What Can Web Traffic Reveal about Air-Travel Demand?*

Alex Marsh Garrett Scott Drew Van Kuiken
Jonathan W. Williams

May 26, 2024

Abstract

We analyze consumer search and purchase behavior using data from a major airline's website over a 6-month period in 2022 and 2023. We find large increases in demand for one-way and nonstop service from single adults and upper-tier loyalty members as the flight departure approaches. Later searches are more likely to result in purchases, consistent with traditional assumptions about business and emergency travelers arriving to airline markets closer to departure dates. We observe that consumers sometimes search for multiple destinations and departure dates within and across browsing sessions, and some consumers are redirected to the airline's website from metasearch websites. By combining the web-traffic data with contemporaneous information from Google Flights, we show that stronger competition results in a more selected sample of consumers to the airline's website and reduces the probability of purchase. Those consumers redirected to the site from metasearch websites are less likely to purchase. We discuss how our findings apply to current models of air-travel demand.

Keywords: Airlines, Demand, Search

JEL Codes: L1, L93

*We thank seminar participants for comments on this paper, and the airline for providing the data for our analysis. 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

Air-travel demand receives tremendous attention from researchers in a variety of disciplines given its important role in the global economy. Modeling efforts range widely in sophistication and breadth of analysis depending on the objective. These include complex models of passenger flows within airline networks to schedule flights (e.g., Barnhart et al. (2003)) and discrete choice models of demand to understand preferences over price and other attributes of service (e.g., Berry and Jia (2010)). Unfortunately, for many important topics related to air-travel demand, data limitations often result in strong assumptions on fundamental aspects of the model with limited empirical support.

These modeling decisions can have important implications for the development of effective public policy directed towards the industry. For example, consolidation in the airline industry has received a great deal of attention, but antitrust policy relies on market definitions that have weak empirical support. Similarly, conclusions regarding the welfare effects of dynamic pricing and other discriminatory pricing practices rely on strong assumptions about the rate and composition of arrivals of potential buyers. Recent contributions like McWeeny (2023), Aryal et al. (2023), and Marsh et al. (2024) demonstrate the value that novel data sources can have on informing modeling decisions that address such important policy questions. In this paper, we provide a detailed analysis of unique web-traffic data from a major airline to offer a stronger empirical foundation for a variety of air-travel demand modeling efforts.

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 this North American 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. The novelty of these data provide a unique opportunity to examine browsing and purchasing patterns of consumers within sessions and across sessions over time, as well as heterogeneity in patterns across markets.

We complement these novel data with information 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 connections like carrier, fare, layover time, total travel time, and connecting airports. By linking these two data sources, we can examine how competition drives selection to an airline's website and its impact on subsequent purchasing behavior.

The airline's prices reveal a considerable amount of variation within and across markets. The airline discriminates intra-temporally by screening customers across vertically-differentiated cabins at different price points, and inter-temporally discriminates by adjusting prices as passenger composition changes. Prices for premium cabin itineraries are more than double the prices for economy cabin itineraries at certain points in some markets. In addition, economy cabin prices climb as high as twice their starting value on some flights. Across markets, the airline varies prices with demand and distance. Long-haul international travelers often have prices more than 50% greater than domestic travelers on a given day before departure. Prices also vary across markets based on competition. On average, the airline's price is 15% greater than its primary competitors, but the gap is almost indistinguishable at the median.

Leveraging the detailed information in the data, we find patterns in the timing of search and the propensity to buy for several key search inputs and consumer characteristics. One-way itineraries consist of less than one quarter of the searches on the airline's website. However, the majority of one-way searches occur within the final month before departure. The rate of conversion on these searches is more than double the conversion rate of round-trip itineraries. One-way searches are typically made by adults traveling alone, a pairing that accounts for 29.1% of the bookings in the data. The mass of late bookings combined with increasing prices results in the median one-way traveler paying 6.4% more than the median traveler in their market.

Further, the data demonstrate the importance of a loyal customer base. Although only 40% of the itineraries are viewed by visitors that log in, these individuals make more than 70% of the airline's bookings. Additionally, they convert on 20% of their searches, four times as often as customers who do not log in. Of the loyalty members who do not purchase, more than 10% can be identified in return searches for flights in the same market. Within sessions, approximately 10% of customers search across multiple airport pairs. It is more common to observe customers search across multiple departure dates, often within a few days of each other. These numbers likely understate the prevalence of dynamic search in the industry, as consumers may browse on the website across many sessions without logging in or may search on other websites altogether. We do observe a portion of the customers that log in after initializing their search, and these individuals book at higher rate than those that never log in.

Loyalty members, especially those of high status, appear very price insensitive. Across all observations, the median upper-tier member pays 25.3% more than the median traveler in their market. When comparing the transaction prices for economy itineraries for these consumers to the prices posted by the airline's competitors at the time of purchase, we find a median premium of 10%, by far the largest of any variable on which we condition. On the other hand, customers redirected from other websites and travel parties with children tend to book at prices that are less than the alternatives available to them through competitors.

Related Literature

Broadly speaking, our research contributes to the rich literature on estimation of demand for air-travel. Many earlier studies utilize discrete choice models in the style of Berry (1990), Berry et al. (2006), Berry and Jia (2010), and White (2023), which assume that consumer choice sets are limited to a particular destination and date of departure from their origin airport. This definition is also fairly standard among antitrust practitioners and academic economists: for example, see Brander and Zhang (1990), Brander and Zhang (1993), Ciliberto and Williams (2014), and Bet (2023). McWeeny (2023) demonstrates the biases regarding the effects of mergers that result from using airport-pair market definitions when in fact many consumers compare across airports. Our analysis goes further and shows that consumers often compare options across multiple destinations and ranges of dates when deciding on an itinerary. Like McWeeny (2023), this finding can help reconcile the difference between the predictions of prospective structural modeling efforts and findings of reduced-form retrospective analysis regarding harm to consumers from consolidation.

We also complement more recent 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. Relatedly, 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.[1] 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. Marsh et al. (2024) demonstrate the value of these data for modeling these arrival processes that are heterogeneous across markets in their study of upgrade processes like auctions and check-in purchases that re-allocate passengers between aircraft cabins.

Our findings also contribute to the broader airline literature by demonstrating the value

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

of dis-aggregated data for a variety of topics. For example, Berry and Jia (2010) use the Department of Transportation's DB1B survey data to estimate demand and 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 differentiated offerings, but much of the differentiation is due to inter-temporal price discrimination by carriers rather than ticket restrictions and other itinerary features. This is consistent with the idea that airlines change the price of each product in response to the meaningful variation in the composition of passengers as the departure date approaches.[2] Observing characteristics of searches and visitors on the site (e.g., type of itinerary, number in party, children, and loyalty status) also creates opportunities for personalized pricing. Babii et al. (2024) demonstrate the challenges of implementing modern machine learning methods to target discounts in the industry.[3] Related, there are numerous recent studies in other industries that have shown the importance for pricing decisions of data that doesn't obscure pre-purchase consumer behavior like Honka et al. (2024a) and Honka et al. (2024b).

2 Data

We use two unique sets of data for our analysis. The first contains detailed information from searches conducted by visitors to a North American airline's website. 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. 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 applicable.

The second set of data consists of a panel of itineraries scraped from the Google Flights website. An observation is a one-way economy class itinerary available at the time of scraping for a given departure date and origin-destination pair in the airline's network. The scraped details include the carrier, the cheapest available price, the time in transit, and the flight connections. Due to computational limitations, we gather the data from sixteen origin-destination pairs (eight bi-directional markets) in the airline's network every seven days for the final two months leading up to departure and every fourteen days up to five months before departure.

[2]See Borenstein and Rose (1994) for the effect that competition has on price dispersion.

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

2.1 Sample Selection

The Google Flights data are available between November of 2022 and March of 2023. Observations are ordered by Google's internal ranking algorithm that places the "best" itineraries first in the search results. We begin by collapsing the data to a unique combination of 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 best overall itinerary, as ranked by Google, and the five best itineraries offered by its competitors. For the remainder of the paper, we use the median price and the median transit time (duration) of the five best alternative itineraries to summarize the competition.

The airline's website data span from November of 2022 to April of 2023.[4] We collapse 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 and received a price quote to be included in our sample. The final sample contains 12,083,366 searches from 8,099 markets (i.e., origin-destination airport pairs). Figure 1 provides an overview of the variation in searches and bookings over the time span of our data. The peak in searches and bookings happens at the end of November, approximately four weeks before the lowest point on December 24^{th}. Searches are largely concentrated in a smaller set of markets: 50% drawn from seventy markets and 75% drawn from 200 markets. There is overlap between our two data sets, allowing us to match searches to alternative itineraries. This subsample contains 197,107 searches across the eight markets in the Google Flights data. The median market has 23,086 observations, while the largest market by observation count has 45,936 searches and the smallest has 4,059 searches.

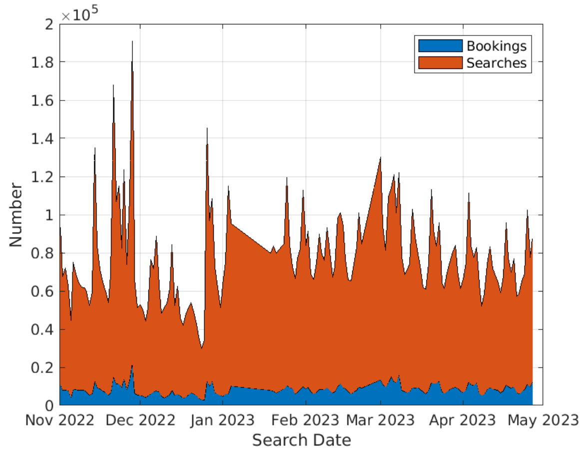
2.2 Descriptive Statistics

In Table 1, we describe the distribution of per-direction prices in the website data. We observe that premium cabin itineraries are more expensive than economy cabin itineraries in the same market and that international travelers typically pay higher prices than domestic travelers. Additionally, the data allow for new insights into the prices quoted to consumers who do not purchase. In particular, we observe that the distribution of prices quoted to consumers that do not purchase has a greater mean and a longer upper tail than the distribution of transaction prices in every itinerary segment of the data.

Table 2 summarizes the airline's prices and durations relative to its competition using the Google Flights data. For each observation in the data, we calculate the airline's price and duration as a percentage of the competition's alternative. Although prices are centered

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

Figure 1: Searches and Bookings by Date



Notes: The figure tracks the total number of searches and bookings over time in the airline's website data. There are two weeks in January of 2023 for which we have no observations in the web-traffic data.

Table 1: Transaction Price and Quoted Price

	Domestic							
	Transactions				**Quotes**			
Variable	**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							
	Transactions				**Quotes**			
Variable	**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. Observations are segmented by destination, cabin, and whether the itinerary was purchased or not.

at zero, the distribution skews heavily right. However, the airline's itineraries are shorter on average than its competition, with the mean duration just under 20% faster. In the second part of the table, we compare the prices faced by users searching for economy class itineraries on the airline's website to the competition's price. We find that the airline's website price is 33% more expensive than the competitors' price on average. The upward shift in the distribution potentially reflects the menu of options within the economy class on the airline's website. The fares that consumers face on the website may be more expensive due to additional features such as extra leg room or priority boarding, or consumers may select more expensive flights that leave at more desirable times during the day.

Table 2: Airline Price and Duration Relative to Competitor

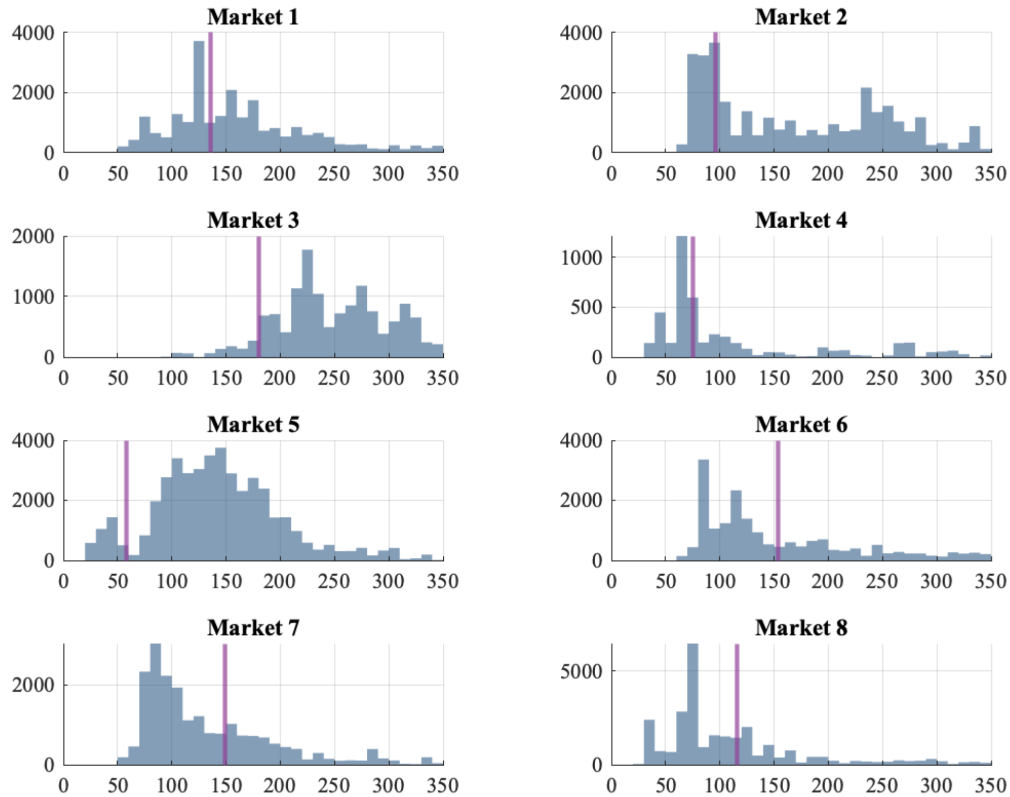
Measure	**Mean**	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Google Data						
Price	0.15	-0.52	-0.27	-0.01	0.34	1.03
Duration	-0.18	-0.72	-0.53	-0.29	-0.02	0.28
Website Data						
Price	0.33	-0.50	-0.26	0.07	0.63	1.40

Notes: The table describes the airline's prices and durations relative to its competition. In the two panels, 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, respectively. The variable is constructed by taking the airline's price minus the competitors' price, then normalizing by the competitors' price for each observation in the respective samples.

In Figure 2, we describe the competitors' prices in the Google Flights data. The histograms show the variation in the range and distribution of prices across the eight markets. Concentration exists in a few of the markets, but overall, the competitors vary prices in much the same way as the airline. The vertical lines indicate the airline's median price within the market using the observations in the Google Flights data. In some markets, the airline is very competitive on prices. However, the figure provides only a glimpse of this competition and does little to describe the time-varying competitiveness.

Figure 3 traces the joint paths of the prices posted by the airline and its competitors 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 its competitors' price decreasing. In most markets, prices hover around the 45-degree line throughout the time 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. However, in other markets, such as 4, 7 and 8, the airline and its competitors 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

Figure 2: Competitor Price by Market

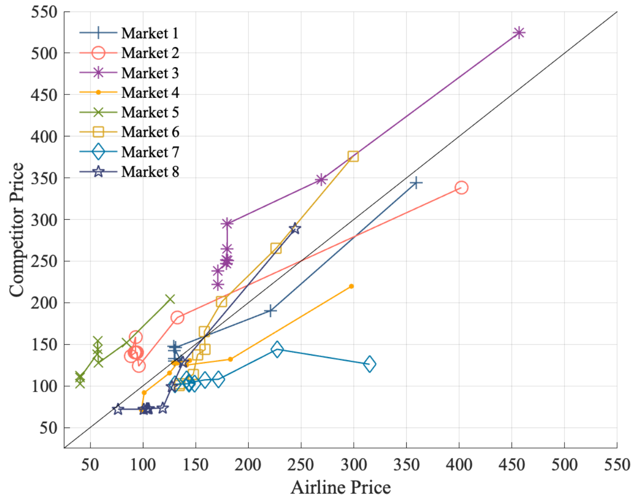


Notes: The figure shows the distribution of the competitors' prices for each market in the Google Flights data. The vertical line represents the airline's median price across all observations in the Google Flights data.

of the flights. In all eight markets, the airline offers a shorter median time in transit than its competitors.

Table 3 presents descriptive statistics from the consumers' search inputs on the airline's website. The mean statistic indicates the share of the 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%. Interestingly, the conversion rate of one-way itineraries is more than two times the conversion rate of round-trip itineraries, and loyalty members convert four times as often as the consumers who do not log in. In the next section, we use the airline's website data and the Google Flights data together to better understand how consumer- and market-level characteristics determine demand.

Figure 3: 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 its competition (y-axis).

Table 3: Itinerary Proportions and Conversion Rates

		Conversion Rate	
Variable	**Mean**	$\mathbb{I}=1$	$\mathbb{I}=0$
Itinerary:			
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. The variables include search inputs and the loyalty status if the consumer logged in. The statistics include the proportion of each variable and the conversion rates of the inclusive ($\mathbb{I}=1$) and excluded ($\mathbb{I}=0$) searches.

3 Analysis of Demand for Air Travel

Much of our understanding about the demand for air travel has been shaped by aggregated transaction data and posted prices. Given the scope of our sample, we are able to analyze the search and purchasing behavior of consumers to study the relationship between prices and bookings.

3.1 Temporal Patterns in Searches and Bookings

Recall that the website search data contain rich information regarding each itinerary searched by consumers on the airline's website. We begin by characterizing the patterns that arise in searches and bookings upon conditioning on some of the available customer information.[5] In Figure 4, 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.

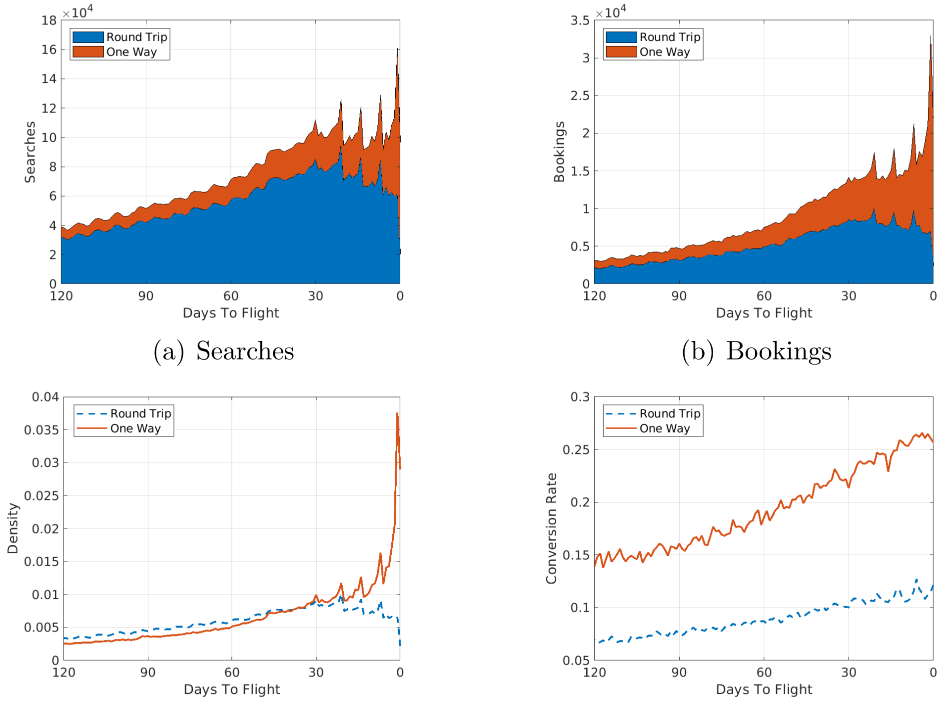
In Figure 5, we consider the same collection of figures for different types of travel parties. Consumers traveling in groups make up the majority of searches and bookings further in advance of the flight. Although the conversion rate remains fairly constant for the travel groups until the final few days before departure, the conversion rate of single adults climbs steadily over time. This increase coincides with an increase in the search density for single 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.

Consumer loyalty status is another important determinant in the search and booking process. As mentioned in the introduction, 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 6, 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.

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

[5] See Appendix A.2 for additional details about the variation in consumer behavior by the time of year and day of week.

Figure 4: Temporal Patterns by Trip Type



(a) Searches (b) Bookings

(c) Search Density (d) Conversion Rate

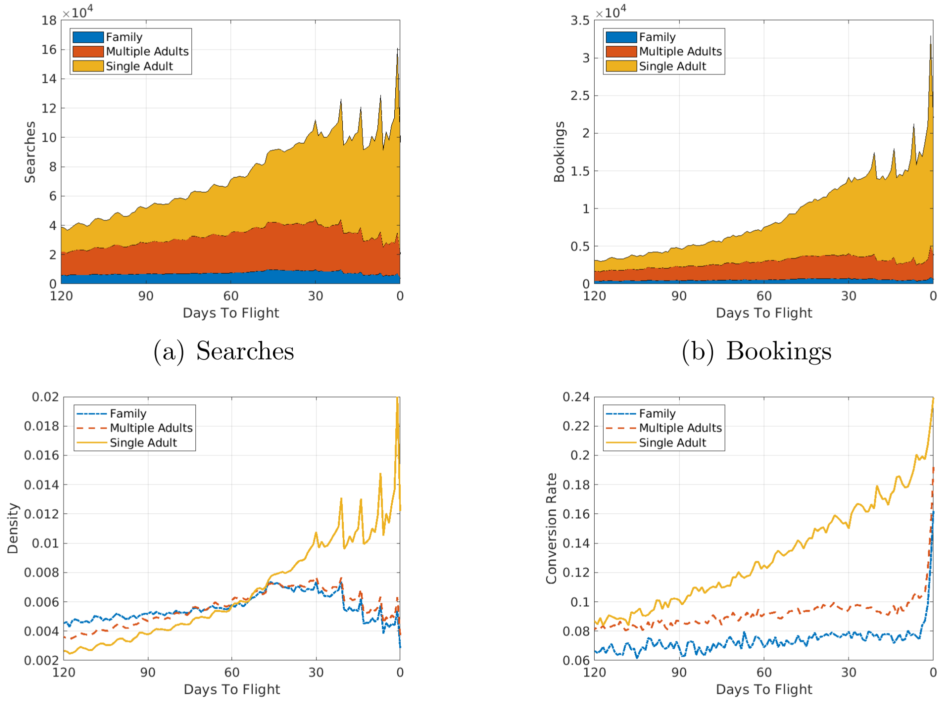
Notes: The figure displays the number of searches (a) and the number of bookings (b), as well as the density of searches (c) and the conversion rate of searches (d) at a given number of days before departure for round-trip and one-way travelers.

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 lead us to suspect that 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.

Figures 4, 5, and 6 show that one-way travelers, single adults, and upper-tier loyalty members book closer to departure on average than other population segments.[6] Because prices increase on average, these consumers pay a higher price. Table 4 allows us to compare

[6]Premium cabin travelers book close to departure, as well, but make up just slightly more than 5% of the searches. In Figure A.1 of Appendix A.1, we consider the temporal patterns in searches and bookings for premium and economy travelers.

Figure 5: Temporal Patterns by Travel Party



(a) Searches

(b) Bookings

(c) Search Density

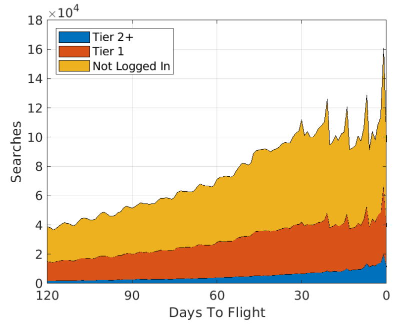
(d) Conversion Rate

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 children (families).

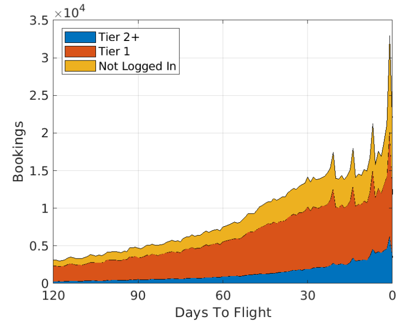
the median transaction price in each of the population segments to the median transaction price in the market in which the itinerary was booked. The median upper-tier loyalty member books 7 days closer to departure and pays a transaction price 25.3% greater than the median traveler in their market. Though smaller in magnitude, one-way travelers and single adults also book closer to departure and have positive premiums on their transaction prices, 6.4% and 3.8%, respectively. On the other hand, the median family books three weeks earlier and pays 15.0% less than the median traveler in their market.

Before discussing competition effects, we further the analysis in Table 4 using information on the originating website from which the consumer was redirected to the airline's website. These data are available in a subset of the sample beginning in January of 2023, with approximately 20% of the observations containing this information. Most consumers in the subset are not redirected from metasearch websites (e.g. Kayak, Skyscanner, etc.) or a Google domain. The transaction premium for these consumers is small, albeit positive. Of those that are redirected, the median discount is at least 15% of the market median. We

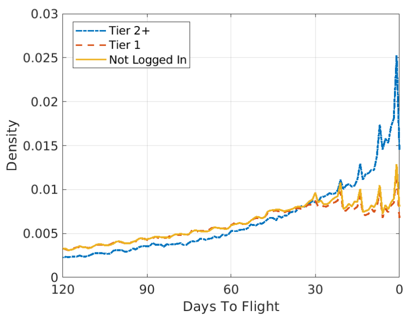
Figure 6: Temporal Patterns by Loyalty Status



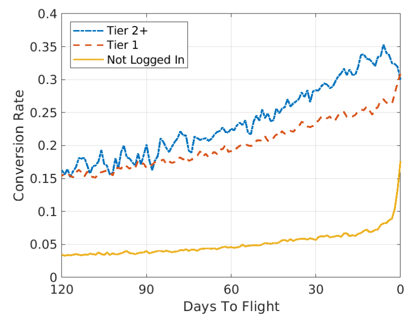
(a) Searches



(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.

plot the search densities and conversion rates of these groups in Figure A.2 of Appendix A.1 and find little variation in the either search rate or conversion rate. Thus, the difference in transaction prices may be driven by flight selection or ticket selection. Consumers arriving from metasearch websites likely book flights that are relatively cheap or choose tickets that have fewer benefits or rewards.

3.2 Competition Effects

The airline's website data illustrate the importance of itinerary and consumer characteristics in understanding demand. However, the booking decision ultimately reflects the utility the consumer receives from the chosen itinerary relative to other options they consider, which may include itineraries from other airlines. Thus far, our analysis of demand has omitted competition between the airline and its competitors, but the merged data allow us to explore the effect of competition on demand. We consider this effect on the volume of searches and

Table 4: Transaction Variables Relative to Market Median

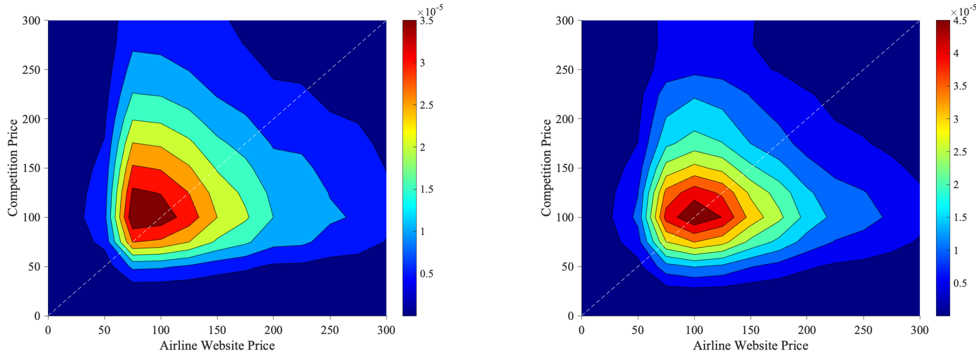
Variable	Transaction Price	Days To Flight	Mean
Trip Type:			
Round Trip	-0.031	7	0.401
One Way	0.064	-10	0.599
Travel Party:			
Single Adult	0.038	-6	0.639
Multiple Adults	-0.035	16	0.292
Family	-0.150	21	0.069
Loyalty Status:			
Not Logged In	-0.039	-4	0.268
Tier 1	-0.028	4	0.587
Tier 2+	0.253	-7	0.145
Redirection:			
Metasearch	-0.151	-3	0.044
Google	-0.229	0	0.235
No Redirection	0.013	0	0.721

Notes: The table describes consumer transaction prices and purchase days before departure relative to the median in the consumer's market. The booking proportion of each variable is provided for reference in the last column. The variable is constructed by taking the observed consumer value minus the market median value. In the case of the transaction price, the difference is then normalized by the market median price. Observations are segmented by trip type, travel party, and loyalty status.

the conversion rate of searches, both of which are impacted by the consumers' method of search. If consumers perform sequential search and use flight metasearch websites (e.g. Google Flights, Kayak, Skyscanner, etc.) to survey their options in the market before arriving to the airline's website, we would expect the volume of arrivals and the conversion rate of the searches to reflect the price of the airline's itineraries relative to its competition.

We begin by comparing the consumers' selected itineraries to the competitors' alternatives available at the time of booking. Figure 7 plots the density of the matched pairing of the price presented to a consumer on the airline's website and the corresponding competitors' price on Google Flights. Density above the 45-degree line indicates a cheaper price on the airline's website at the time of search. Most of the density is clustered around the 45-degree line, but the mass towards the upper limit of the airline's price drives the skewed distribution described in Table 2. 45% of all searches in the merged data feature a price below the competitors' price, with a medium price premium of 7%. Thus, the majority of visits to the airline's website occur when the airline's price is reasonably competitive relative to the alternatives. Panel (b) displays the subsample of observations that were converted. We find that 48% of all bookings occur when the airline's price is below the competitors' price, and the median price premium is only 2%. This finding indicates the effect of the airline's prices on conversion rates is largest when the airline's price is relatively more expensive.

Figure 7: Airline-Competition Price Joint Density



(a) All Searches (b) Bookings

Notes: The figure displays the joint density of quoted prices on the airline's website and prices scraped from Google Flights for the same market. The white dashed line is the 45-degree line, representing equivalent prices by the airline and its competition.

Although we observe redirections in our data, we cannot clearly distinguish between consumers performing sequential search and those arriving directly to the airline's website. We find that 57% of redirected bookings occur when the airline website price is below the competitor price. In addition, fewer bookings happen in the right tail of the distribution. The median redirected transaction is 6% cheaper than the median competitor price. One would assume that redirected consumers differ from the broader population along a number of dimensions, including the price elasticity of their demand. However, to the extent that redirection indicates consumers are performing search, the data are consistent with a sequential search story. When the airline's prices are expensive relative to its competitors, redirected consumers still arrive on the airline's website. In fact, the consumers generally arrive when the airline's prices are relatively high, but typically book when the airline's prices are relatively low. Such behavior is difficult to justify if consumers include all products in the market in their consideration set.

We expand the scope of our analysis by exploring price elasticities across different segments of the population. Table 5 presents descriptive statistics for a larger set of itinerary and consumer characteristics. We include the redirection statistics discussed above for reference. We calculate the median search and purchase price premiums relative to the competitors' price across various characteristic groups. We first consider these relative prices by itinerary details. Recall that in Section 3.1 we observed one-way travelers booking closer to departure on average and the median one-way traveler paying 6.4% more than the median traveler in their market. When comparing the price that one-way travelers pay relative to the options available at the time of booking (Round Trip, $\mathbb{I} = 0$), we find almost no difference in the fares.

Table 5: Search Price and Transaction Price Relative to Competitor Price

	All Searches		Transactions	
Variable	$\mathbb{I}=1$	$\mathbb{I}=0$	$\mathbb{I}=1$	$\mathbb{I}=0$
Itinerary:				
Nonstop	0.08	0.05	0.01	0.06
Round Trip	0.06	0.10	0.03	0.00
Travel Party:				
Single Adult	0.07	0.07	0.02	-0.01
Family	0.04	0.08	-0.08	0.02
Loyalty Status:				
Loyalty Member	0.04	0.10	0.02	0.00
Tier 2+	0.12	0.07	0.10	0.00
Redirection:				
Metasearch	0.13	0.07	-0.06	0.02
Google	0.20	0.07	-0.02	0.02

Notes: The table describes the quoted prices and transaction prices in the website data relative to the competitors' price. The variable is constructed by taking the observed consumer price minus the competitors' price, then normalizing by the competitors' price. The statistics are provided for the inclusive ($\mathbb{I} = 1$) and excluded ($\mathbb{I} = 0$) observations. Observations are segmented by itinerary, travel party, and loyalty status.

The second panel in Table 5 focuses on the composition of travel parties. As seen previously, single adults book closer to departure and pay higher prices on average than other travel parties within their market. This pattern seems to hold when comparing transaction prices to the broader set of alternative itineraries at the time of booking. Single adults booking on the website purchase more expensive tickets relative to the alternatives than other parties, whereas families purchase the relatively cheapest tickets of any characteristic grouping on average. The last panel of Table 5 explores the relative prices paid for a given loyalty status. We find little difference between the relative transaction price paid loyalty members and those that do not log in (2% premium versus 0% premium, respectively), but the premium of 10% paid by upper-tier loyalty members is more substantial. The upper-tier loyalty members also show a smaller gap between median relative search and transactions prices than consumers who do not log in (2 percentage point difference versus 10 percentage point difference). Together, these findings suggest that the upper-tier loyalty members are less price elastic and more likely to purchase tickets when the airline's price exceed the competitors' price.

The analysis in Sections 3.1 and 3.2 highlights the impact that the airline's prices have on demand. For the remainder of this section, we focus on understanding other drivers of purchasing behavior. With detailed information about website sessions, we can ask, conditional on searching for a flight, how do a consumer's characteristics, a flight's characteristics, and the characteristics of the other options available to consumers shift the rate at which consumers purchase tickets from the airline?

Overall, 15% of searches convert to bookings in the merged data. To capture the effects of the variables of interest, we estimate the probability that any given visit to the airline's website results in a booking. In Table 6, we report the average marginal effects, which represent the average response of the purchasing probability to each variable, estimated by a logistic regression. Each specification includes a market fixed effect, as our analysis is focused on within market variation in purchase probabilities. The figures in Section 3.1 suggest that the conversion rate rises as departure nears across most segments of the population. The first column captures this effect by simply controlling for the number of days out that the search occurred. We create four time periods and use the searches beyond 30 days before departure as our excluded group. We find that the time period effect size doubles in each successive period, moving from 1% higher than the baseline for bookings between 15 and 30 days before departure to 5% higher in the last three days before departure.

The second and third columns include consumer and itinerary characteristics. In column (2), we find that single adult, nonstop, and loyalty member searches are more likely to convert to sales on average. However, family, economy cabin, and round trip searches are less likely to convert. Conversion probability is decreasing in the airline's price relative to the competitors' price, though the magnitude of the effect is small. Surprisingly, the date indicators become negative or insignificant. This suggests that the increase in conversion rate over time comes from changes in the composition of consumers.

As departure approaches, the share of consumers with a high propensity to purchase (e.g. one-way travelers, single adults, and upper-tier loyalty members) increases in the searching population. We find that one-way travelers drive a majority of the changes in conversion rates over time. One-way travelers comprise roughly 30% of all purchases three months before departure; one day before departure, they comprise 70% of all purchases. The coefficient on the round trip variable is negative and large in magnitude. We hypothesize that this is due to the situations in which one-way tickets are typically purchased, which are often price-insensitive situations, like a flight change, a business trip, or a family emergency. Alternative specifications of the regression with the round trip variable excluded yield positive and significant coefficients on the time period indicators.

Loyalty member status increases purchase probability by 22% on average, a very large increase consistent with Figure 6. In the robustness check performed in column (3), we investigate whether conversion rates differ across loyalty tiers and find a positive relationship,

Table 6: Logistic Regression of Purchase Probabilities

	Dependent Variable: Booking			
Independent Variable	(1)	(2)	(3)	(4)
1-3 NDO	0.05***	-0.02*	-0.02*	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)
4-14 NDO	0.02***	-0.00	-0.01	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
15-30 NDO	0.01***	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Economy		-0.03**	-0.02**	-0.04***
		(0.01)	(0.01)	(0.01)
Nonstop		0.05***	0.05***	0.04***
		(0.01)	(0.01)	(0.01)
Round Trip		-0.14***	-0.14***	-0.13***
		(0.01)	(0.01)	(0.01)
Single Adult		0.06***	0.06***	0.05***
		(0.01)	(0.01)	(0.01)
Family		-0.02***	-0.02***	-0.02***
		(0.00)	(0.00)	(0.01)
Loyalty Member		0.22***		0.22***
		(0.01)		(0.01)
Tier 1			0.21***	
			(0.01)	
Tier 2			0.24***	
			(0.01)	
Tier 3			0.25***	
			(0.01)	
Tier 4			0.27***	
			(0.02)	
Price Relative to Comp.		-0.01***	-0.01***	-0.01***
		(0.00)	(0.00)	(0.00)
Metasearch				0.04***
				(0.01)
N	190,356	190,356	190,356	133,205

Notes: The table shows the results of a logistic regression, which calculates the probability that a search on the airline's website resulted in a booking. Data is limited to website visits within 90 days of departure. We include controls for days before departure, itinerary and consumer characteristics, the price premium relative to the price available from competitors, and an indicator for whether the consumer was redirected from a metasearch website.

indicating that purchase probabilities increase in loyalty tier. Consistent with specifications in columns (1) and (2), overall loyalty member conversion rates are roughly 20% higher than the conversion rates of consumers who do not log in. Conversion rates increase in tier status, as Tier 4 loyalty members convert at a 6% higher rate than Tier 1 members.

In the last column, we include an indicator for redirections, which become available in January of 2023. The indicator captures the effect of a consumer being redirected to the airline's website on the conversion rate. We find that redirections convert at a 4% higher rate than those that arrive directly to the website. This finding is interesting in the context of Table 5. Redirected consumers typically observe a relatively high price compared to the alternatives when they arrive on the website. However, these consumers are more likely to book a ticket, and the transaction price is typically less than the alternatives.

4 Conclusion

Many researchers and practitioners have dedicated substantial effort to understanding the demand for air travel. In this paper, we used novel web-traffic data from a major airline to analyze the search and booking behavior of air travelers. We show that patterns in this behavior vary with itinerary and consumer characteristics. One-way travelers, single adults, and loyalty members arrive closer to departure, book relatively more frequently, and pay more on average than other consumers within their market. In a subsample of markets for which we obtained posted price data of competitors, we found that loyalty members often pay higher prices through the airline than the prices available through the airline's competition. On the other hand, travel parties with children and consumers redirected to the airline's website typically transact at lower prices than the market alternatives.

We have shown that a portion of consumers do not limit their search to a single destination or a single 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. These findings challenge common assumptions made in the academic literature and in public policy settings, but they also invite thought and discussion about the limitations and costs of such assumptions. We hope this work encourages the continued use of novel data sources to inform air-travel demand modeling efforts.

References

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*.

Barnhart, Cynthia, Peter Belobaba and Amedeo R. Odoni (2003). "Application of Operations Research in the Air Transport Industry." *Transportation Science*, 37(4): 368–391.

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.

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

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

Brander, James A. and Anming Zhang (1990). "Market Conduct in the Airline Industry: An Empirical Investigation." *The RAND Journal of Economics*, 21(4): 567–583, URL `http://www.jstor.org/stable/2555469`.

Brander, James A. and Anming Zhang (1993). "Dynamic oligopoly behaviour in the airline industry." *International Journal of Industrial Organization*, 11(3): 407–435, URL `https://www.sciencedirect.com/science/article/pii/0167718793900177`.

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.

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

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*.

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

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

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*.

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

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

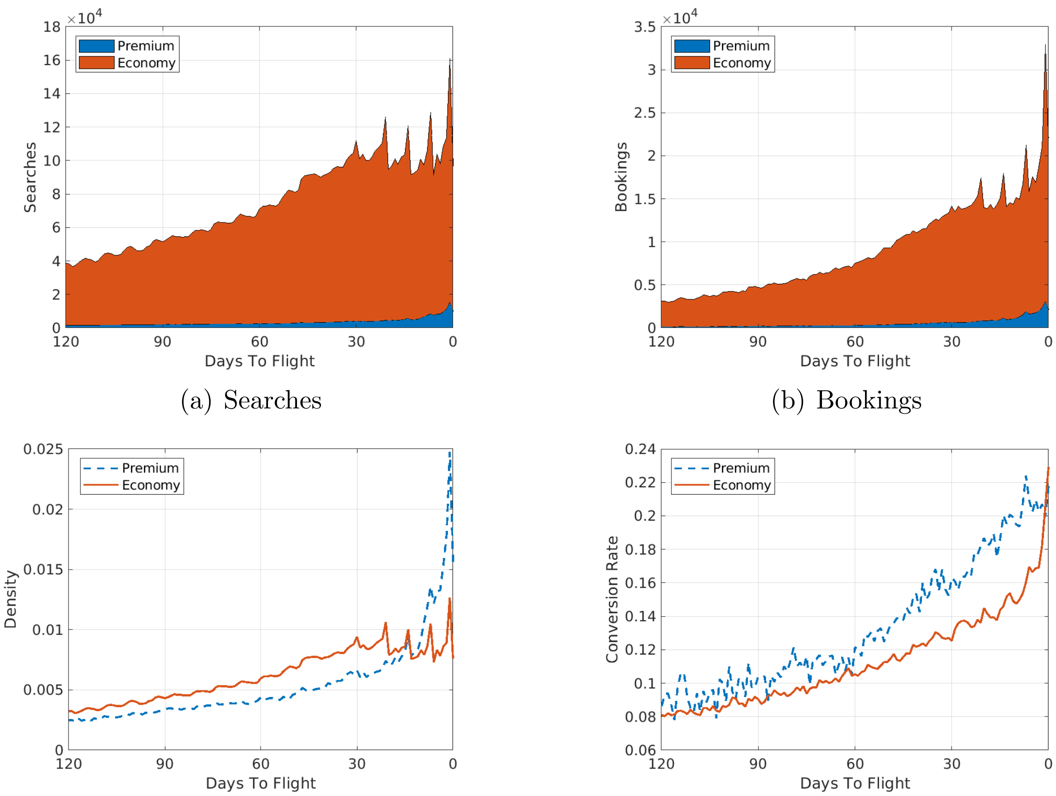
Williams, Kevin R. (2022). "Dynamic Airline Pricing and Seat Availability." *Econometrica*, 90(2): 831–858.

Appendix A Additional Empirical Findings

A.1 Temporal Patterns in Searches and Bookings

Figures A.1 and A.2 show search and booking patterns by cabin and redirection, respectively. We observe in Figure A.1 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.1 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.1: Temporal Patterns by Cabin

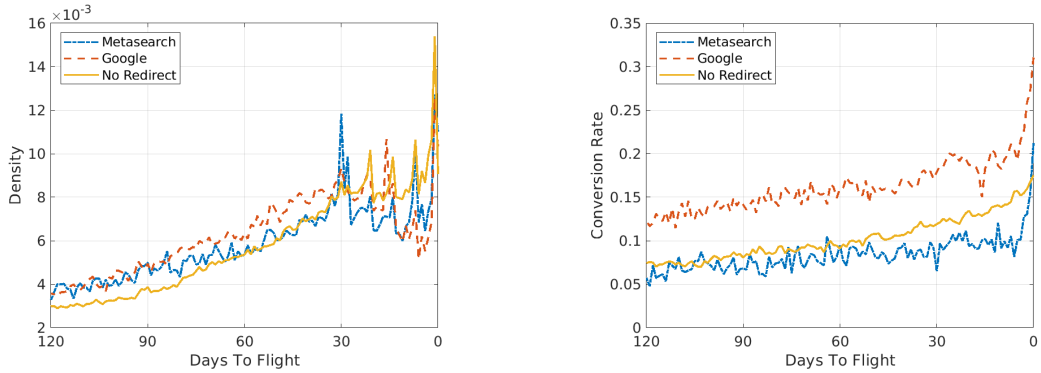


(a) Searches (b) Bookings

(c) Search Density (d) Conversions

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.2: Temporal Patterns by Redirection



(a) Search Density (b) Conversions

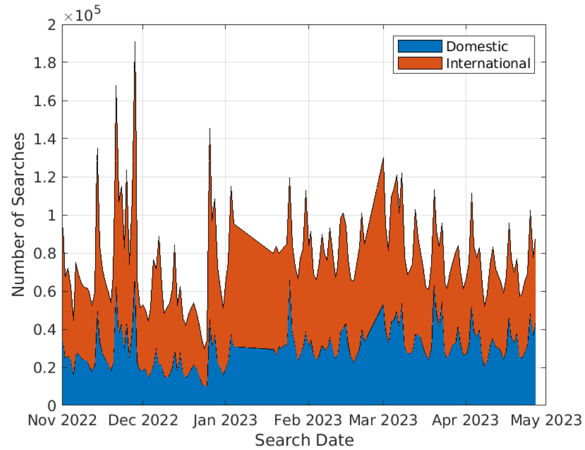
Notes: The figure displays the density of searches (a) and the conversion rate of searches (b) at a given number of days before departure by redirected website.

A.2 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.3. 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.

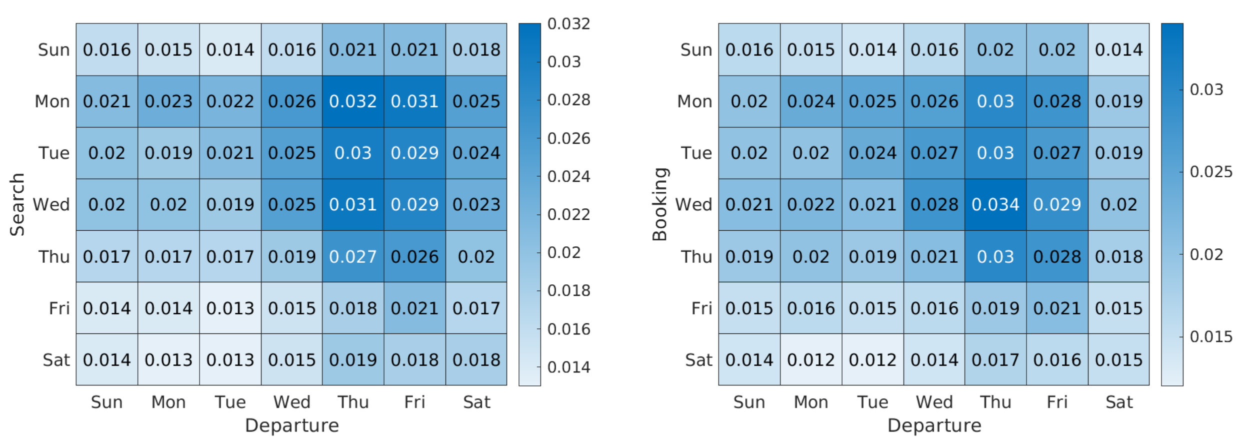
We detail our findings on within-week patterns in the website data in Table A.1 and Figure A.4. To summarize, more than one-third of the consumers search for Thursday or Friday flights, and the median consumer is quoted a price within 1% of the market median quoted price. The consumers that book flights 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.

Figure A.3: 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.

Figure A.4: 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.

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

Day of Week	**Queries**	**Δ Quoted Price**	**Bookings**	**Δ Transaction Price**
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.