Should I Stay or Should I Go? An Empirical Analysis of Consumer Behavior Using Airline Web-Traffic Data*

Alex Marsh Garrett Scott Drew Van Kuiken
Jonathan W. Williams

April 1, 2025

Abstract

We analyze consumer search and purchase behavior in response to airline revenue-management practices using data from a major carrier's website and Google Flights. We first describe patterns in search timing, purchase decisions, and paid fares. Then we estimate a multinomial logistic regression to identify factors driving search timing, finding that single adults with loyalty status, especially booking one-way nonstop itineraries, tend to search closer to departure. Next, we use a binary logistic model of conversions of searches to sales, showing that competitors' prices and changing customer composition explain rising conversion probabilities as departure nears. Finally, using a fixed-effects regression, we reveal how search and booking patterns affect prices paid. Late-arriving travelers, particularly single adults with loyalty status, pay substantially more, consistent with the airline's pricing strategies that segment more inelastic customers. Overall, our findings underscore how revenue-management, competitor fares, and consumer characteristics jointly shape online search and purchase behavior.

Keywords: Airlines, Search, Bookings, Demand

JEL Codes: L1, L93

*We thank seminar and conference participants for helpful comments on this paper, and the airline for providing the data for our analysis. We are particularly grateful to two anonymous referees and the editor for a constructive review process that substantially improved the paper. Garrett Scott and Jonathan Williams appreciate the support of NSF awards 2417693 and 2417692, respectively. Marsh: Department of Economics, University of North Carolina at Chapel Hill, alex.marsh@unc.edu; Scott: Department of Economics, University of Mississippi, gescott1@olemiss.edu; Van Kuiken: Department of Economics, University of North Carolina at Chapel Hill, drew.van.kuiken@unc.edu; Williams: Department of Economics, University of North Carolina at Chapel Hill, jonwms@unc.edu.

1 Introduction

Airlines rely on sophisticated revenue-management practices to both adjust prices to demand conditions and price discriminate. Yet, little is known about how these practices influence decisions by consumers regarding when to search and whether to buy, and the resulting heterogeneity in prices paid. A better understanding of these relationships can provide insight into the effectiveness of modern pricing practices at achieving airlines' objectives and offer guidance regarding their welfare implications to policy authorities. In this paper, we improve the current understanding of these issues through a detailed empirical analysis of the complex interaction between consumers and the revenue-management system of an airline.

Central to our analysis are novel data on consumers' visits to a major North American airline's website that include searches and booking information for all markets served by the carrier. The data for our analysis span six months from the beginning of November of 2022 until the end of April of 2023. Time-stamped records provide detailed information from each webpage viewed by all visitors to the airline's site, including itinerary-specific information like origin-destination pair, fare, indicators for round-trip and nonstop, number of passengers (infants, children, and adults), date of departure (and return if round-trip), and loyalty status if the customer logged in. In addition, we observe whether each viewed itinerary was purchased. Records for a browsing session are linked by a session identifier, and multiple sessions by logged-in customers can be linked by the customer's hashed loyalty identifier. We complement the web-traffic data with information on competitors' offerings from Google Flights for a subset of the markets served by the airline that provided the web-traffic data. The scraped data include the details of all nonstop and one-stop itineraries, such as carrier, fare, total travel time, and connecting airports.

Our empirical analysis proceeds in two steps. We first perform a detailed descriptive analysis to provide an overview of our sample and document patterns and sources of variation in searches, purchases, and prices paid. Next, we use a variety of econometric models to identify the relationship between the characteristics of itineraries and consumers and the timing of search, decision to buy, and price paid.

Our descriptive analysis reveals interesting patterns and substantial heterogeneity in the timing of searches, propensity to buy, and pricing in the market. Generally, searches and bookings increase as a flight's departure date approaches, as does the conversion rate of searches to bookings. In addition, the composition of searches changes in multiple ways. Closer to departure, there are more searches for one-way and premium itineraries by single adults with loyalty status. These customers are also more likely to purchase, which explains the overall trend in conversions of searches to bookings. Searches by families and multi-adult travel parties peak about forty days before departure and then decrease until departure.

Prices paid by consumers vary substantially within and across markets. Generally, average prices within a market increase as the departure date nears, and the price for an itinerary in the premium cabin is often double the economy cabin. Across markets, there is considerable, but intuitive, price variation. Longer-haul international flights are typically more than 50% more expensive than domestic travel. Relative to the competition in the subset of markets for which we have information on competitors' offerings from Google Flights, the airline is about 15% more expensive. Within each of these markets, the fares of the airline and its competitors have a strong positive correlation as a departure date approaches.

To go beyond our descriptive analysis and provide further insight into the relationship between consumers' behavior and the airline's revenue-management practices, we first use a multinomial logistic regression to relate the timing of searches to attributes of consumers and types of itineraries. Specifically, the regression predicts the timing of a search (i.e., among a discrete set of time windows before departure) using features reported in the web-traffic data. We find statistically significant evidence that searches for nonstop and one-way itineraries, particularly for the premium cabin, increase monotonically as the flight date approaches. Similarly, we find that searches by single adults with loyalty status increase monotonically, while families and those being re-directed from metasearch engines (e.g., Google Flights or Kayak) are less likely to search close to the departure date. Taken together with the descriptive analysis, these results show that there is a substantial and statistically significant change in the composition of consumers as flight dates approach. Further, our findings are consistent with later arrivals to the airline's website likely being more inelastic, as they consist largely of emergency travelers and business travelers with loyalty status.

Next, to provide additional evidence regarding the importance of different factors that influence the conversion of searches to bookings, we estimate a binary logistic regression model. The dependent variable in the regression is an indicator for whether the search resulted in a booking, while the explanatory variables include consumer and itinerary characteristics, a relative price measure to capture the attractiveness of competitors' offerings, and a variety of fixed effects. In our most parsimonious specification with only indicators for time windows prior to a flights departure and market, day-of-week, and month-year fixed effects, we find a monotonically increasing conversion rate as a flight's departure date approaches. However, after the itinerary and consumer characteristics of the search are added to this baseline regression, the time trend in conversions is effectively eliminated. Searches for nonstop and one-way itineraries by single adults with loyalty status are the most likely to convert to a booking, while family travel parties and those re-directed remain the least likely. In our final regression specification, estimated on the subsample of markets with accompanying Google

Flight's data, we find that stronger competition decreases the probability of a conversion while the other coefficient estimates remain largely unchanged.

The results from the regressions for searches and bookings imply that the airline's revenue-management practices induce a strong sorting of customers with respect to search timing and that those searching later are substantially more likely to book. In the last step of our empirical analysis, we use a fixed-effects regression framework to measure the heterogeneity in the price paid by customers that arises due to the airline's screening efforts. We find that the price paid for seats in the economy cabin averages 53% less than the premium cabin. Further, customers purchasing in the last and second-to-last weeks before departure pay 95% and 39%, respectively, more than those purchasing more than 30 days in advance. Conditional on the chosen cabin and timing of purchase, families pay 13% less than parties with only adults, while single adults pay 4% more. Those re-directed from metasearch platforms pay 12% less, while loyalty members pay 7% more. That there is such strong heterogeneity based on the characteristics of the travel party is interesting because it reveals the intensity of selection driven by the revenue-management system. In particular, the airline presents the same price to customers searching at the same time for a given flight (i.e., airline does not personalize price), but our analysis pools flights within a market during our sample. This implies that when prices are greater, certain consumers opt not to book or select less expensive alternatives (e.g., a different flight or date or ticket class).

Collectively, our results offer new insights into the effectiveness of revenue-management practices and directions for improvement, along with some evidence on the distributional impacts for policy authorities. We find that existing practices are quite effective at extracting higher prices from late-arriving inelastic customers by prompting earlier searches and bookings by others through higher prices close to departure. However, our results are also suggestive of opportunities for airlines. Elastic customers pay less conditioning on time and cabin because when prices are high they either don't buy or choose less-expensive options. If airlines are intent on capturing additional rents from these customers, there are some options, such as characteristics-based personalized pricing that targets discounts to lessen the frequency of the no-booking decision and targeted adjustment of prices for the most likely substitutes for a particular customer. Regarding the welfare implications of the revenue-management practices, the large difference in prices paid by customers is consistent with meaningful distributional consequences but also a substantial market-expansion effect.

Related Literature

Our paper contributes to several literatures, most importantly that on consumer behavior and pricing in the airline industry. Broadly speaking, our analysis provides a better understanding of how revenue-management practices of airlines influence and impact consumers.[1] As shown in theoretical studies such as Gale and Holmes (1993) and Gallego and van Ryzin (1994), optimal prices may not be constant in these finite horizon settings. Many papers (e.g., Dana (1998); Su (2007); Nocke et al. (2011)) have shown that advance purchase discounts (APDs) result in increasing prices that screen consumers based on valuation and arrival date. Bertsimas and de Boer (2005) and van Ryzin and Vulcano (2008) extends this line of research to the network context of the airline industry using simulation-based optimization to find optimal booking limits.

Empirical studies like Bilotkach et al. (2015), Alderighi et al. (2015), and Escobari (2012) have used various combinations of scraped and administrative data to empirically study the revenue-management practices of legacy carriers and low-cost carriers alike. More recently, there have been a number of structural modeling efforts that attempt to capture the rich dynamic interaction between consumers and price-discriminating airlines. Lazarev (2013) develops and estimates a model to measure the welfare impact of inter-temporal price discrimination, as well as counterfactual calculations that allow for resale. Williams (2022) finds that dynamic pricing policies benefit early-arriving passengers at the expense of late-arriving business travelers, and that such policies improve welfare relative to uniform pricing. Aryal et al. (2023) enrich these studies by considering both inter-temporal and intra-temporal price discrimination, where an airline sells seats in vertically-differentiated cabins up until departure. Each of these studies make strong assumptions about the arrival rate and composition of preferences for consumers that consider purchasing from the airline at each point in time before departure.[2] Our unique data on consumers' activity on the airline's website provide guidance on this crucial model feature, and inform more nuanced assumptions about search activity and the timing of purchases. In this way, our work also complements recent studies like Honka et al. (2024a) and Honka et al. (2024b) that demonstrate the importance of data that do not obscure pre-purchase consumer behavior for a firm's pricing decision. These types of data are crucial to airlines' efforts to implement further discriminatory pricing strategies like upgrade auctions (e.g., Marsh et al. (2024)) and personalized pricing (e.g., Babii et al. (2024)).[3]

Our work also complements the extensive literature that models demand for air-travel by measuring the influence of various factors on consumer choice. Chiou and Liu (2016) predict demand for air travel between Taipei and Macau as a function of fare class, ticket price, and seasonality using a multinomial logit approach. Wen and Chen (2017) take a similar approach to estimate the effect of ticket promotions on demand. Grosche, Tobias

[1]See Talluri and Ryzin (2006) and Chiang et al. (2006) for an overview of the revenue management framework.

[2]See Li et al. (2014) for estimates of the fraction of consumers that are strategic in the timing of purchase.

[3]See Dubé and Misra (2023) and Seiler et al. (2023) for studies of personalized pricing in practice.

and Heinzl (2007) provide an excellent summary of the gravity model approach that models demand as a function of endpoint characteristics. Many other studies of demand utilize a discrete-choice framework and aggregate data to understand the characteristics of an airline's service that generate market share (e.g., (Berry, 1990; Berry et al., 2006; Berry and Jia, 2010; Ciliberto and Williams, 2014; White, 2023; Bet, 2023)). These studies typically assume that the choice sets of consumers are limited to a particular destination and date of departure from their origin airport. We show that consumers often violate this assumption and this can have important implications for measuring market power and the effects of mergers. In a small set of markets, McWeeny (2023) shows accounting for a richer choice set can help reconcile the difference between the predictions of prospective structural modeling efforts and the findings of reduced-form retrospective analysis regarding harm to consumers from consolidation.

Our findings also contribute to the broader airline literature by demonstrating the value of dis-aggregated data for studying a variety of topics. For example, studies like Berry and Jia (2010) that rely on the Department of Transportation's DB1B survey data to estimate demand assume that price variation for a carrier within a market and quarter is largely due to differentiated products. We show that there is meaningful price variation between offerings, but much of this difference is due to inter-temporal price discrimination by carriers rather than ticket restrictions and other itinerary features. Additional careful studies like Luttmann (2019), Lewis (2021), and Escobari et al. (2021) are needed to disentangle discounts due to itinerary features (e.g., layovers and overnight stays) from inter-temporal adjustments to prices.[4] Related, Orlov (2011) shows that a higher level of internet penetration in a geographical location leads to more competitive markets and lower prices. This is similar to our finding that users of metasearch engines (e.g., Google Flights and Kayak) pay substantially lower fares.

The remainder of the paper proceeds as follows. In Section 2, we present the data sources and provide a descriptive analysis of the sample. Section 3 describes our empirical approach and discusses the results. In Section 4, we conclude.

2 Data and Descriptive Analysis

In this section, we introduce our data sources, discuss sample selection, and present descriptive statistics from the data. We exploit detailed information in prices and passenger

[4] These type of discounts are found to shrink when competition increases. See Borenstein and Rose (1994) and Gerardi and Shapiro (2009) for two studies that use the DB1B survey to study the effect of competition on price dispersion.

itineraries from two data sources: a North American airline and the Google Flights website. Leveraging rich itinerary information, we identify patterns in consumer search and booking behavior along several dimensions of consumer heterogeneity. We also find variation in prices across markets and evidence of dynamic pricing behavior within markets. Our descriptive findings motivate the empirical analysis in Section 3.

2.1 Data Sources

We use two unique sets of data for our analysis. The first contains detailed information from the searches conducted by all visitors to a North American airline's website over a six month span. We observe the visitors' search inputs, including origin and destination, date of departure (and return if round trip), and the number of adults and children in the travel party. Each observation in the data represents a time-stamped web page result (e.g., flight listing, itinerary summary, itinerary confirmation), which can be linked to an individual via a website session ID.[5] Other details of interest include the price per direction presented to the visitor after selecting their flight(s), the fare class associated with the price, the visitor's hashed loyalty ID and status if logged in, and the redirecting website if available. The airline provided these data from their records.

The airline's web-traffic data span from November of 2022 to April of 2023, with flight departure dates ranging from November 2022 to March 2024.[6] We begin by collapsing the data to a search, which we define as a combination of session ID, origin and destination, departure date, travel party composition, price, and fare class. This means a customer must have selected flights, reached an itinerary summary page, and received a price quote to be included in our sample. The full sample contains 12,083,366 searches from 8,099 origin-destination airport pairs. However, searches are largely concentrated in a smaller set of airport pairs within the airline's network: 50% of searches are drawn from just seventy of the 8,099 pairs and 75% are drawn from 200 pairs. 40% of the searches are for domestic trips, but domestic travel accounts for 60% of the bookings.

The second set of data consists of a panel of itineraries we scraped from the Google Flights website over an overlapping five month span. Our algorithm searches for and scrapes every one-way economy class itinerary available at the time of scraping for a given departure date and origin-destination pair. Due to our computational limitations, we collect the data from sixteen origin-destination airport pairs (eight bi-directional markets) in the airline's network

[5]The session ID is a multiple character code that is unique to a browsing session and refreshes whenever a search is resumed, or a new search is begun, at a later time. It does not contain information regarding the point of sale or the consumer's IP address.

[6]There are two weeks in January of 2023 for which we have no observations in the web-traffic data.

every seven days for the final two months leading up to departure and every fourteen days up to five months before departure. The airline provided guidance on which markets to select given our limitations and interest in market competition.[7] In these data, an itinerary is our unit of observation. The scraped details include the carrier, the cheapest available itinerary price, the time in transit (also referred to as duration), and the flight connections.[8]

The Google Flights data are available between November of 2022 and March of 2023, with flight departure dates ranging from November 2022 to May 2023. Observations are ordered by Google's internal ranking algorithm that places the three "best" itineraries based on price and convenience first in the search results followed by all remaining itineraries ordered by ascending price.[9] We begin by collapsing the data to a unique combination of bi-directional market, departure date, and days to departure. We observe the airline in 99% of the triplets. For each triplet, we subset the data to include the airline's first itinerary in the Google ranking and the first five itineraries offered by the airline's competitors in the Google ranking. For the remainder of the paper, we use the median price and the median transit time (duration) of the first five alternative itineraries to summarize the competition.

There is overlap between our two data sets, allowing us to match searches in the web-traffic data to alternative itineraries on Google Flights. This merged subsample contains 197,107 website searches and alternative itinerary options in the eight markets from the Google Flights data. The median market has 23,086 observations, the largest market by observation count has 45,936 searches and the smallest has 4,059 searches. In the subsample, 50% of the searches are for domestic trips, and domestic travel accounts for 65% of the bookings.

2.2 Web-Traffic Descriptives

Figure 1 explores patterns in searches and bookings over the time span of our data. In panel (a), we find that the busiest day on the website is near the end of November, approximately four weeks before the day with the least website activity, December 24$^{\text{th}}$. Despite the relatively short time horizon available to us, panel (b) shows a notable peak in the searched and booked flights near the end of December 2022. However, the most searched and booked flights in our sample are in April 2023. This is unsurprising given this set of flights contains

[7]Per our data use agreement, we cannot reveal the markets. All eight of the markets feature one of the airline's hubs, and six feature at least one major competitor's hub. Four of the markets are domestic, and prices in these markets are slightly less-expensive on average than, but similar in distribution to, the full sample of domestic markets. The remaining four markets are international, leisure markets. These markets are less expensive on average than the full sample of international markets.

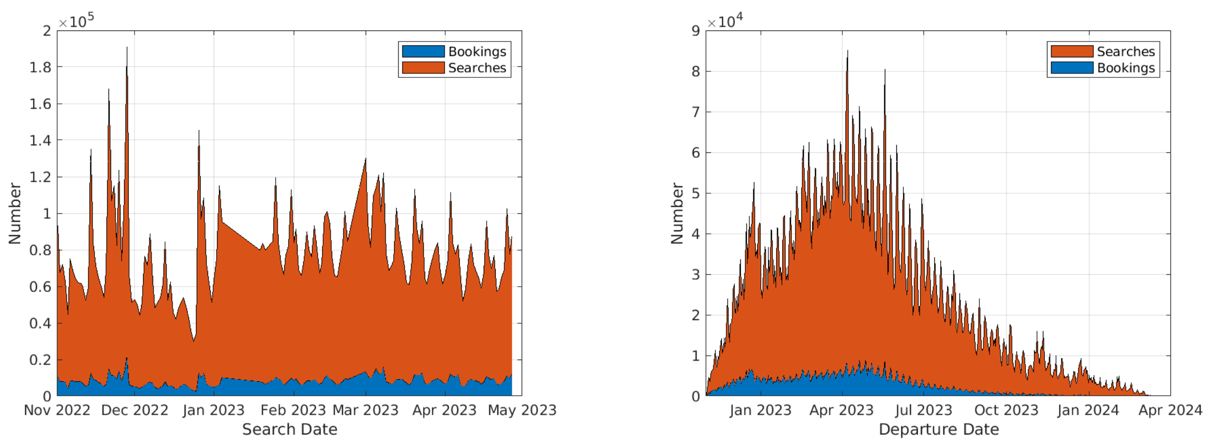
[8]Google partners directly with airlines to provide access to itinerary and pricing information.

[9]"Best flights are ranked based on the best trade-off between price and convenience factors such as duration, number of stops, and airport changes during layovers." (Google Flights)

the longest time horizon of searches leading up to departure, and also includes the final weeks before departure. We show shortly that the final weeks before departure account for the majority of searches and bookings.[10]

Panel (b) of Figure 1 reveals one drawback of our sample. Air travel is very seasonal, and although we have some insight into how consumers make decisions from the late fall through the early spring, it would be interesting to see whether holiday travel is booked particularly far in advance relative to other trips, which may have a higher concentration of business travel. To the extent that we do have insight into this seasonality, we find that consumers search for flights well in advance of summer and winter holiday seasons in 2023, though the conversion rate for these searches is quite low.

Figure 1: Searches and Bookings by Search and Departure Date



(a) Search Date

(b) Departure Date

Notes: The figure tracks the total number of searches and bookings over time in the airline's web-traffic data. Panel (a) characterizes the web access dates of the searches and bookings. The time horizon covers that of the website data. Panel (b) refers to the departure dates that consumers considered in the website data. There are two weeks in January of 2023 for which we have no observations in the website data.

We next explore the detailed consumer information in the searched itineraries. Table 1 presents descriptive statistics from the consumers' search inputs on the airline's website. The mean statistic indicates the share of searches with the selected input. We also record the proportion of searches that are converted into a purchase ($\mathbb{I} = 1$) and the conversion rate of the excluded set of searches ($\mathbb{I} = 0$). To illustrate, 52% of searches are made by a single adult, and 13% of these searches are converted. However, the excluded set of searches (i.e. multiple adults and travel parties with children) have a conversion rate of only 8%.

[10] See Appendix A.1 for additional details about the variation in consumer behavior by the time of year and day of week.

Interestingly, the conversion rate of one-way itineraries is more than twice the conversion rate of round-trip itineraries, and loyalty members convert four times as often as consumers who do not log in. In general, the behavior of one-way travelers, single adults, and loyalty members plays a pivotal role in our sample, as we will discuss throughout this section and Section 3.

Table 1: Itinerary Characteristics and Conversion Rates

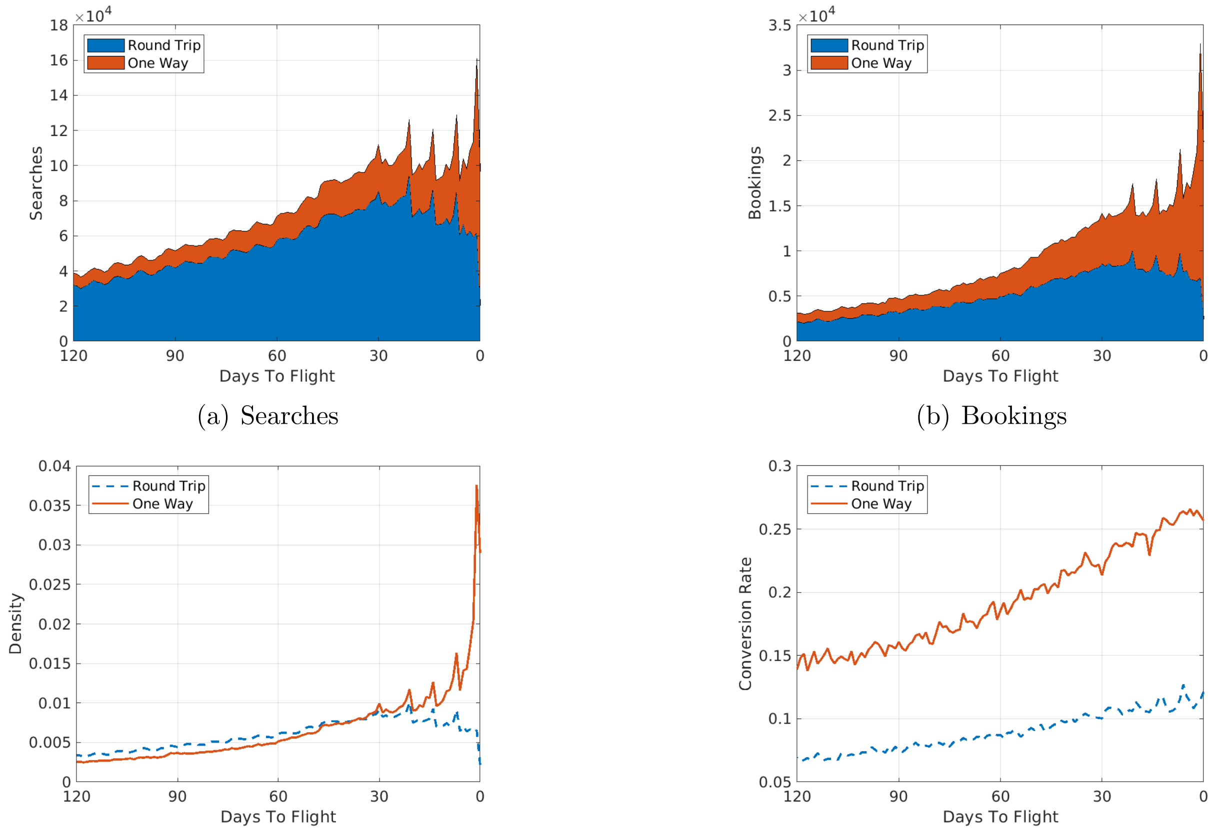
		Conversion Rate	
	Mean	$\mathbb{I} = 1$	$\mathbb{I} = 0$
Itinerary Type			
Domestic	0.396	0.161	0.072
Economy	0.948	0.106	0.124
Nonstop	0.643	0.122	0.081
Round Trip	0.782	0.082	0.197
Travel Party			
Single Adult	0.519	0.132	0.081
Multiple Adults	0.371	0.085	0.121
Family	0.110	0.067	0.112
Loyalty Status			
Not Logged In	0.599	0.048	0.196
Tier 1	0.336	0.188	0.067
Tier 2	0.027	0.203	0.105
Tier 3	0.019	0.238	0.105
Tier 4	0.019	0.288	0.104

Notes: The table contains descriptive statistics for itinerary searches in the web-traffic data. The variables include search inputs and the loyalty status if the consumer logged in. The statistics include the share of each itinerary characteristic and the conversion rates of the included ($\mathbb{I} = 1$) and excluded ($\mathbb{I} = 0$) searches.

We now characterize the patterns that arise in searches and bookings leading up to departure, conditioning on some of the available customer information. In Figure 2, we compare the search and booking patterns of round-trip and one-way travelers. Consumers search for round-trip itineraries at a relatively higher frequency throughout the time horizon. However, the densities indicate that the majority of one-way searches occur close to departure, after the rate of round-trip searches begins to decline. The high rate of conversion of one-way searches coupled with the mass of late-arriving consumers indicates a sense of urgency amongst many one-way travelers.

In Figure 3, we consider the same collection of figures for different types of travel parties: single adult, multiple adults, and families (travel parties with at least one adult and one child). Consumers traveling in groups make up the majority of searches and bookings further in advance of the flight. Although the conversion rate remains fairly constant for the travel groups until the final few days before departure, the conversion rate of single adults climbs steadily over time. This increase coincides with an increase in the search density for single

Figure 2: Temporal Patterns by Trip Type



(a) Searches (b) Bookings

(c) Search Density (d) Conversion Rate

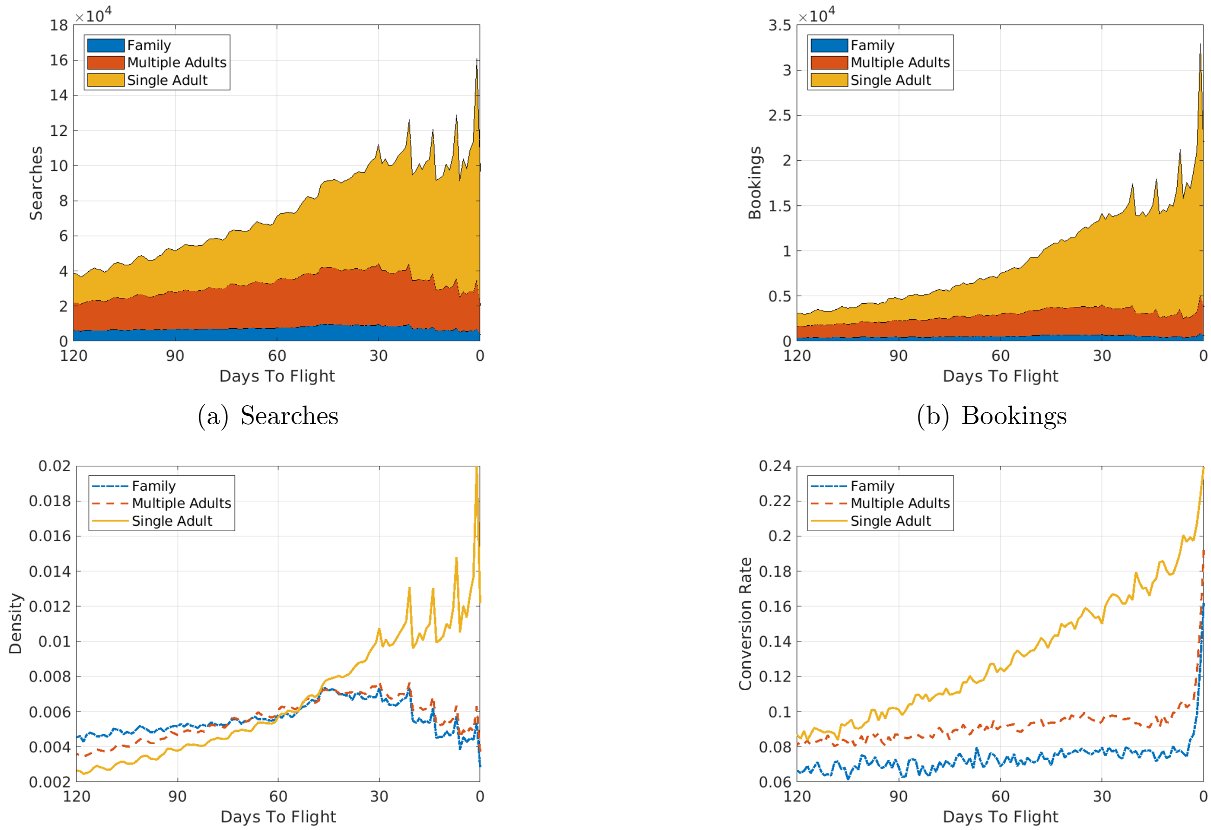
Notes: The figure displays the number of searches (a) and the number of bookings (b), as well as the density of searches (c) and the conversion rate of searches (d) at a given number of days before departure for round-trip and one-way travelers. These figures are created using the web-traffic data.

adults. These patterns appear similar to those of one-way travelers, which is unsurprising given that 29.1% of bookings in the data are made by single adults on one-way itineraries.

We find that customer loyalty status is another important determinant in the search and booking process. 60% of the searches are made by consumers who do not log in, yet they account for less than 30% of the bookings. According to Figure 4, the search rate of these consumers varies little over the final few months before departure, a pattern shared by the lower-tier loyalty members. However, these two segments of the population differ quite drastically in their conversion rates. In fact, the conversion rate of the lower-tier loyalty members mirrors that of the upper-tier loyalty members and tends to be much higher than the conversion rate of the consumers who do not log in.[11]

[11]Premium cabin travelers book close to departure, as well, but make up just slightly more than 5% of the searches. In Figure A.3 of Appendix A.2, we consider the temporal patterns in searches and bookings

Figure 3: Temporal Patterns by Travel Party



(a) Searches

(b) Bookings

(c) Search Density

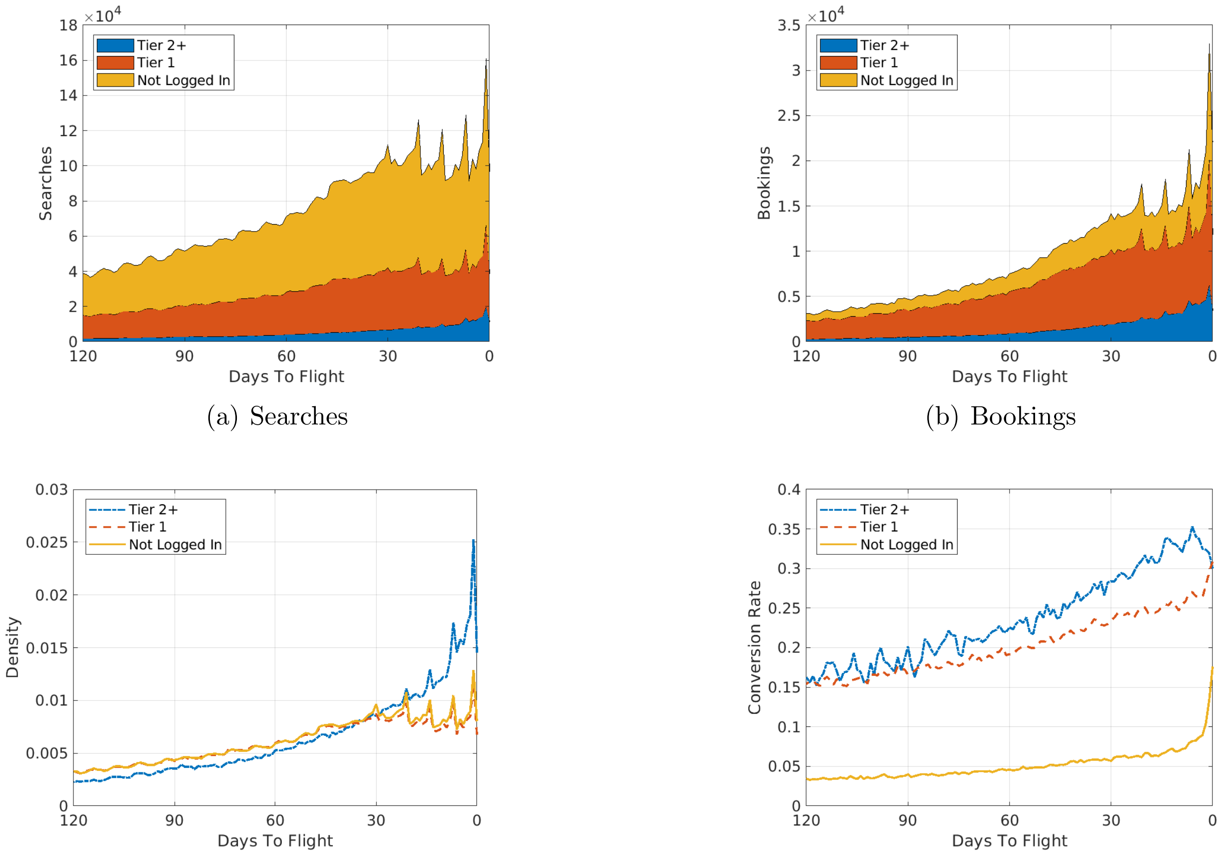
(d) Conversion Rate

Notes: The figure displays the number of searches (a) and the number of bookings (b), as well as the density of searches (c) and the conversion rate of searches (d) at a given number of days before departure for travel parties with a single adult, multiple adults, and parties with children (families). These figures are created using the web-traffic data.

The webpage format of the data provides a potential explanation for these findings. Exploiting the website session and loyalty identification numbers, we track consumers within and across sessions. Within a session, we observe that approximately 10% of consumers search across multiple airport pairs. A slightly larger percentage search across multiple departure dates. Additionally, we observe loyalty members who log in during their website session and find that these individuals purchase at a higher rate than the individuals that never log in. Across sessions, we observe individual loyalty members search the same flights multiple times before purchasing. More than 10% of the loyalty members that do not book are identified later in the sample searching in the same market. These findings suggest that

for premium and economy travelers.

Figure 4: Temporal Patterns by Loyalty Status



(a) Searches

(b) Bookings

(c) Search Density

(d) Conversion Rate

Notes: The figure displays the number of searches (a) and the number of bookings (b), as well as the density of searches (c) and the conversion rate of searches (d) at a given number of days before departure for travelers with upper-tier loyalty status, base-tier loyalty status, and those that did not log in. These figures are created using the web-traffic data.

a number of the searches made by "non-loyalty" members are simply individuals who have yet to commit to purchasing and wish to evaluate options and prices without logging in.

In an effort to understand how prices may impact consumer behavior (and vice versa), we survey the price information in our two data sets. In Table 2, we describe the distribution of per-direction prices from the airline's website data. The data corroborate a number of stylized facts from the airline industry, including that premium cabin itineraries are more expensive than economy cabin itineraries in the same market.[12] We also find that international travelers typically pay higher prices than domestic travelers in this airline's network. Additionally, the data allow for new insights into the prices quoted to consumers who do

[12]See Aryal et al. (2023) and Marsh et al. (2024) for studies on price discrimination using multiple cabins.

not purchase. In particular, we observe that the distribution of these searched prices has a greater mean and a longer upper tail than the distribution of transaction prices in every itinerary segment of the data.

Table 2: Booked and Searched Prices

	Domestic							
	Bookings				***Searches (not Booked)***			
	Mean	25^{th}	50^{th}	75^{th}	Mean	25^{th}	50^{th}	75^{th}
Economy								
Fare (*per direction*)	212.46	122.44	183.59	269.42	231.88	139.49	205.11	288.32
Observations	732,859				3,870,043			
Premium:								
Fare (*per direction*)	521.73	377.68	481.38	623.40	617.05	416.42	534.68	760.04
Observations	37,810				139,821			
	International							
	Bookings				***Searches (not Booked)***			
	Mean	25^{th}	50^{th}	75^{th}	Mean	25^{th}	50^{th}	75^{th}
Economy								
Fare (*per direction*)	321.17	177.26	259.97	404.00	394.01	220.28	328.45	491.05
Observations	485,790				6,360,888			
Premium								
Fare (*per direction*)	664.10	433.88	584.39	819.40	845.11	532.61	748.82	1,078.42
Observations	40,823				415,332			

Notes: The table describes the distribution of itinerary prices in the airline's web-traffic data. Observations are segmented by destination, cabin, and whether the itinerary was purchased or not.

2.3 Google Flights Descriptives

In Table 3, we use the Google Flights data to summarize the airline's prices and durations on Google Flights relative to its competition. For each observation in the data, we calculate the airline's price and duration as a percentage of the competition's alternative. For example, the airline's posted prices are 15% more expensive on average, but their itineraries are 18% faster on average. Although prices are centered near zero, the distribution skews heavily right. In more than 10% of the observations, the airline's price is double its competition's. However, the airline's itineraries are shorter on average than its competition's, with the median duration just under 30% faster.

In the second part of Table 3, we compare the prices faced by users searching for economy class itineraries on the airline's website to the competition's price found on Google Flights on the day of of booking using the merged subsample. We find that, although the distribution is still centered near zero, the price quoted to website visitors is 33% more expensive

Table 3: Airline Price and Duration Relative to Competitor

	Mean	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Google Data						
Price	0.15	-0.52	-0.27	-0.01	0.34	1.03
Duration	-0.18	-0.72	-0.53	-0.29	-0.02	0.28
Website Data						
Price	0.33	-0.50	-0.26	0.07	0.63	1.40

Notes: The table describes the airline's prices and durations relative to its competition. We use the airline's best price for each observation in the Google Flights data and the prices observed by consumers in the airline's website data. The variable is constructed by taking the airline's price in the respective sample minus the competition's price in the Google Flights data, then normalizing by the competition's price. The top panel uses the Google Flights data, and the bottom panel uses the merged subsample of the web-traffic data.

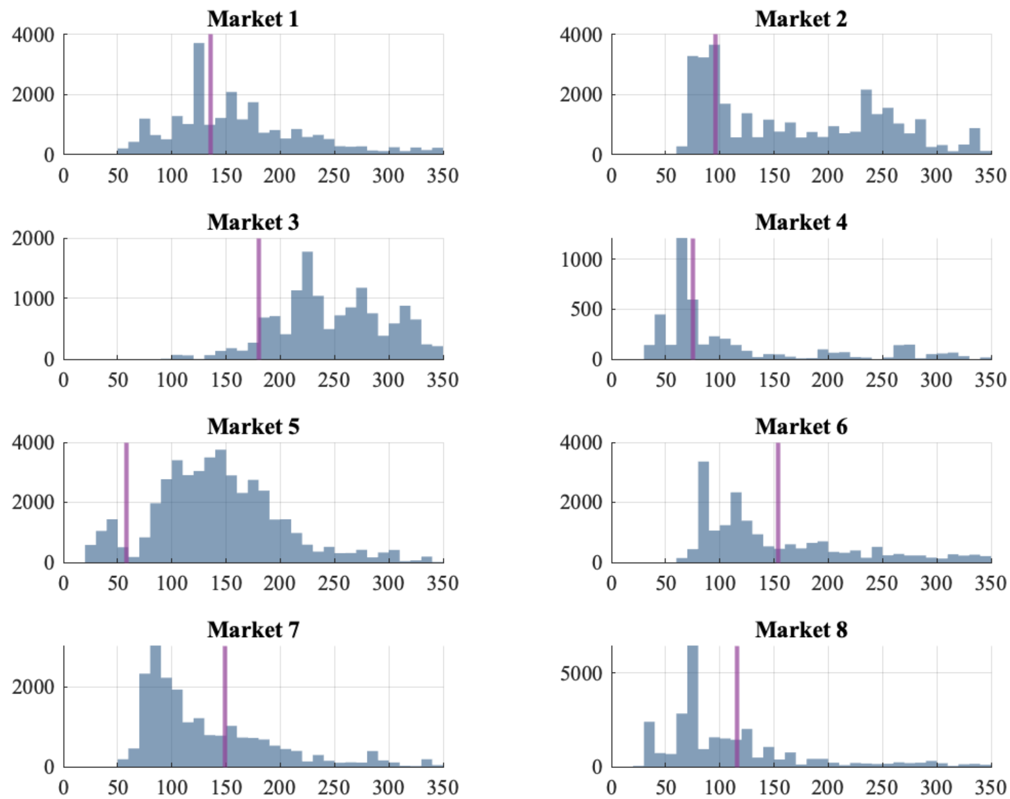
on average than the competitors' price from Google Flights. The shift in the distribution potentially reflects the menu of options within the economy class on the airline's website. The fares presented to consumers on the website may be more expensive than the cheapest available option due to self-selection into additional features such as extra leg room or priority boarding, or consumers selecting more expensive flights that leave at more desirable times during the day.

In Figure 5, we describe the competitors' prices across the eight markets in the Google Flights data. The histograms show the variation in the range and distribution of prices across the sample in each market. The vertical lines show the airline's median price within the market using the observations in the Google Flights data. The competition varies prices within each market in much the same way as the airline, but in some markets, the airline is much more competitive on prices. This figure provides only a glimpse of this competition and does little to describe the time-varying aspect of pricing and competitiveness, which is more relevant when considering the choices a consumer faces when purchasing a ticket.

Figure 6 traces the joint paths of the prices posted by the airline and its competition across time in each of the eight markets. Each marker represents the pair of average prices calculated every seven days leading up to departure. We find that both groups increase prices on average in all markets as departure nears. Posted prices on Google Flights rarely decrease in our sample. Of the 519 market-departure date combinations we observe, only fifty feature an instance of either the airline's price or a competitor's price decreasing between periods.[13] In most markets, prices hover around the 45-degree line throughout the time

[13] Alderighi et al. (2015) discuss this phenomenon in detail. The authors find that fares increase monotonically with capacity. Holding capacity fixed, prices follow a U-shaped profile, where prices are lowest between 35 and 14 days before departure. Because we do not have information on the number of seats available in a given fare class, we cannot separate the capacity and temporal effects on prices.

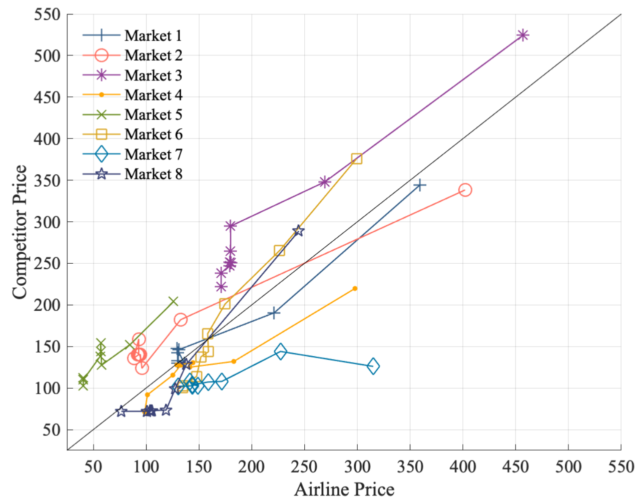
Figure 5: Competition Price by Market



Notes: The figure shows the distribution of the competition's prices for each market in the Google Flights data. The vertical line represents the airline's median price across all observations in the Google Flights data.

horizon, indicating the two groups typically offer similar prices. We find that the airline offers cheaper prices than its competitors on average in markets 3 and 5. In other markets, such as 4, 7 and 8, the airline and its competition post similar prices early in the time horizon, but the airline becomes relatively more expensive as the departure date nears. The variation across markets in the degree of price competitiveness does not seem to be driven by the duration of the flights. In all eight markets, the airline offers a shorter median time in transit than its competitors.

Figure 6: Evolution of Posted Prices



Notes: The figure shows the paths of the prices posted by the airline and its competition over the final fifty-six days leading up to departure for each market in the Google Flights data. The markers indicate the pair of average prices across observations in a given week before a flight for the airline (x-axis) and the competition (y-axis).

3 Empirical Analysis of Pricing, Search, and Bookings

In this section, we build upon the descriptive analysis of Section 2 to provide further insight into the dynamics of pricing practices, consumer search, and bookings. We begin by estimating a multinomial logistic regression to concisely characterize the changing composition of consumers searching for flights as the departure date nears. We then use a binary logit regression to measure the effect of different factors on the decision to buy for consumers searching on the airline's website. Finally, we identify the importance of different determinants of transacted prices via a fixed-effects regression. In each of these regressions, we control for the same set of indicator variables, which include the full set of itinerary and consumer characteristics available in the web-traffic data.[14]

3.1 Timing of Searches

To concisely describe the changing patterns in the attributes of customers searching on the airline's website as the departure day for a flight approaches, we estimate a multinomial logistic regression. Specifically, we express the probability of a given search occurring in one of $K = 4$ time periods as

[14]Table 5 includes one additional regressor, Price Relative to Comp., which is unique to that table.

$$P(Y=k|Z_i) = \frac{\exp(Z_i\gamma_k)}{1+\sum_{j=1}^{K-1}\exp(Z_i\gamma_j)}, \tag{1}$$

where $Y \in \{1,..,K\}$ is a discrete outcome that indicates the search period: 0-7, 8-14, 15-30, and 30 or more days before departure. The vector Z_i includes a constant and characteristics of the consumer and itinerary, and γ_k is the k^{th} period-specific vector of coefficients to be estimated. This regression is effectively a classification exercise, where consumer and itinerary characteristics are used to predict the timing of a search. As required, we normalize the coefficients for the K^{th} time period to zero (i.e., $\gamma_{30+} = 0$), which makes 30 or more days the base case against which others are measured.

Table 4: Timing of Search Multinomial Logistic Regression

	15-30 NDO	8-14 NDO	0-7 NDO
Variables			
Intercept	-1.05***	-1.61***	-0.81***
	(0.01)	(0.01)	(0.01)
Nonstop	0.12***	0.14***	0.18***
	(0.00)	(0.00)	(0.00)
Round Trip	-0.20***	-0.42***	-1.13***
	(0.00)	(0.00)	(0.00)
Economy	-0.10***	-0.43***	-0.77***
	(0.00)	(0.01)	(0.00)
Loyalty Member	0.03***	0.09***	0.14***
	(0.00)	(0.00)	(0.00)
Family	-0.06***	-0.11***	-0.07***
	(0.00)	(0.01)	(0.01)
Single Adult	0.42***	0.64***	0.86***
	(0.00)	(0.00)	(0.00)
Metasearch	-0.21***	-0.39***	-0.54***
	(0.01)	(0.01)	(0.01)
Fit Statistics			
N: 7,672,158			
BIC: 16,715,529			

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The table shows the results of a multinomial regression for consumer and ticket characteristics on temporal arrival rates using the web-traffic data. For each variable, the omitted group is website visits between 31 and 90 days before departure. Website visits earlier than 90 days before departure are excluded. NDO is shorthand for Number of Days Out from departure.

We organize the coefficient estimates of the multinomial logistic regression corresponding to equation 1 in Table 4 to ease interpretation. The omitted group for the travel party indicators (Family and Single Adult) is the set of itineraries with multiple adults in the party.

Each row reports the coefficients for a particular covariate across the different time periods. For example, the first row shows that most searches occur more than 30 days before departure (i.e., coefficients for remaining periods are all negative), while the probability a search occurs in the remaining periods increases as the departure date nears (i.e., the coefficient for the final week is less negative than the others). Similarly for the remaining covariates, the magnitude of the coefficients from left to right across columns within a row indicates whether a particular type of search was more or less likely to have occurred nearer to departure. We find that searches for nonstop flights by single adults with a loyalty status are increasingly likely to have occurred nearer to departure (i.e., coefficients are monotonically increasing off a baseline of zero). Conversely, searches for round-trip travel, economy itineraries, and redirections from metasearch engines are less likely closer to a flight's departure (i.e., coefficients are monotonically decreasing off a baseline of zero).[15]

The estimates in Table 4 contribute to our discussion at the end of Section 2. In particular, those whom we assume are less price-sensitive (e.g., one-way travelers, single adults, and loyalty members) are more likely to search on the airline's website as the departure date nears and prices rise. On the other hand, perhaps anticipating increasing prices, more price-sensitive consumers (e.g., families and those searching on metasearch websites) are more likely to search far in advance of departure when prices are relatively cheap.

3.2 Decision to Purchase

Next, we consider the decision to purchase. To describe the relationship between consumer and itinerary characteristics and whether a search is converted to a purchase, we estimate a binary logistic regression model. This corresponds to specifying a probability of purchase equal to

$$P(Y_{imdt} = 1|W_i) = \frac{\exp(W_i\theta + \lambda_t + \nu_d + \omega_m)}{1 + \exp(W_i\theta + \lambda_t + \nu_d + \omega_m)}, \tag{2}$$

where Y_{imdt} is an indicator for whether a search by consumer i in market m for a flight departing on day of week d in period t results in a purchase. We let λ_t, ν_d, and ω_m denote month-year, day-of-week, and market fixed effects, respectively. The vector W_i includes indicators capturing characteristics of the itinerary and customer, and θ is the vector of coefficients to be estimated.

We report the marginal effects for different specifications of equation 2 in Table 5. In

[15] Appendix A.3 includes the same multinomial logit model, but run on only observations which are matched to the Google Flights data. We find that coefficients are of similar direction and magnitude, though standard errors are larger due to the smaller sample size.

Table 5: Ticket Purchase Probability Binary Logistic Regressions

	Full Dataset		***Google***
	(1)	(2)	(3)
Variables			
0-7 NDO	0.05***	0.01**	-0.02*
	(0.00)	(0.00)	(0.01)
14-30 NDO	0.01***	0.01***	-0.00
	(0.00)	(0.00)	(0.00)
8-14 NDO	0.03***	0.01***	-0.01*
	(0.00)	(0.00)	(0.00)
Nonstop		0.04***	0.05***
		(0.00)	(0.02)
Round Trip		-0.10***	-0.14***
		(0.00)	(0.01)
Economy		0.00	-0.03***
		(0.00)	(0.01)
Loyalty Member		0.17***	0.22***
		(0.00)	(0.03)
Family		-0.02***	-0.02**
		(0.00)	(0.01)
Single Adult		0.04***	0.06***
		(0.00)	(0.01)
Metasearch		-0.02***	-0.02
		(0.00)	(0.01)
Price Relative to Comp.			-0.01***
			(0.00)
Fixed-effects			
Market	Yes	Yes	Yes
Day of Week	Yes	Yes	Yes
Month-Year	Yes	Yes	Yes
Fit Statistics			
N	7,654,470	7,654,470	190,356
R^2	0.04	0.14	0.19
Within R^2	0.05	0.16	0.19

Clustered (Market) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The table shows the average marginal effects for a regression of various consumer- and flight-level characteristics on website visit purchase probabilities. Columns (1) and (2) use the web-traffic data, and the column (3) uses the merged subsample. Day of week and month-year fixed effects pertain to the departure date of a flight.

column 1, we report a parsimonious specification to capture the general time trends in the probability of a search resulting in a purchase, while including fixed-effects to control for market, day-of-week, and month-year of the flight. Relative to the excluded indicator for a search more than 30 days before departure, we find that the probability of a search resulting in a purchase is monotonically increasing as the departure date nears. This is consistent with the results in Section 3.1 that show the composition of customers is increasingly made up of those individuals that we expect to be more inelastic and likely to purchase despite prices that rise on average.

In column 2 of Table 5, we report results from a specification that builds on the baseline reported in column 1 by also including characteristics of the itinerary and consumer. Interestingly, we find that the inclusion of these covariates effectively explains away the time trend in the probability of purchase. Searches for nonstop and one-way itineraries are more likely to convert to a purchase, as are searches by single adults with loyalty status. Searches for families and those redirected from metasearch engines are less likely to convert to a purchase, which again we largely attribute to their price sensitivity.

In column 3 of Table 5, we rerun the regression from column 2 but restrict our analysis to the Google Flights sample to include the relative price measure discussed in Section 2. Most of the statistically significant results from column 2 remain unchanged with similar magnitudes. The relative price metric has the expected sign and is statically significant, i.e., prices greater than the competition result in fewer searches being converted to purchases. The economic significance is also meaningful. The mean of the metric is 0.33, and the standard deviation is 0.97. Our results indicate that increasing relative prices by one standard deviation decreases conversion rates by one percent. Note that the comparison between website prices and Google Flights prices is subject to the same drawbacks that we discussed in Section 2: Google Flights prices are for the cheapest possible fare class, while website prices incorporate add-ons like additional leg room and upgrades like sitting in first class. We also do not have insight into the route that consumers on the website are considering.

3.3 Transacted Prices

Finally, we examine factors influencing the heterogeneity in prices paid by analyzing the relationship between itinerary- and consumer-specific attributes and transaction prices. Publicly available data like the Department of Transportation's DB1B Survey are aggregated to a level that prevents meaningful insight into many sources of variation in transacted prices prior to a flight's departure. In contrast, our data include information on the timing of the purchase and characteristics of the traveling party, which offers an opportunity to identify heterogeneity in prices paid by passengers across time within a market.

To do so, we begin by estimating a fixed-effects regression of the form

$$p_{imdt} = X_i\beta + \lambda_t + \nu_d + \omega_m + \epsilon_{imdt}, \tag{3}$$

where p_{imdt} is the price paid by passenger i in market m for a flight departing on day of week d in period t, λ_t, ν_d, and ω_m are month-year, day-of-week, and market fixed effects, respectively, and ϵ_{imt} is an idiosyncratic error term. The vector X_i includes a sequence of indicators for the timing of purchase relative to flight departure, indicators for whether the itinerary is nonstop, roundtrip, and economy, and indicators that describe travel party details such as whether they are a loyalty member, family, single adult, and purchasing following a metasearch redirection.

The estimates of equation 3 are presented in Table 6 for two alternative specifications using two different samples. In column 1 and 2 of Table 6, we present coefficient estimates from the full sample when price is measured in levels and logged, respectively. Columns 3 and 4 report analogous results to columns 1 and 2, but the sample is restricted to those markets for which we have accompanying data from Google Flights. The results are qualitatively quite similar across the four columns of Table 6 with some loss of statistical power when the analysis is restricted to the Google Flights sample. Given this similarity and for ease of interpretation of the coefficient estimates, we focus our discussion on column 2.

We find that relative to those purchasing more than 30 days prior to departure, customers purchasing in the last week before departure pay approximately 95% (i.e., $e^{0.67} - 1$) more than those purchasing 30 days in advance. The relationship is monotonic; those purchasing 8-14 days and 15-30 days before departure pay 39% and 11% more, respectively. Like other studies (e.g., Ito and Lee (2007)), we find that purchases of nonstop itineraries are discounted by 23% relative to connecting itineraries purchased in the same market for the carrier, though the results for round-trip itineraries are inconsistent and statistically insignificant in some specifications. The price paid for tickets for the economy cabin average 53% less than those for the premium cabin, representing a substantial willingness to pay for quality. Regarding attributes of the travel party, we find that single adults pay 4% more than multiple adults, while families (at least one adult with at least one child) pay 13% less. Customers with the highest frequent flier loyalty status tiers pay 7% more than others, while those originating from a metasearch platform pay 12% less.[16]

Taken together, our results demonstrate the important changes in passenger composition that arise due to selection. Specifically, identification in the fixed-effects regression relies

[16] In Appendix A.3, we present results for the specification in Column 1 where each covariate is interacted with time.

Table 6: Transaction Pice Regressions

	Full Dataset		***Google***	
	Price	***Log(Price)***	***Price***	***Log(Price)***
	(1)	(2)	(3)	(4)
Variables				
0-7 NDO	181.66***	0.67***	133.43***	0.71***
	(7.79)	(0.03)	(24.31)	(0.05)
8-14 NDO	78.63***	0.33***	55.80***	0.36***
	(3.34)	(0.02)	(9.03)	(0.07)
15-30 NDO	25.98***	0.10***	22.34***	0.15***
	(1.40)	(0.01)	(4.17)	(0.03)
Nonstop	-68.18***	-0.26***	-35.74***	-0.19***
	(4.66)	(0.02)	(8.74)	(0.04)
Round Trip	-11.43***	0.00	-1.09	0.03***
	(1.33)	(0.01)	(1.68)	(0.01)
Economy	-293.15***	-0.76***	-226.53***	-0.85***
	(7.09)	(0.02)	(25.73)	(0.04)
Loyalty Member	11.27***	0.07***	7.37*	0.08**
	(1.02)	(0.01)	(3.41)	(0.02)
Family	-26.79***	-0.14***	-1.26	-0.08**
	(1.72)	(0.01)	(6.64)	(0.03)
Single Adult	9.74***	0.04***	11.20**	0.06***
	(1.02)	(0.00)	(3.86)	(0.02)
Metasearch	-40.39***	-0.13***	-13.54*	-0.04
	(2.94)	(0.01)	(6.49)	(0.04)
Fixed Effects				
Market	Yes	Yes	Yes	Yes
Day of Week	Yes	Yes	Yes	Yes
Month-Year	Yes	Yes	Yes	Yes
Fit Statistics				
N	990,071	990,071	30,020	30,020
R^2	0.53473	0.59260	0.45146	0.44199
Within R^2	0.34204	0.35694	0.36570	0.32905

Clustered (Market) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The table shows the results of four regressions on transacted average prices (columns (1) and (3)) and log average prices (columns (2) and (4)). All regressions are subset to purchases which happen within 90 days of departure. Columns (1) and (2) show results for the full sample of web-traffic data; columns (3) and (4) use only the merged subsample. NDO is shorthand for Number of Days Out from departure. Day of week and month-year fixed effects pertain to the departure date of a flight.

upon there being multiple options within each market because there is no variation in prices faced by different customers searching at the same time for a given class on a flight (i.e., airline does not personalize price). Thus, observing a particular type of passenger paying less on average for travel occurs because passengers choose whether to purchase and which ticket to purchase based on the price. For example, families redirected from metasearch platforms pay considerably less than single adults with the highest frequent flier loyalty status. This difference may reflect both the decision not to purchase when prices are greater and the choice of less-attractive options due to departure times, ticket class restrictions, or other factors. Separating these different explanations requires something like experimental data that randomizes prices for a flight so that the elasticity to the no-purchase option and alternative flights for different groups can be separately identified.[17]

4 Conclusion

Many researchers and practitioners have dedicated substantial effort to understanding consumer behavior in the airline industry. In this paper, we use novel web-traffic data from a major airline to analyze the search and booking behavior of air travelers. We also pair the web-traffic data with information from the Google Flights website to study the effects of prices and competition on searches and bookings.

We show that patterns in consumer behavior vary with itinerary and consumer characteristics. One-way travelers, single adults, and loyalty members arrive closer to departure and have a higher propensity to book than other consumers within the market. We find that late-arriving consumers often pay more than other consumers, which is unsurprising given that prices increase on average across markets leading up to departure. However, we find that one-way travelers, single adults, and loyalty members pay more on average after conditioning on the booking date, indicating that these consumers are more willing to choose expensive options when presented with choices. On the other hand, we find that consumers redirected to the airline's website from metasearch websites (e.g., Google Flights) have a slightly lower propensity to book and typically transact at lower prices than other consumers, highlighting their price sensitivity.

Our findings have important implications for algorithmic pricing, targeted pricing, and price matches. In particular, sophisticated pricing strategies that are able to identify consumer "types" that are only likely to purchase when prices are relatively low, or are more willing to purchase when prices are relatively high, create opportunities for firms to extract

[17] See Babii et al. (2024) for a study that uses experimental price variation to identify purchase elasticities that vary by customer characteristics.

additional surplus from consumers. The evidence we have shown indicates that details provided by consumers, either necessarily or willingly, contain a great deal of information about price sensitivity.

Additionally, our findings challenge common assumptions made in the academic literature and in public policy settings regarding market definition in the airline industry, but they also invite thought and discussion about the limitations and costs of such assumptions. In particular, choice sets are unlikely to be uniform across consumers and may not be limited to flights, origin-destination pairs, or dates. For example, we have provided some evidence that a portion of consumers explore more than one destination or more than one departure date within website sessions. Across website sessions, we have used loyalty identification numbers to demonstrate that some consumers track the same or similar flights over time without purchasing.

Although the web-traffic data are sourced from a single airline's website, they identify important segments of the consumer population that drive the booking and pricing decisions that we observe in less granular data. Admittedly, our data are not best suited to tackle issues of market definition, but they indicate that future research should approach choice set formation more carefully. In general, we hope this work encourages the continued use of more granular novel data sources to inform air-travel demand modeling efforts.

References

Alderighi, Marco, Marcella Nicolini and Claudio A. Piga (2015). "Combined Effects of Capacity and Time on Fares: Insights from the Yield Management of a Low-Cost Airline." *The Review of Economics and Statistics*, 97(4): 900–915.

Aryal, Gaurab, Charles Murry and Jonathan W Williams (2023). "Price Discrimination in International Airline Markets." *The Review of Economic Studies*: rdad037, URL `https://doi.org/10.1093/restud/rdad037`.

Babii, Andrii, Rohit Kumar, Alex Marsh, Garrett Scott and Jonathan W. Williams (2024). "Targeting Discounts." *Working Paper*.

Berry, Steven, Michael Carnall and Pablo T. Spiller (2006). "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity." *Competition Policy and Antitrust*: 183–213.

Berry, Steven T. (1990). "Airport Presence as Product Differentiation." *American Economic Review*, 80(2): 394–399.

Berry, Steven T. and P. Jia (2010). "Tracing the Woes: An Empirical Analysis of the Airline Industry." *American Economic Journal: Microeconomics*, 2(3): 1–43.

Bertsimas, Dimitris and Sanne de Boer (2005). "Simulation-Based Booking Limits for Airline Revenue Management." *Operations Research*, 53(1): 90–106.

Bet, German (2023). "Market Power in the U.S. Airline Industry." *Working Paper*, URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3913695`.

Bilotkach, Volodymyr, Alberto A. Gaggero and Claudio A. Piga (2015). "Airline pricing under different market conditions: Evidence from European Low-Cost Carriers." *Tourism Management*, 47: 152–163.

Borenstein, Severin and Nancy L. Rose (1994). "Competition and Price Dispersion in the US Airline Industry." *Journal of Political Economy*, 102(4): 653–683.

Chiang, Wen-Chyuan, Jason C.H. Chen and Xiaojing Xu (2006). "An overview of research on revenue management: current issues and future research." *International Journal of Revenue Management*, 1(1): 97–128.

Chiou, Yu-Chiun and Chia-Hsin Liu (2016). "Advance purchase behaviors of air tickets." *Journal of Air Transport Management*, 57: 62–69.

Ciliberto, Federico and Jonathan W. Williams (2014). "Does Multimarket Contact Facilitate Tacit Collusion? Inference on Conduct Parameters in the Airline Industry." *RAND Journal of Economics*, 45(4): 764–791.

Dana, James D. (1998). "Advance-Purchase Discounts and Price Discrimination in Competitive Markets." *Journal of Political Economy*, 106: 395–422, URL https://www.jstor.org/stable/10.1086/250014.

Dubé, Jean-Pierre and Sanjog Misra (2023). "Personalized Pricing and Customer Welfare." *Journal of Political Economy*, 131(1).

Escobari, Diego (2012). "Dynamic Pricing, Advance Sales and Aggregate Demand Learning in Airlines." *The Journal of Industrial Economics*, 60(4): 697–724.

Escobari, Diego, Paan Jindapon and Nicholas G. Rupp (2021). "Bundling in Advance Sales: Theory and Evidence from Round-trip vs Two One-way Tickets." *working paper*, URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3857885.

Gale, Ian L and Thomas J Holmes (1993). "Advance-purchase discounts and monopoly allocation of capacity." *The American Economic Review*: 135–146.

Gallego, Guillermo and Garrett van Ryzin (1994). "Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons." *Management Science*, 40: 999–1020, URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.40.8.999.

Gerardi, Kristopher S. and Adam Hale Shapiro (2009). "Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry." *Journal of Political Economy*, 117(1): 1–37.

Grosche, Tobias, Rothlauf Franz and Armin Heinzl (2007). "Gravity models for airline passenger volume estimation." *Journal of Air Transport Management*, 13: 175–183.

Honka, Elisabeth, Raluca Ursu and Stephen Seiler (2024a). "Consumer Search: What Can We Learn from Pre-Purchase Data?" *Journal of Retailing*.

Honka, Elisabeth, Raluca Ursu and Stephen Seiler (2024b). "The Sequential Search Model: A Framework for Empirical Research." *Working Paper*.

Ito, Harumi and Darin Lee (2007). "Domestic Code Sharing, Alliances, and Airfares in the U.S. Airline Industry." *The Journal of Law and Economics*, 50(2): 355–380, URL https://www.jstor.org/stable/10.1086/511318.

Lazarev, John (2013). "The Welfare Effects of Intertemporal Price Discrimination: An Empirical Analysis of Airline Pricing in U.S. Monopoly Markets." *Working Paper*.

Lewis, Matthew S. (2021). "Identifying airline price discrimination and the effect of competition." *International Journal of Industrial Organization*, 78, URL `https://www.sciencedirect.com/science/article/pii/S0167718721000540`.

Li, Jun, Nelson Granados and Serguei Netessine (2014). "Are Consumers Strategic? Structural Estimation from the Air-Travel Industry." *Management Science*, 60(9): 2114–2137.

Luttmann, Alexander (2019). "Are Passengers Compensated for Incurring an Airport Layover? Estimating the Value of Layover Time in the U.S. Airline Industry." *Economics of Transportation*, 17: 1–13.

Marsh, Alex, Garrett Scott and Jonathan W. Williams (2024). "Revenue Management with Reallocation." *Working Paper*.

McWeeny, Dennis B. (2023). "Spatial Competition in the Airline Industry?" *forthcoming RAND Journal of Economics*.

Nocke, Volker, Martin Peitz and Frank Rosar (2011). "Advance-purchase discounts as a price discrimination device." *Journal of Economic Theory*, 146: 141–162, URL `http://www.sciencedirect.com/science/article/pii/S0022053110001018`.

Orlov, Eugene (2011). "How does the Internet Influence Price Dispersion? Evidence from the Airline Industry." *The Journal of Industrial Economics*, 59(1): 21–37, URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6451.2011.00444.x`.

Seiler, Stephen, Adam Smith and Ishant Aggarwal (2023). "Optimal Price Targeting." *Marketing Science*, 42(3).

Su, Xuanming (2007). "Intertemporal Pricing with Strategic Customer Behavior." *Management Science*, 53(5): 726–741, URL `https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1060.0667`.

Talluri, Kalyan T. and Garrett J. van Ryzin (2006). *The Theory and Practice of Revenue Management*, volume 1. Boston: Kluwer Academic Publishers.

van Ryzin, Garrett and Gustavo Vulcano (2008). "Simulation-Based Optimization of Virtual Nesting Controls for Network Revenue Management." *Operations Research*, 56(4): 865–880.

Wen, Chieh-Hua and Po-Hung Chen (2017). "Passenger booking timing for low-cost airlines: A continuous logit approach." *Journal of Air Transport Management*, 64: 91–99.

White, Quinton (2023). "Taxation of Multi-Product Firms with Cost Complementarities." *Working Paper*.

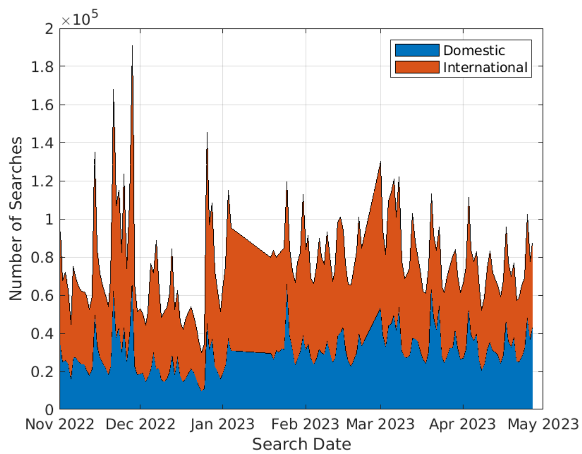
Williams, Kevin R. (2022). "Dynamic Airline Pricing and Seat Availability." *Econometrica*, 90(2): 831–858.

Appendix A Additional Empirical Findings

A.1 Time of Year & Day of Week

The six months of search data allow for a unique perspective on the seasonality of demand for air travel. We consider the impact of destination on search in Figure A.1. We find that searches peak in the final two weeks of November, approximately a month before the end-of-year holiday travel. In November, searches climb to more than twice the daily average in the sample. After a decline through most of December, searches rise in the new year until the beginning of March as consumers prepare for spring and summer travel. During this time, a larger proportion of consumers search for international flights.

Figure A.1: Searches by Date



Notes: The figure tracks the total number of searches over time by destination: domestic or international. There are two weeks in January of 2023 for which we have no observations in the web-traffic data.

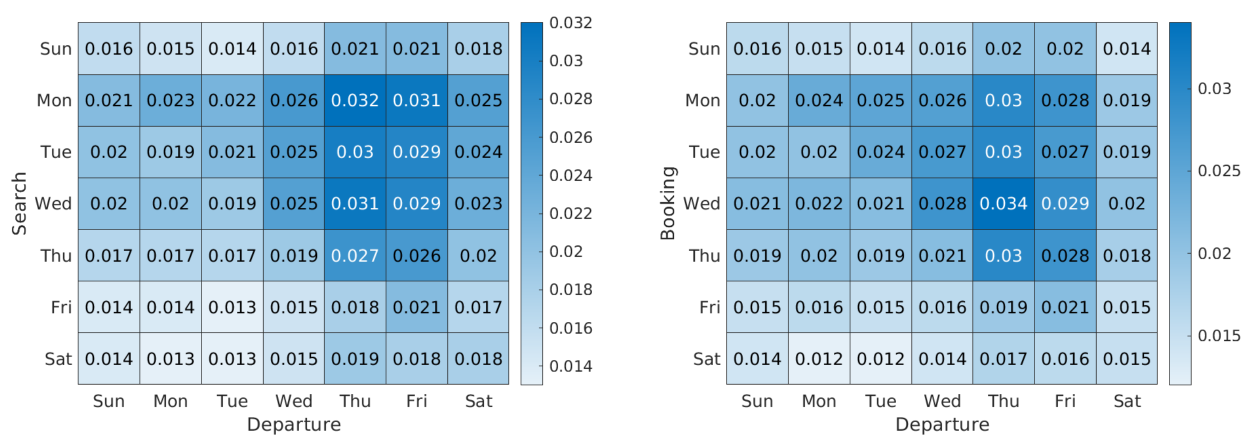
We detail our findings on within-week patterns in the website data in Table A.1 and Figure A.2. To summarize, more than one-third of the consumers search for Thursday or Friday flights, and the median consumer is quoted a price within 1% of the market median quoted price. The consumers that book flights for Thursday or Friday tend to find cheaper fares than other consumers. The median transaction price is between 2% and 3% less than the market median transaction price. Much like Friday, Saturday is popular flight day for visitors to the website, and the median quote price awaiting these consumers is 2.4% greater than the market median. The high price seems to deter some consumers because the proportion of bookings for Saturday flights is less than all other days. Monday is the most popular day to conduct a search across all departure days.

Table A.1: Prices by Day of Week of Departure

	Queries	Δ Quoted Price	Bookings	Δ Transaction Price
Day of Week				
Monday	0.121	0.001	0.129	0.051
Tuesday	0.119	-0.027	0.130	0.000
Wednesday	0.141	-0.027	0.147	-0.019
Thursday	0.177	-0.004	0.180	-0.027
Friday	0.175	0.010	0.169	-0.018
Saturday	0.145	0.024	0.120	-0.003
Sunday	0.122	0.019	0.125	0.051

Notes: The table provides the proportion of queries, the proportion of bookings, and the median deviation from the market median quoted/transaction prices for a given day of departure. The variable is constructed by taking the price minus the market median price, then normalizing by the market median price.

Figure A.2: Day of Week Statistics



(a) Searches & Departures

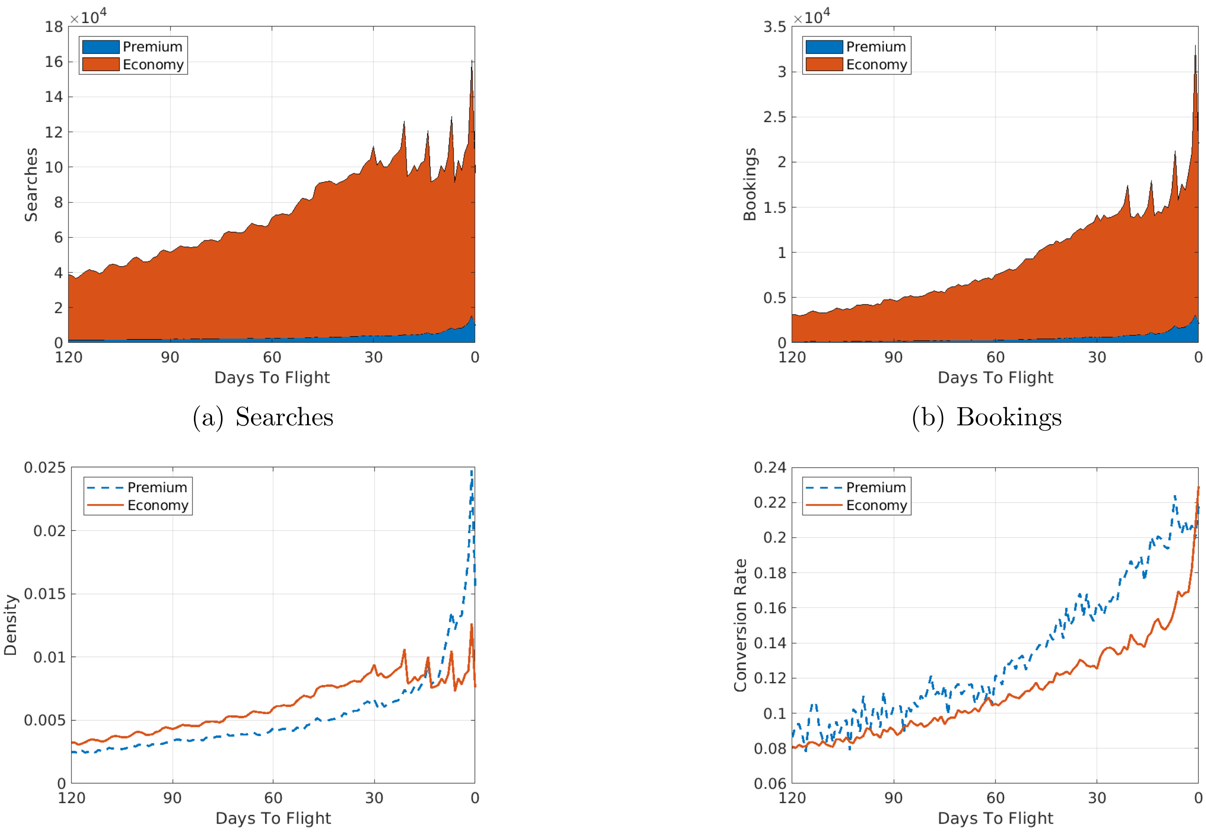
(b) Bookings & Departures

Notes: The figure displays heat maps of the proportion searches (a) and bookings (b) by day of week for a given departure day of week.

A.2 Temporal Patterns

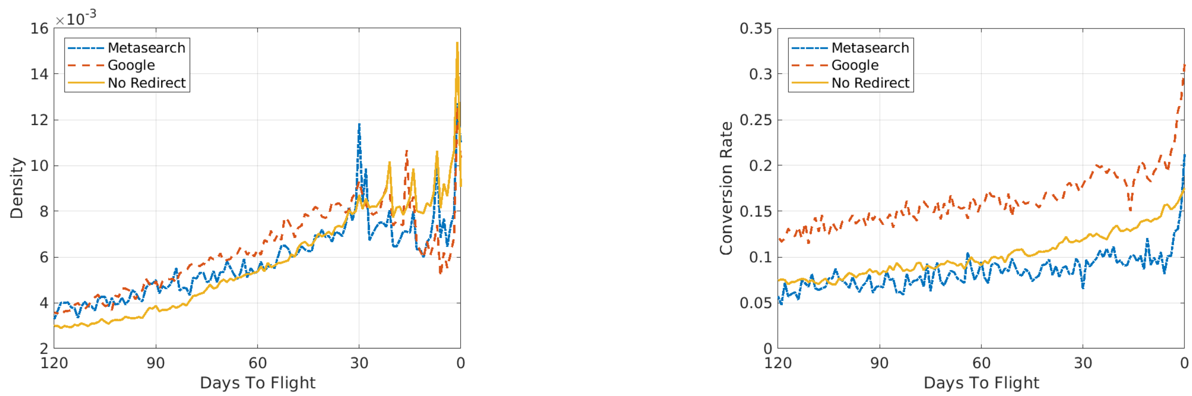
Figures A.3 and A.4 show search and booking patterns by cabin and redirection, respectively. We observe in Figure A.3 that the economy cabin is searched and booked relatively more frequently than the premium cabin. However, many premium cabin searches occur in the last week before departure, and the conversion rate of these searches is up to five percentage points greater than the conversion rate of economy cabin searches. Figure A.3 shows that redirected searches peak roughly one month before departure, but searches without a redirection continue to climb. Loyalty members often search later in the time horizon, and given their familiarity with the airline, they would have no need to be redirected to the airline's website.

Figure A.3: Temporal Patterns by Cabin



Notes: The figure displays the number of searches (a) and the number of bookings (b), as well as the density of searches (c) and the conversion rate of searches (d) at a given number of days before departure by cabin.

Figure A.4: Temporal Patterns by Redirection



(a) Search Density

(b) Conversions

Notes: The figure displays the density of searches (a) and the conversion rate of searches (b) at a given number of days before departure by redirected website.

A.3 Regressions with Time Interactions

Table A.2: Timing of Search Multinomial Logistic Regression (Merged Sample)

	15-30 NDO	8-14 NDO	0-7 NDO
Variables			
Intercept	-0.48***	-1.08***	-0.56***
	(0.04)	(0.05)	(0.04)
Nonstop	-0.01	0.01	0.09***
	(0.02)	(0.02)	(0.02)
Round Trip	-0.27***	-0.52***	-1.13***
	(0.01)	(0.02)	(0.02)
Economy	-0.23***	-0.44***	-0.72***
	(0.03)	(0.04)	(0.03)
Loyalty Member	-0.04***	0.09***	0.13***
	(0.01)	(0.02)	(0.02)
Family	0.07***	-0.07**	0.04
	(0.02)	(0.04)	(0.04)
Single Adult	0.31***	0.49***	0.72***
	(0.01)	(0.02)	(0.02)
Metasearch	-0.19***	-0.30***	-0.52***
	(0.07)	(0.10)	(0.09)
Fit statistics			
N: 190,356			
BIC: 440,865.6			

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The table shows the results of a multinomial regression for consumer and ticket characteristics on temporal arrival rates. For each variable, the omitted group is website visits between 31 and 90 days before departure. Website visits earlier than 90 days before departure are excluded.

Table A.3: Transaction Price Regression (Time Interactions)

	15-30 NDO	8-14 NDO	0-7 NDO
Variables			
Intercept	323.51***	383.72***	518.69***
	(7.20)	(9.10)	(13.36)
Nonstop	-38.93***	-55.59***	-61.93***
	(2.29)	(4.00)	(8.32)
Round Trip	-2.32	-5.21*	-43.47***
	(1.57)	(2.04)	(3.95)
Economy	-295.48***	-294.35***	-287.09***
	(7.22)	(8.58)	(10.23)
Loyalty Member	14.93***	14.74***	5.80*
	(1.47)	(1.61)	(2.66)
Family	-29.58***	-33.63***	-61.20***
	(2.40)	(4.35)	(4.37)
Single Adult	12.34***	20.27***	5.49*
	(1.54)	(2.21)	(2.64)
Metasearch	-34.65***	-49.83***	-65.56***
	(4.03)	(7.64)	(7.76)
Fixed-effects			
Market: Yes			
Day of Week: Yes			
Month-Year: Yes			
Fit statistics			
N: 990,071			
R^2: 0.49			
Within R^2: 0.27			

Clustered (Market) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: The table shows the results from explanatory variables in column (1) of Table 6 interacted with indicators based on how many days are between purchase dates and departure dates. The omitted category is purchases made by 31 and 90 days before departure.