using online data to explore competitive airline pricing policies: a case study approach. Transp. Res. Rec. 2184 (1), 1–12.
Mumbower, S., Garrow, L.A., Higgins, M.J., 2014. Estimating flight-level price elasticities using online airline data: a first step towards integrating pricing, demand, and revenue optimization. Transp. Res. Part A 66, 196–212.
Newey, W., 1985. Generalized method of moments specification testing. J. Economet. 29 (3), 229–256.
Pagan, A.R., Hall, D., 1983. Diagnostic tests as residual analysis. Economet. Rev. 2 (2), 159–218.
Pekgün, P., Griffin, P.M., Keskinocak, P., 2013. An Empirical Study for Estimating Price Elasticities in the Travel Industry. Working Paper at University of South Carolina. <http://www2.ie.psu.edu/griffin/price.pdf> (accessed 08.02.14).
Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. J. Mark. Res. 47 (1), 3–13.
PhoCusWright, 2004. The PhoCusWright Consumer Travel Trends Survey. PhoCusWright. <http://store.phocuswright.com/> (accessed 08.02.14).
Pope, S., Garrow, L., Guin, A., Leonard, J., Bankston, L., Campbell, P., 2009. A conceptual framework for collecting online airline pricing data: challenges, opportunities, and preliminary results. Transp. Res. Rec. 2016 (1), 30–37.
Post, D., 2010. Variable opaque products in the airline industry: a tool to fill the gaps and increase revenues. J. Revenue Pricing Manage. 9 (4), 292–299.
Puller, S.L., Taylor, L.M., 2012. Price Discrimination by Day-of-week of Purchase: Evidence from the U.S. Airline Industry. Working Paper at Texas A&M University. <http://econweb.tamu.edu/puller/AcadDocs/PT_weekend_pricing.pdf> (accessed 08.02.14).
Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New York.
Smith, B.C., Darrow, R., Elieson, J., Guenther, D., Rao, B.V., Zouaoui, F., 2007. Travelocity becomes a travel retailer. Interfaces 37 (1), 68–81.
Stern, S., 1996. Market Definition and the Returns to Innovation: Substitution Patterns in Pharmaceutical Markets. Working Paper, Sloan School of Management, Massachusetts Institute of Technology.
Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, New York, pp. 80–108.
Stokey, N.L., 1979. Intertemporal price discrimination. Quart. J. Econ. 93 (3), 355–371.
Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press, New York, NY (Chapter 13).
Train, K.E., Winston, C., 2007. Vehicle choice behavior and the declining market share of U.S. automakers. Int. Econ. Rev. 48 (4), 1469–1496.
U.S. Census Bureau, 2010. American Fact Finder. <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml> (accessed 08.02.14).
Villas-Boas, J.M., Winer, R.S., 1999. Endogeneity in brand choice models. Manage. Sci. 45 (10), 1324–1338.
Weinstein, D., Keller, J., 2012. Findings from RSG's bi-annual air passenger study. Proc., Choice Based Revenue Management. Georgia Institute of Technology, p. 15.
Zhang, J., Fang, X., Sheng, O., 2007. Online consumer search depth: theories and new findings. J. Manage. Inform. Syst. 23 (3), 71–95.