Value of Search Aggregators

Selin Akca
UFSP Social Networks
University of Zurich

Anita Rao
Booth School of Business
University of Chicago

November 2017

Abstract

Airlines are increasingly questioning the value of search aggregators, preferring to bypass online ticket aggregators and sell directly to consumers. However, airlines need aggregators to gain access to a wider audience. Similarly, aggregators need the presence of all airlines to make them comprehensive and useful to consumers. We investigate who benefits most in the airline-aggregator relationship. Specifically, we ask what would happen to airline and aggregator site visits and purchases in the absence of a comprehensive aggregator. We first explore consumers' search patterns on Southwest: an airline that has never been part of any aggregator. In a descriptive exercise, we find that consumers who book on Southwest are least likely to visit aggregator sites. Second, we use the 2011 American dispute with Orbitz as an exogenous event, which led to American fares no longer being displayed on Orbitz for five months. The dispute creates a clear shift in the choice set available to consumers visiting Orbitz, with the timing being driven entirely by the contract renegotiation deadline, which creates a quasi-experimental setting. We use this dispute to identify who was hurt the most – the aggregator or the airline - in the months following the dispute. Our findings indicate that the aggregator loses the most when it is not comprehensive.

1 Introduction

Search aggregators, such as Google, Kayak and Expedia, improve consumers' search experience by providing a quick and comprehensive view of all available options resulting in better matches between consumers and products. Firms also benefit because aggregators help consumers discover new or unfamiliar products (e.g., a niche hotel, a new airline route). However, firms have begun questioning the value of search aggregators because they can reduce direct traffic to firms' websites (e.g., United.com, Hilton.com) and increase comparison search both of which can negatively impact firm profits. As a result, many firms are moving online and using the internet to disintermediate and sell directly to consumers.

Focusing on the airline industry, this paper examines the value of search aggregators. Enabled by digital technology, airlines are beginning to bypass online ticket aggregators and global distribution systems, preferring to sell directly to consumers. For example, Southwest sells its tickets only on southwest.com, and is not available for purchase on any flight aggregator. Disputes between aggregators and airlines over fees have also been on the rise: American pulled its listings from Orbitz in 2011 following a dispute over distribution fees on flight reservation systems; Delta recently delisted its flights from several online travel websites.

In this paper we investigate who benefits most in the airline-aggregator relationship. While aggregators help airlines by enabling consumers to find their best match; the presence of more airlines benefits an aggregator by making it comprehensive. With airlines questioning the value of search aggregators, we ask what would happen to airline and aggregator site visits and purchases in the absence of a comprehensive aggregator.

Using detailed browsing data from Comscore, we first document that most consumers use aggregators before making a final purchase at an airline's website. Southwest, which sells tickets only directly through its website, sees an exception to this general behavior: its users are the least likely to visit aggregators.

We next turn to identifying a causal link between aggregator search and airline visits. To do so, we take advantage of a dispute between American airlines and Orbitz surrounding the payment of distribution fees. Because of this dispute American tickets were not displayed on Orbitz for nearly 5 months from December 21, 2010 - June 2, 2011. Expedia delisted American's flights for a shorter 3 month period from January 1, 2011-April 5, 2011. A Time (2011) article reported "In bypassing the online travel agents, American saves on distribution costs, but can also raise its ticket prices more easily, since its fares won't be displayed directly beside those of its competitors."

1.1 Contribution

Closely related to the incentives at work in this industry, Baye and Morgan (2001) theoretically show that, in a homogenous product market, an aggregator has incentives to gain full consumer participation but keep firm participation partial. This is because when all firms participate prices drop to marginal costs, removing incentives for the firm to pay a fee to the aggregator or for the consumer to use the aggregator's site. The gatekeeper would rather have some firms not participate to encourage price dispersion. This perhaps can explain why Southwest is not part of an aggregator and advertises itself as a "low-cost" carrier. This can also explain why the disputes lasted only a few months, and did not lead to a new equilibrium with airlines permanently absent from aggregators. Our goal is to empircally identify who benefits most in the airline-aggregator relationship and identify how consumers react to the absence of a prominent airline from an aggregator.

Importance of aggregators has been empirically analyzed in the news industry (Athey et al 2016, Chiou and Tucker 2017, Calzada and Gil 2016) where similar disputes have arisen between aggregators and news outlets. However, while news aggregators can act as complete substitutes to news outlets threatening the outlets' revenue stream, in our setting the aggregator merely serves as an additional channel with airlines still earning most of the revenues.

2 Institutional Details

Understanding the reason for the disputes between airlines and aggregators requires an understanding of the way tickets are distributed and the various revenue models in place.

Revenue sources of OTAs and meta-search sites

Airline search aggregators consist of online travel agencies (OTAs) such as Orbitz and Expedia, and meta-search sites such as Kayak. When a consumer purchases a ticket from an OTA, the OTA earns a commission from the airline. If a consumer uses an OTA to find her best match and then goes on to the airline website to buy her ticket, the OTA does not get paid. Meta-search sites on the other hand earn revenues through referral fees from directing consumers to the airline's website, irrespective of whether a purchase is made.

Both OTAs and meta-search sites also earn revenues from ad placements. Booking fees were an additional source of revenue for the OTAs, eliminated prior to the timespan we study in our paper: Priceline and Expedia's Hotwire stopped charging fees in 2007 and Travelocity and Orbitz eliminated fees in 2009. Recently, some of the aggregators have reinstated these

booking fees.

Distribution fees

All OTAs use Global Distribution Systems (GDSs) which are flight reservation systems. Every time a ticket is purchased through an OTA, the airline not only pays a commission to the OTA (˜$5), but also a fee to the GDS (˜$9). The GDS then passes approximately 50% of this to the OTA. To avoid paying the distribution fee, airlines have been trying to create their own reservation systems. Another reason for airlines to own their reservation systems is that GDSs do not have the capabilities to showcase add-on features such as upgrades. For these two reasons, American, at the time of its contract renewal in December 2010, wanted Orbitz to use its Direct Connect reservation system instead of relying on Sabre's GDS. Orbitz on the other hand resisted making the change, because it would mean loss of nearly half of its revenues whenever an American ticket is booked.

3 Data

To understand how consumers search and buy tickets online, we need detailed browsing data that tracks consumers' website visits across various domains. We use comScore's web behavior database that consists of a sample of 50,000 internet users randomly chosen from a cross-section of more than 2 million internet users (comScore Database Manual). We supplement this dataset with the Airline Origin and Destination Survey (DB1B) from the Bureau of Transportation Statistics to get measures of demand and prices. The DB1B database is a 10% sample of airline tickets consisting of origin, operating carrier, number of passengers and itinerary fare aggregated at the quarter-level every year. The data covers 37 airlines and over 500 airports.

From the comScore databse, we select those searches and transactions that pertain to airlines and aggregators. We select the top airlines, by passengers flown as reported in the DB1B survey, which have non-zero transactions in the comScore database. We further select those domains which have at least 10,000 searches across our panel: the cutoff is at 16,756 searches for Alaska Airlines, with the next popular domain being cheapair with only 8,240 searches. Table 1 shows the included websites, the total number of search-related visits to each domain and the number of transactions.

The number of unique users that visit airline or aggregator websites ranges from 20,000-29,000 across the years 2010-2013. comScore tracks the exact timestamp a particular website was visited. Each visit comprises a search session, and comScore tracks the duration spent, the number of pages viewed, and whether a transaction was made in that session. Table 2

summarizes these statistics across all users as well as conditional on those users who made a purchase. Furthermore, user demographics such as income, education and zipcode are present in the dataset.

A majority of visits to airline/aggregator websites (˜ 55%) occur in the month before purchase, with 80% occurring within 4 months of the purchase date. We assume that all of these visits and searches are related to the final purchased itinerary. Our dataset also includes those searches that did not result in a purchase. Of these, 35% just visit a travel website once and never return over the course of the panel, 12% search over the course of a month, and the remainder search repeatedly for more than a month but never purchase. Note, we do not know if these are directed searches toward a purchase, or random visits.

Table 1: Total Number of Searches and Transactions on Travel Domains, 2010-2013

Airlines			Aggregators		
Domain	Searches	Transactions	Domain	Searches	Transactions
aa.com	71,811	1,769	cheapflights.com	20,596	-
airtran.com	25,685	959	cheapoair.com	81,675	-
alaskaair.com	15,731	1,025	expedia.com	234,595	2,277
continental.com	23,774	629	hotwire.com	62,332	362
delta.com	82,300	1,964	kayak.com	76,976	-
jetblue.com	46,570	932	orbitz.com	115,385	1,750
southwest.com	125,276	8,429	priceline.com	138,454	1,862
united.com	45,626	1,253	travelocity.com	84,383	1,165
usairways.com	30,254	975			

Table 2: Descriptive Statistics: ComScore data

	All searches		Conditional on purchase	
	Median	Mean	Median	Mean
Number of websites visited	2	3.75	3	5.18
Time spent on search (minutes)	9	27.48	43	66.35
Total pages viewed	10	26.25	36	54.47
Transactions	0	0.9	2	4.91
Number of unique users	98313		12090	
Total purchases			24056	

3.1 Descriptive Evidence

3.1.1 Search Patterns – Aggregator visits and domain visits

For every purchased ticket, we identify how many visits were made to an aggregator, the focal airline, and any non-purchased airline, prior to making a ticket purchase. Figure 1 highlights all possible browsing patterns for a typical airline, using Delta as an example. We see that for a typical airline, the most common search pattern involves visits to all three types of sites: aggregators, the purchased airline and other airlines. 73% of all searches involves the use of aggregators at least once. The only airline in contrast to this search pattern is Southwest, where the most common pattern is an user going to Southwest and making a purchase directly without any search. Only 44% of searches use aggregators. Note that these patterns could also result if Southwest's customers are very different from other airlines' customers.

Next, we turn to identifying when in the search history aggregators were visited relative to the purchased airline. If consumers are fairly well-informed of their destination airline and are merely browsing other websites before making a purchase, we should see the focal airline being visited earlier. On the other hand, if a consumer searches an aggregator first and then subsequent airlines, we should see aggregators being visited earlier. To study this pattern, we document for all purchased tickets on an airline's domain, the rank order of the focal/purchased airline and of any aggregator. Figure 2 plots these ranks. We observe that for most airlines, the aggregator is visited first and then the focal airline. Exceptions to this pattern are southwest and jetblue, the low cost airlines, where consumers are more likely to visit the focal airline first.

While these patterns are indicative of the extent to which customers use aggregators to find their ultimate match, and that being absent from an aggregator is not necessarily harmful to an airline, they do not establish any causal effect. To do so, we turn to the American-Orbitz dispute which resulted in American being unavailable on Orbitz's website for over 5 months.

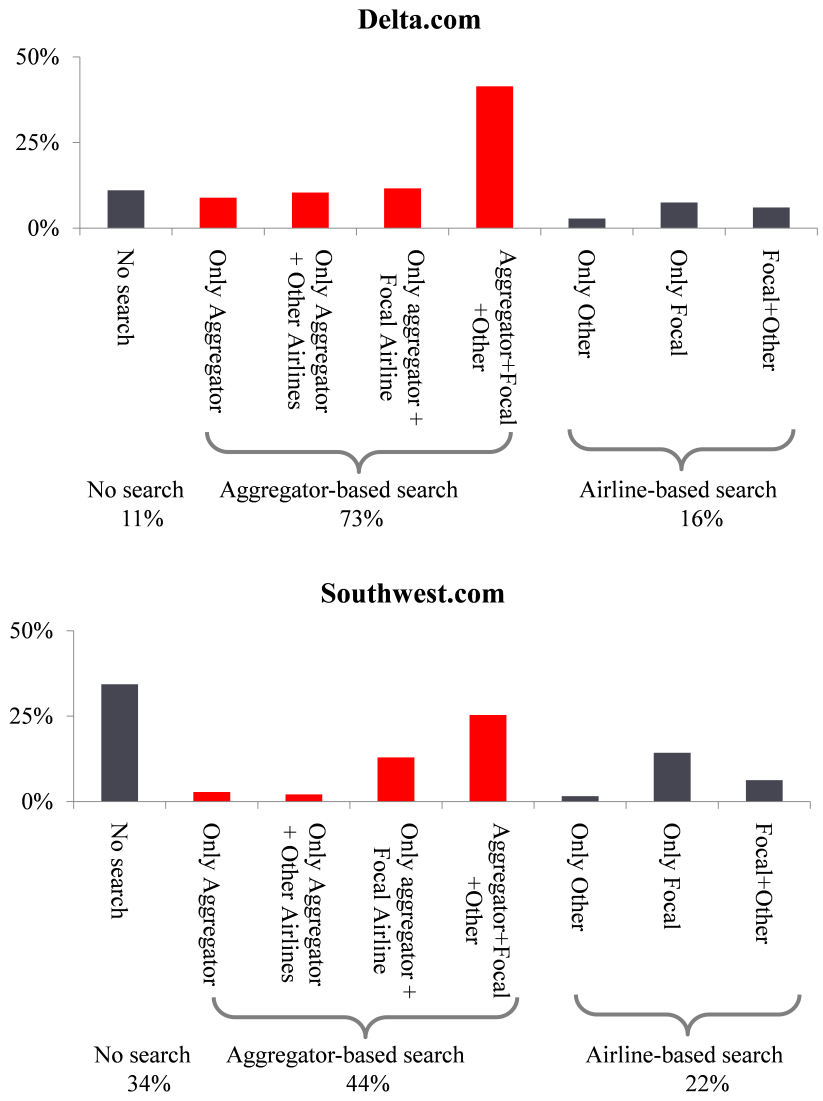


Figure 1: Southwest users go directly to southwest.com; All other airlines search an aggregator before making a purchase

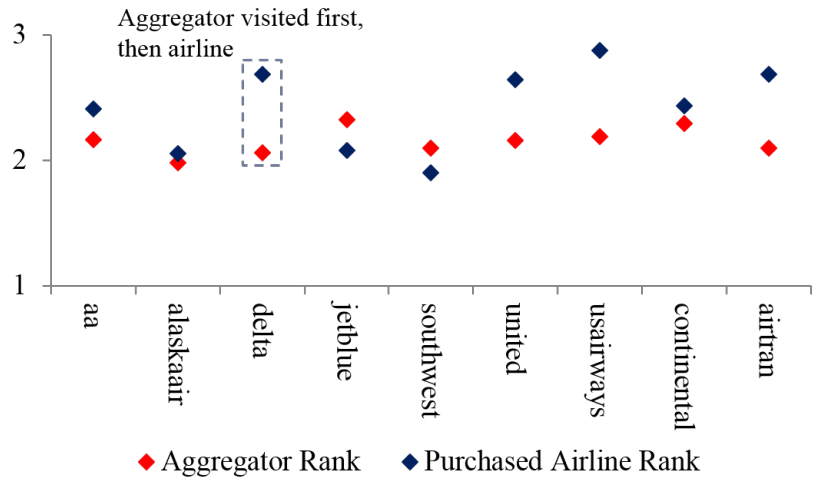


Figure 2: Consumers start their search at an aggregator, except Southwest and Jetblue users

4 Empirical Evidence

In the absence of an aggregator, airline visits can decrease, increase or stay the same: they could decrease if the value of visiting the airline's website is too low to justify the cost of visiting the site, they could increase if the value of visiting the airline is informative to the consumer, and they stay the same of visiting additional sites adds no value. Similarly aggregator visits when the aggregator is not comprehensive can either decrease or stay the same. A simple model in the Appendix illustrates these possibilities. Therefore, the true impact of an airline's absence from an aggregator is an empirical one and is dependent on the context (e.g. the airline industry is likely to behave very differently from the hotel industry).

We use the American-Orbitz dispute to identify the effect of the absence of a major airline from a popular aggregator. We examine both search and purchase behavior during this dispute period, and compare it to the same period one-year before and one-year after. For each consumer's purchased itinerary, we construct[1] three dependant variables: *prop_sites*: the number of times a domain's website was visited relative to all airline/aggregator websites in a month, *prop_pages*: the number of pages viewed at the domain's website relative to all pages visited that month, *prop_durn*: the duration spent on the website relative to the total duration spent on all airline/aggregator related websites.

We focus on only those months when an individual engaged in a travel-related search. If there is a month in which no travel website is visited by an individual, we do not include

[1]When our analysis includes all searches (including ones where there was no final purchase), we group all the non-purchase-related searches into one itinerary.

that individual-month in the data[2]. We first present how the impacted domains - aa.com, orbitz.com, and expedia.com - perform relative to the control period. As controls, we choose the same months when the dispute occurred: January-March but in the previous year and the future year: 2010 and 2012. We focus on the same set of months as the dispute period to hold any seasonality patterns that might exist constant. We conduct both pooled (treating 2010 and 2012 equally), and un-pooled regressions (treating the two years separately). In Section 4.2we compare how the impacted domains fare relative to competitors as well. Below we report results from the pooled regressions, while the Appendix provides the un-pooled regressions.

4.1 Search intensity relative to control period

For each of the impacted domains, we evaluate the search intensity during the dispute period relative to the same period in the past (2010) and future (2012) years. In the regressions below, we pool together the control years; in the Appendix we present the results where both controls are evaluated separately. For each of the impacted domains, we estimate the following regression separately:

$$y_{ijt} = \kappa_j^{\text{dispute}} \mathrm{I}\left(\mathrm{Disp}_t\right) + \boldsymbol{\kappa}_j^{\text{ctrl_dispute}} \mathrm{I}\left(\mathrm{Ctrl}_t\right) + \alpha_i \tag{1}$$

where i is the individual, j is the domain, t is time in months,
$\mathrm{I}\left(\mathrm{Disp}_t\right)$ is an indicator which equals 1 if t belongs to the dispute period: Jan-Mar 2011,
$\mathrm{I}\left(\mathrm{Ctrl}_t\right)$ is an indicator which equals 1 if t belongs to the control period: Jan-Mar 2010 or Jan-Mar 2012,
κ^{dispute} is the estimate for domain j during the dispute period, $\kappa^{\text{ctrl_dispute}}$ is the estimate for domain j during the control period, and
α_i is the individual-fixed effect.

The relevant estimate reported in Table 3 is $\kappa^{\text{dispute}} - \kappa^{\text{ctrl_dispute}}$, which indicates the proportion of searches on the impacted domains relative to the control period. We find that the duration spent on aa.com increases, and that this increase is statistically significant. The number of aa.com sites visited also increase, but is significant at a lower 90% confidence level. The aggregator, Orbitz, is significantly and negatively impacted: both duration spent

[2]In a robustness check, we expand the data to include those months when no travel-related search was performed: this allows for total searches to increase/decrease. When an individual did not search or visit any given domain in a month, the dependant variables take on the value 0. Doing so allows for the scenario where consumers increase their total search volume at a given domain when they can no longer find aa.com on an aggregator. The results are fairly similar in this specification but are quantitavely smaller in magnitude because of the large number of additional 0's.

as well as the number of sites visited decreases. Expedia on the other hand does not seem to be impacted during the dispute period.

However, it is possible that other players, e.g., delta.com, priceline.com, are also impacted similarly. To rule out any such industry-wide trend, we turn to comparing these differences relative to competitors of the impacted domains.

Table 3: Searches during dispute period on impacted domains, relative to control periods: $\kappa^{\text{dispute}} - \kappa^{\text{ctrl_dispute}}$

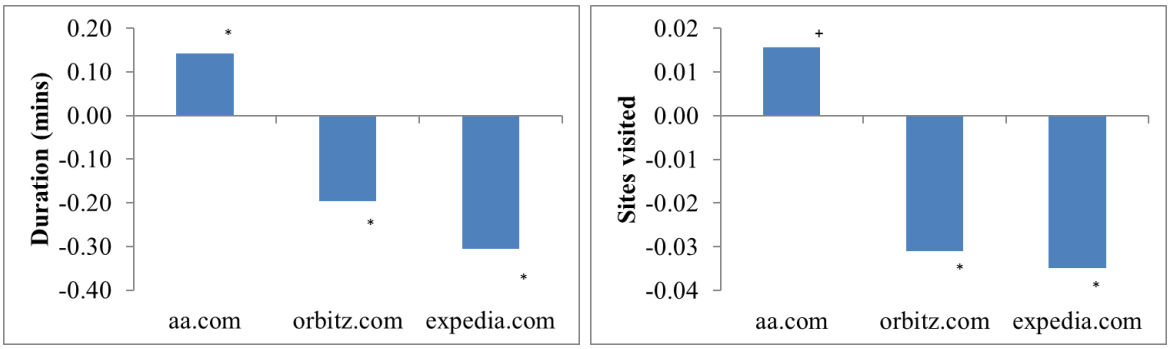
	aa.com		orbitz.com		expedia.com	
	coeff	t-stat	coeff	t-stat	coeff	t-stat
prop_sites	0.0035	1.82	-0.0053	-2.46	-0.0028	-0.83
prop_durn	0.0049	2.39	-0.0063	-2.75	-0.0008	-0.24
prop_pages	0.0014	0.66	-0.0031	-1.35	-0.0009	-0.26
Fixed-effects	machine_id					
N	258,201					
N obs	82,886					
Cluster	machine_id, purchase_id					

4.2 Search intensity relative to competitors and control period

In this analysis, we include competitors and measure the change relative not only to the control periods, but also relative to the competitors. Note that doing so measures the net competitive effect of the dispute-period on the impacted domains. In other words, if consumers shift their searches from orbitz to priceline during the dispute period, we are measuring the decrease in orbitz relative to the increase in priceline, relative to the baseline years. We further control for time-trends by adding in year-month fixed-effects, to allow for any industry-wide changes in demand. The diff-in-diff regression we estimate for American Airlines is:

$$y_{ijt} = \kappa^{\text{dispute,all}} + \kappa^{\text{dispute,F}} \mathrm{I}\left(\text{Focal}_t\right) + \boldsymbol{\kappa}^{\text{ctrl_dispute,all}} \mathrm{I}\left(\text{Ctrl}_t\right) + \boldsymbol{\kappa}^{\text{ctrl_dispute,F}} \mathrm{I}\left(\text{Ctrl}_t\right) \mathrm{I}\left(\text{Focal}_t\right) + \alpha_t + \alpha_i \tag{2}$$

where $\kappa^{\text{dispute,all}}$ captures the average estimate of all airline websites visited during the dispute period (relative to the estimate in the non-dispute period),
$\kappa^{\text{dispute,F}}$ is the additional change specific to American's website,
$\mathrm{I}\left(\text{Focal}_t\right)$ is an indicator which equals 1 if airline j is the focal airline, i.e. American,



Notes: + significant at the 90% level. * significant at the 95% level.

Figure 3: Duration spent and Sites visited Increase on aa; Decrease on orbitz and expedia

$\boldsymbol{\kappa}^{\text{ctrl_dispute,all}}$ captures the average estimate of all airline websites in the control period - preceding and succeeding years are used as controls,
$\boldsymbol{\kappa}^{\text{ctrl_dispute,F}}$ is the additional change to this vector, specific to American's website,
α_i and α_t are individual- and time- fixed-effects.

The estimate of interest is $\kappa^{\text{dispute,F}} - \boldsymbol{\kappa}^{\text{ctrl_dispute,F}}$ which measures the incremental change on aa.com relative to competitors and the base line years. We run a similar regression for the aggregators Orbitz and Expedia, where the competitors are the set of all other aggregators.

Table 4 reports the estimates of interest, $\kappa^{\text{dispute,F}} - \boldsymbol{\kappa}^{\text{ctrl_dispute,F}}$. We find that during the dispute period, both Orbitz and Expedia see a significant drop in number of pages browsed, duration spent and number of visits to its website, compared to all other aggregators, and compared to the same period the year before/after. American sees an increase in duration spent on its website, and an increase in number of sites visited (significant only at the 90% level), compared to all other airlines, and compared to the same period the year before/after. Figure 3 plots the actual (not proportion) increase/decrease in duration and site visits across all three websites.

4.2.1 Searches conditional on a purchase

We next condition on a purchased itinerary on the impacted domain and estimate Equation 2 on the resulting subset. In contrast to the finding in Table 4, Table 5 shows that for those consumers that actually made a purchase at aa.com the time spent on the domain, the pages browsed and number of times aa.com was visited are all lower in the dispute period. Orbitz and expdia do not experience a significant change, conditional on a purchase.

It is possible that during the dispute period, aa.com sees more of its loyal consumers visiting its website and purchasing: these consumers do not need to spend as much time browsing or viewing their options and make their decisions quickly explaining the decrease

Table 4: Searches on impacted domains, relative to competitors and pooled control period: $\kappa^{\text{dispute,F}} - \boldsymbol{\kappa}^{\text{ctrl_dispute,F}}$

	aa.com		orbitz.com		expedia.com	
	coeff	t-stat	coeff	t-stat	coeff	t-stat
prop_sites	0.0038	1.85	-0.0076	-3.60	-0.0085	-2.53
prop_duration	0.0048	2.21	-0.0067	-2.96	-0.0103	-3.02
prop_pages	0.0003	0.15	-0.0084	-3.77	-0.0148	-4.17
Fixed-effects	machine_id, year-month					
N	4,389,417					
Nobs	82,886					
Cluster	machine_id					

in all the search-related dependent variables. It is possible that other consumers, especially those who need to spend more time browsing, are less likely to find American's fares. In the enxt section, we explore possible sources of such heterogeneity.

Table 5: Searches conditional on purchase at impacted domain: relative to competitors and pooled control period:

$$\kappa^{\text{dispute,F}} - \boldsymbol{\kappa}^{\text{ctrl_dispute,F}}$$

	aa.com		orbitz.com		expedia.com	
	coeff	t-stat	coeff	t-stat	coeff	t-stat
prop_sites	-0.0645	-1.96	0.0234	0.69	-0.0216	-0.81
prop_duration	-0.0683	-1.92	0.0089	0.24	-0.0081	-0.28
prop_pages	-0.1002	-2.87	-0.0009	-0.02	-0.0205	-0.72
Fixed-effects	machine_id, year-month					
N	60,333		60,146		81,498	
Nobs	933		1,070		1,404	
Cluster	machine_id					

4.3 Heterogeneity: Distance from Hubs

We next turn to investigating who is likely to use the aggregators less when it is not comprehensive. One dimension of heterogeneity we investigate is how close a consumer lives to an airport where American is a prominent airline. We use the Airline Origin and Destination Survey (DB1B) database to infer the importance of each airport to each airline. We create an importance score imp_{mj} for each airline j -airport m combination as follows:

$$\text{imp}_{mj} = \frac{d_{mj}}{\sum_j d_{mj}}$$

where d_{mj} is the total number of passengers that flew from origin airport m in airline j across the years 2010-2013, $\sum_j d_{mj}$ sums across all passengers for that airport across all top airlines.

The Comscore data provides us with the zipcode of each individual user in the panel. From this, we can infer the closest airport to each user[3]. We then merge the importance measure to the Comscore data to identify if a consumer is close to an airport that is likely to be a focal airport for American Airlines. This provides a continuous measure of the relevance of an airport to American Airlines, and does not include just hubs.

We run the same regression as in Equation 2, but with an additional interaction that interacts the coefficients with the importance measure.

$$y_{ijt} = \widetilde{\kappa}^{\text{dispute}} + \widetilde{\kappa}_h^{\text{dispute}} \text{imp}_{aj} + \alpha_t + \alpha_i \tag{3}$$

$\widetilde{\kappa}_h^{\text{dispute,F}} - \widetilde{\kappa}_h^{\text{ctrl_dispute,F}}$

where $\widetilde{\kappa}^{\text{dispute}} = \left[\kappa^{\text{dispute}}, \kappa^{\text{dispute,F}}, \kappa^{\text{ctrl_dispute}}, \kappa^{\text{ctrl_dispute,F}}\right]$

We find evidence that most of the effect is driven by consumers who live far from airports where American is prominent. In other words, if a consumer is close to an airport where American is a prominent airline, during the dispute period she is less likely to be impacted, likely because she already uses aa.com to book her tickets directly. However, if she lives far away, she needs to conduct a comprehensive search to evaluate her options and uses Orbitz and Expedia less because they are not as informative anymore.

Table 6: American searches: close to AA airports vs. far: $\widetilde{\kappa}_h^{\text{dispute,F}} - \widetilde{\kappa}_h^{\text{ctrl_dispute,F}}$

	aa.com			
	close		far	
	coeff	t-stat	coeff	t-stat
prop_sites	0.0021	0.27	0.00092	1.95
prop_durn	0.0033	0.40	0.00118	2.38
prop_pages	-0.0016	-0.19	0.00029	0.57

4.4 Demand

We estimate demand using a multinomial logit, assuming a consumer's utility is given by:

[3]We compute the Euclidean distance between every airport and the center of the user's zipcode. We use this distance measure to infer the nearest airport to the user

Table 7: Orbitz searches: close to AA airports vs. far
$\widetilde{\kappa}_h^{\text{dispute,F}} - \widetilde{\kappa}_h^{\text{ctrl_dispute,F}}$

	orbitz.com			
	close		far	
	coeff	t-stat	coeff	t-stat
prop_sites	0.0030	0.69	-0.00186	-3.56
prop_durn	0.0018	0.41	-0.00168	-3.01
prop_pages	0.0027	0.58	-0.00204	-3.69

Table 8: Expedia searches: close to AA airports vs. far
$\widetilde{\kappa}_h^{\text{dispute,F}} - \widetilde{\kappa}_h^{\text{ctrl_dispute,F}}$

	expedia.com			
	close		far	
	coeff	t-stat	coeff	t-stat
prop_sites	-0.0077	-1.04	-0.00251	-2.83
prop_durn	-0.0093	-1.29	-0.00288	-3.30
prop_pages	-0.0076	-1.01	-0.00388	-4.21

$$u_{ijt}(\theta) = \alpha_j^{\text{dispute}} + \alpha_j^{\text{ctrl,dispute}} \mathrm{I}(\mathrm{Ctrl}_t) + \varepsilon_{ijt} \tag{4}$$

Assuming that the unobserved (to the researcher) shocks follow a Type-1 extreme-value distribution, the probability that individual i books at airline/aggregator j in month t is given by

$$\Pr_{ijt}(\theta) = \frac{e^{u_{ijt}(\theta)}}{\sum_{k \in C} e^{u_{ikt}(\theta)}}$$

The overall log-likelihood across all individuals can then be written as

$$LL(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{C} log \Pr_{ijt}^{\mathrm{I}_{ijt}}(\theta)$$

where I_{ijt} is 1 if individual i booked using airline/aggregator j in month t, C is the choice set available to the individual which includes all airlines, all aggregators and the outside option (no purchase).

We find that purchases on aa.com did not increase/decrease significantly compared to all other airlines and the outside option. However, we find that purchases at orbitz decreases. We compare the purchases at Orbitz.com relative to all other aggregators (as one group), all

airlines (as one group) and the outside option. Table 9 reports $\alpha_j^{\text{dispute}} - \alpha_j^{\text{ctrl,dispute}}$. Figure 4 translates these to market share differences to depict where the lost Orbitz purchases are moving to. Overall, we see that the market share of the outside option increases during the dispute period: this is evidence that the aggregator loses demand. To the extent that the tickets booked on orbitz might have been American tickets, this is also potential evidence that american might also be impacted negatively.

Therefore, to verify if total demand for American Airlines changes, we use the DB1B database which tracks passengers (demand) by origin-carrier-quarter. We run a similar regression but control for time and origin fixed-effects, as well as the mean fare charged. We find no evidence supporting a change in demand for American Airlines during the quarter of the dispute.

Table 9: Demand Estimates, $\alpha_j^{\text{dispute}} - \alpha_j^{\text{ctrl,dispute}}$

	coeff	t-stat
Outside option	0.438	2.77
Other aggregator	0.414	2.32
Airline	0.353	2.13

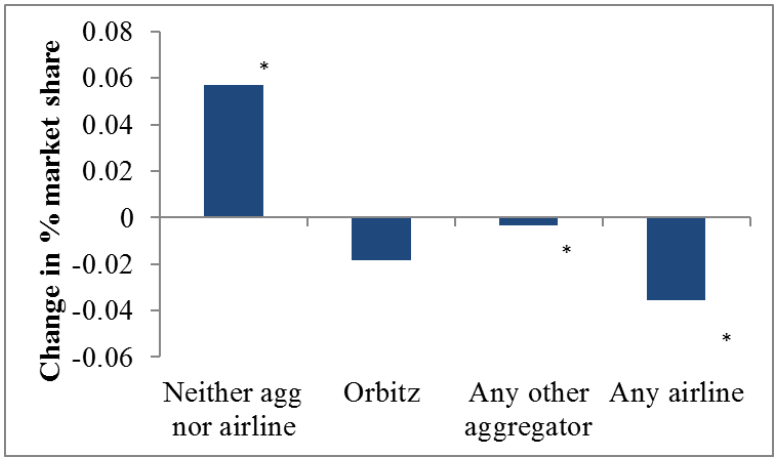


Figure 4: Market share changes compared to the control period

5 Prices

In response to the dispute, American could have altered its marketing variables, so as not to impact overall demand. We therefore check for price responses, where prices are observed using the DB1B database. This exercise is similar in spirit to Bilotkach et al (2014) who find that American's fares dropped during the dispute period. Similar to their approach, we

1) drop observations with fares which are less than \$0.02 and more than \$1 per mile which are flagged as questionable, and 2) drop markets with less than 100 passengers in total in all those 3 years. We use roundtrip itineraries to calculate passenger-weighted average fares at the airline-year-quarter-origin-destination-number of stops level. The difference between our approach and theirs is we include the future year (2012) as an additional conrol and include not just five, but the top nine airlines (to be consistent with the demand estimation).

The price regression we estimate is:

$$p_{mjkt} = \beta^{\text{dispute,all}} + \beta^{\text{dispute,F}} \mathrm{I}\left(\text{Focal}_t\right) + \beta^{\text{ctrl_dispute,all}} \mathrm{I}\left(\text{Ctrl}_t\right) + \beta^{\text{ctrl_dispute,F}} \mathrm{I}\left(\text{Ctrl}_t\right) \mathrm{I}\left(\text{Focal}_t\right) + \alpha t + \alpha_{mk} + \alpha_j \tag{5}$$

where m is an Origin-Destination pair, j is the airline, k is the number of stops and t is the quarter,

p_{mjkt} is the passenger-weighted airfare charged by airline j on market m with number of stops k in quarter t .

Figure plots the estimates of interest $\beta^{\text{dispute,F}}$ as well as $\beta^{\text{ctrl_dispute,F}}$ for 2010 and 2012. As can be seen, American fares in 2011 seems to decrease relative to 2010 and this decrease is significant. It appears that American, in response to the dispute, might have lowered its airfares. However, note that in our demand estimation (using the DB1B databse) we control for prices and find no effect of the dispute period on American's demand.

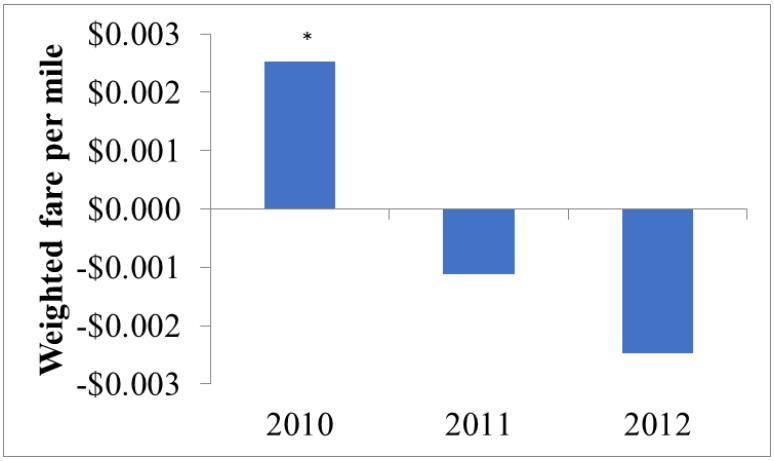


Figure 5: American prices relative to other airlines

6 Conclusion

We find that the aggregator stands to lose, in terms of consumers visiting its website and the duration spent there, when it is not comprehensive. For an aggregator where a sizable portion of the revenue comes from ad placements, this can imply a revenue decline. One caveat is that there are other aggregators that consumers can go to, making any individual agggregator relatively powerless. The findings might loook very different in a category where there is only one aggregator.

We also find evidence that indicates consumers closer to American-prominent airports are less likely to make use of aggregators when they do not contain American's fares. We are currently exploring how price changes during the dispute period, and incorporating prices in our regressions as additional controls.

References

[1] Athey, S., M. Mobius and J. Pal (2017), "The Impact of Aggregators on Internet News Consumption", Stanford GSB working paper.

[2] Baye, M.R. and J. Morgan (2001), "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets", The American Economic Review, 91(3), 454-474

[3] Bilotkach, V., N.Rupp and V. Pai (2014), "Value of a Platform to a Seller: Case of American Airlines and Online Travel Agencies," working paper.

[4] Calzada, J. and R. Gil (2016), "What do News Aggregators Do? Evidence from Google News in Spain and Germany", SSRN working paper.

[5] Chiou L. and C. Tucker (2017), "Content aggregation by platforms: The case of the news media", Journal of Economics and Management Strategy, 00:1–24

[6] Time (2011), "How Travelers Could Lose in American's Web Ticket War", by Kayla Webley, Jan 06 2011.

A Illustrative Model

Consumers typically visit both aggregators and airlines to learn about their best match value. Let the information gain from an aggregator be $N\mu$, where N is the number of airlines on the aggregator's website. Let information available at an airline's website be $\mu + mu_n$, where μ can be learned at an aggregator but to learn additional airline-specific details mu_n, the consumer has to visit the airline. In other words, if a consumer visits an aggregator she learns μ specific to an airline; if she then visits that airline's website she learns the additional mu_n. If she visits only the airline without ever going to the aggregator she learns $\mu + mu_n$ at the airline's website.

Consider a path where a consumer visits the aggregator and one airline before making her purchase. Her total utility is $N\mu - c + mu_1 - c_1$ where c and c_1 are the costs associated with visiting the aggregator and the airline's website respectively. We now delineate cases where in the absence of an aggregator, airline visits stay the same, increase or decrease:

1. Airline visits stay the same. This would occur when the information gain from the second airline adds little value to the consumer: the utility from visiting the first airline is positive, but the utility from visiting the second airline is negative:

$$\mu + mu_1 - c_1 > 0 \text{ and } \mu + mu_2 - c_2 < 0$$

2. Airline visits increase. Under this scenario, visiting an additional airline is informative to the consumer:

$$(\mu + mu_1 - c_1) + (\mu + mu_2 - c_2) > 0 \text{ and } \mu + mu_1 - c_1 > 0$$

3. Airline visits decrease. This occurs when the value of information in any given airline's website is too low to justify the cost of visiting the website. However, had the aggregator been present the consumer would have incurred the cost becuase she would have learnt about multiple airlines (at a lower cost) which is valuable:

$$\mu + mu_1 - c_1 < 0$$

Aggregator visits on the other hand can either decrease or stay the same:

1. Aggregator visits stay the same. This would occur if the absence of the airline does not change the information value drastically, i.e., it is still worth learning $(N-1)\,\mu$(since one airline is absent, the information available is $(N-1)\,\mu$ and not $N\mu$.

2. Aggregator visits decrease. When the aggregator is not comprehensive, the information available at the aggregator is lower and the utility gain does not justify the cost:

$$(N-1)\,\mu - c + mu_1 - c_1 < 0 \text{ and } \mu + mu_1 - c_1 > 0$$

This simple model illustrates that the absence of an aggregator can impact airlines in very different ways depending on the range of parameter values.

B Analysis with un-pooled control periods

The diff-in-diff regression we estimate for American Airlines is:

$$y_{ijt} = \kappa^{\text{dispute,all}} + \kappa^{\text{dispute,F}} \mathrm{I}\left(\text{Focal}_t\right) + \boldsymbol{\kappa}^{\text{ctrl_dispute,all}} \mathrm{I}\left(\text{Ctrl}_t\right) + \boldsymbol{\kappa}^{\text{ctrl_dispute,F}} \mathrm{I}\left(\text{Ctrl}_t\right) \mathrm{I}\left(\text{Focal}_t\right) + \alpha_t + \alpha_i \tag{6}$$

where $\kappa^{\text{dispute,all}}$ captures the average estimate of all other airline websites visited during the dispute period (relative to the estimate in the non-dispute period),
$\kappa^{\text{dispute,F}}$ is the additional change specific to American's website,
$\mathrm{I}\left(\text{Focal}_t\right)$ is an indicator which equals 1 if airline j is the focal airline, i.e. American,
$\boldsymbol{\kappa}^{\text{ctrl_dispute,all}} = \begin{matrix} \kappa^{\text{ctrl_disp_before}} \\ \kappa^{\text{ctrl_disp_after}} \end{matrix}$ is the vector of coefficients that captures the average estimate in the control period - preceding and succeeding years are used as controls,
$\boldsymbol{\kappa}^{\text{ctrl_dispute,F}}$ is the additional change to this vector, specific to American's website,
α_i and α_t are individual- and time- fixed-effects.

Table 10: Searches on impacted domains: Diff-in-diff estimates

		aa.com		orbitz.com		expedia.com	
		coeff	t-stat	coeff	t-stat	coeff	t-stat
prop_pages	$\kappa^{\text{before,F}}$	-0.0048	-1.86	-0.0098	-3.79	-0.0398	-9.58
	$\kappa^{\text{after,F}}$	0.0059	2.21	-0.0069	-2.65	0.0123	3.04
prop_durn	$\kappa^{\text{before,F}}$	0.0058	2.36	-0.0103	-3.94	-0.0167	-4.17
	$\kappa^{\text{after,F}}$	0.0038	1.44	-0.0028	-1.07	-0.0037	-0.92
prop_sites	$\kappa^{\text{before,F}}$	0.0030	1.28	-0.0112	-4.58	-0.0269	-6.81
	$\kappa^{\text{after,F}}$	0.0047	1.90	-0.0037	-1.50	0.0113	2.94
N		4,389,417					
N obs		9 82,886					
Cluster		machine_id					