

Airline Pricing, Price Dispersion, and Ticket Characteristics On and Off the Internet[†]

By ANIRBAN SENGUPTA AND STEVEN N. WIGGINS*

This paper uses transaction data to investigate the effects of Internet purchase on airline fares. Our data include ticket characteristics, restrictions, flight load factors, and dates and channel of purchase. Controlling for ticket and flight characteristics, online purchasers pay about 11 percent less than offline purchasers, which seems rooted in more efficient shopping. The results do not support a spillover in terms of reduced fares or dispersion from greater Internet shopping. The paper also uses the data to reevaluate the relationship between market concentration and fares, but fails to identify any statistically significant, robust relationship. (JEL D83, L11, L86, L93)

This paper uses a unique, individual transaction level dataset to contribute to our understanding of the impact of Internet purchases on airline ticket prices. The analysis considers both the direct effect of Internet purchase on prices paid, and the indirect effect of increased route shares of total Internet purchases on both the level and dispersion of prices more generally. The analysis also extends previous work by Borenstein (1989) and Borenstein and Rose (1994) regarding how market concentration and hubbing affect the level and dispersion of airline prices. Our data enables us to investigate these issues while controlling for ticket characteristics and restrictions, time of purchase, and estimated load factors.

The sale of products on the Internet has dramatically risen over the past decade. Internet sales compete directly with traditional outlets for the sale of numerous products. Consumers purchase books and CDs online, as well as electronic products through shopbots, and a host of travel and entertainment services. Internet penetration is particularly substantial in airlines where for particular city-pairs (routes) the Internet share exceeds 50 percent of transactions. The Internet theoretically reduces search costs and enables buyers to identify low-price sellers, enhancing competition and reducing prices and price dispersion.

As noted by many, however, the empirical results supporting these theoretical predictions have not generally been strong or consistent, although data limitations have played an important role in limiting the analysis of various markets. Early studies

* Sengupta: Amazon Corporate LLC., 500 Boren Ave. N, Seattle, WA 98106 (e-mail: anirbans@amazon.com); Wiggins: Department of Economics, Texas A&M University, College Station, TX 77843-4228 (e-mail: swiggins@neo.tamu.edu).

[†] To comment on this article in the online discussion forum, or to view additional materials including the author disclosure statement(s), visit the article page at <http://dx.doi.org/10.1257/pol.6.1.272>.

found mixed evidence regarding prices and price dispersion.¹ Morton, Zettlemeyer, and Silva-Rissio (2001) use micro-level transaction data and they find that consumers who searched for automobiles online paid about 2 percent less than consumers who did not. Brown and Goolsbee (2002) investigate the effects of Internet search on the prices of term life insurance. They compare prices in geographic areas and for demographic groups where there are high levels of Internet search for life insurance with areas and groups where Internet search is low. They find that the Internet has led to a substantial reduction in both the level and dispersion of term life insurance prices. Baye and Morgan (2001) use posted prices and investigate the role of shopbots and other middlemen, finding high levels of price dispersion on the Internet. Similarly, Scholten and Smith (2002) examine cross-market variation and find greater price dispersion for Internet purchases for a wide variety of goods as compared to prices drawn from heterogeneous local geographical settings.²

These studies provide important information regarding varied aspects of Internet pricing, but due to data limitations they do not provide a comprehensive, industry-wide analysis of the impact of the Internet on Internet prices, offline prices, and price dispersion for any industry. Instead these analyses are each drawn from different industries, and each investigates a particular aspect of how the Internet has affected prices. Further, the data used often consist of posted prices rather than transaction prices.³ These analyses also at times compare “national” Internet prices with offline prices drawn from potentially heterogeneous local markets.

This paper provides, to our knowledge, the most comprehensive analysis of the impact of Internet purchases on pricing for a single industry, using a unique dataset from the airline industry.⁴ The central goal of the paper is to investigate the effects of Internet sales on prices paid for airline tickets.

The dataset consists of actual individual transactions purchased through a large Computer Reservation System (CRS).⁵ These actual transactions data offer substantial advantages over data collected from posted prices because they include actual purchases and reflect substantial heterogeneities in units purchased for tickets with different characteristics and prices. The data include whether a ticket was purchased online or offline and numerous observable ticket characteristics and restrictions, such as refundability, advance purchase requirements, and travel restrictions. The analysis also controls for estimated load factors at the time of purchase, network

¹Lee (1998) focused on cars and Bailey (1998) focused on books, CDs and software, finding higher online prices and equal price dispersion across these channels. Brynjolfsson and Smith (2000) focus just on books and CDs and find lower prices but find that internet price dispersion is quite high.

²Chen (2006) finds that average prices on the Internet for airline tickets differ by about 3 percent among the online travel agents, consistent with the findings of Clemons, Hann, and Hitt (2002).

³Note that much of the existing literature on Internet pricing relies on posted prices as opposed to transaction data (Pan, Ratchford, and Shankar 2001, 2004).

⁴Note that Verlinda (2004) and Orlov (2004) have analyzed the effects of the Internet on airline pricing. Their analyses match data from DB1B with data measuring the average share of people using the Internet to purchase airline tickets in a given geographical area. These analyses are limited because they do not use observed Internet transactions. More importantly, their analyses do not consider route-level variation in Internet share from a given origin, and they also do not have available other ticket characteristics of the type considered here. The analysis in Section C, Table 3 shows that Internet savings will be overestimated if one does not control for ticket characteristics.

⁵These data were matched with ticket characteristics using a procedure described in more detail below.

peak times, market structure, and route characteristics.⁶ The results show that ticket characteristics account for the large majority of observed variation in ticket prices; a regression of individual ticket prices on these characteristics, route, and carrier dummies yields an R^2 of approximately 0.7.

Our analysis measures the direct impact of Internet purchase for those customers who buy on the Internet, controlling for ticket, flight, and market characteristics. We also investigate how increased shares of Internet purchases influence both average prices online and offline and examine the impact of Internet purchases on the dispersion of prices for online, offline, and overall fares. Finally, the analysis provides a more detailed investigation of the relationship between market structure, price levels, and dispersion, while controlling for ticket and flight characteristics.

The results show that when controlling for ticket characteristics, tickets purchased on the Internet cost about 11 percent less than tickets purchased offline. The Internet also enables consumers to find the lowest-priced package of characteristics. If one does not control for ticket characteristics, estimated Internet savings rise to as much as 30 percent or more.⁷ Our data and analysis, however, do not support the presence of a spillover effect of high levels of Internet purchases on prices more generally. Hence our results do not confirm a general market effect as found by Brown and Goolsbee (2002).

Figure 1 presents some illustrative data. The figure shows mean fares measured for various days prior to departure for online versus offline sales for the four largest city-pair/carrier combinations in our data. These data indicate that Internet purchasers generally pay substantially lower fares than offline purchasers, regardless of the number of days in advance when the purchase occurs.

Table 1 provides a further look at these data, investigating the general relationship between online and offline fares for tickets containing various restrictions. The table shows average online and offline fares for the largest carrier on our nine largest city-pairs for tickets with various characteristics. The characteristics considered include class of service, refundability, advance purchase requirements, Saturday stayover, and other ticketing restrictions. The results show that Internet fares are consistently significantly lower in these categories than offline fares. The analysis below provides an in-depth investigation of these differences.

The paper also contributes significantly to our understanding of the airline industry by providing a more complete analysis of how market structure affects the level and dispersion of airline prices. The analysis builds on the work of Borenstein (1989) and Borenstein and Rose (1994), who analyze the effects of market structure on price levels and price dispersion for airlines. Borenstein (1989) shows that hub dominance and high route shares lead to higher average prices. Borenstein and Rose

⁶The data were provided by one of the major computer reservation systems (CRS) and include individual transaction data for the fourth quarter of 2004. The data indicate city-pair, carrier, flight number, dates of purchase, departure and return, booking class, and whether the ticket was purchased online or offline. These data were then matched with contemporaneous data from another CRS that included full details on individual ticket characteristics such as refundability, booking class, advance purchase requirements, travel and stay restrictions, and Saturday night stayover.

⁷Note that this result is driven, in part, by the adverse selection problem where customers who want lower fares and are willing to accept various restrictions are more prone to shop on the Internet. See Section C for further discussion.

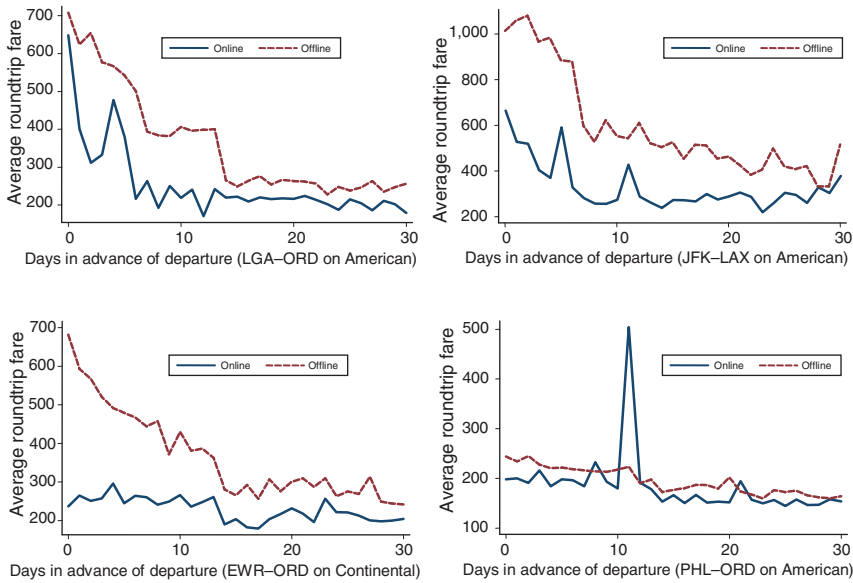


FIGURE 1. COMPARISON BETWEEN ONLINE AND OFFLINE DAILY AVERAGE FARES

(1994) show that price dispersion decreases as routes become more monopolistic. Gerardi and Shapiro (2009) however find a negative effect of competition on price dispersion, attributing it to the heterogeneous elasticity of demand across routes.

Due to data limitations these studies do not control for important ticket characteristics or flight load factors, which have a significant effect on both price levels and price dispersion. Similar limitations apply to most existing studies of airline pricing.⁸ Our analysis investigates the effect of market structure variables while controlling for Internet purchases and ticket characteristics, purchase days in advance, estimated flight level load factors, and other peaking variables. Our results do not indicate a statistically significant relationship between market structure variables and fares.

I. Airline Pricing

Airlines offer a wide variety of different fares for travel on the same flight and the same day. The available evidence indicates that airlines offer tickets for sale in a conceptual series of “bins” or “buckets,” where a bucket is defined by a series of ticket characteristics including class of travel, refundability, advance purchase requirements, and travel and stay restrictions such as minimum and maximum stays and/or Saturday stayover.⁹ The received wisdom is that airlines limit the quantity

⁸Note that Stavins (2001) overcomes a part of this limitation by using some ticket characteristics in explaining the relationship between market concentration and prices in the airlines market. Stavins (2001), however, uses certain posted prices and only a subset of ticket characteristics. Using these posted prices, Stavins (2001) shows that price discrimination decreases as market concentration rises while increases in the route share of a carrier allows it to price discriminate more among consumers.

⁹See Smith et al. (2001).

TABLE 1—COMPARISON OF ONLINE AND OFFLINE MEAN FARES CONDITIONAL ON TICKET CHARACTERISTICS

	Various routes			
	Atlanta–LaGuardia DL		Boston–Reagan National US	
	Offline	Online	Offline	Online
3-day advance	351.7	240.7	691.4	682.0
7-day advance	264.8	233.5	384.1	242.3
14-day advance	191.3	185.2	300.2	239.9
Refundable	1,365.1	426.0	435.4	394.5
Nonrefundable	276.9	216.1	275.3	211.9
Saturday stayover	240.4	202.4	295.4	224.7
First class	1,360.0	438.0	736.1	304.5
Business class	1,105.4	—	680.7	652.2
Minimum stay restriction	181.5	175.0	235.1	207.7
Maximum stay restriction	181.5	175.0	274.6	211.6
	Los Angeles–Newark CO		Boston–Chicago O’Hare AA	
	Offline	Online	Offline	Online
	3-day advance	492.1	370.0	396.5
7-day advance	307.2	283.1	269.3	198.2
14-day advance	293.8	292.2	292.1	202.8
Refundable	1,869.4	—	1,029.9	—
Nonrefundable	490.7	331.1	384.2	230.2
Saturday stayover	482.9	298.8	269.1	205.0
First class	2,128.4	—	737.7	—
Business class	782.7	386.3	916.8	—
Minimum stay restriction	360.9	309.9	277.0	204.3
Maximum stay restriction	440.8	412.3	281.1	204.6

(Continued)

of low price tickets by limiting the number of tickets in low price buckets. For example, certain combinations of characteristics may only be used during certain days of the week (e.g., TWF), and certain tickets may only be available for round trips. Certain fares may not be available on certain flights for a period of time, and then later once again become available. High priced tickets are sometimes sold far in advance of departure, and deeply discounted tickets in certain bins may be available on the day of departure.

Airlines can alter the prices passengers ultimately pay for tickets both by changing the price of tickets within a given bucket and by rationing the number of tickets in that bucket.¹⁰ The general analysis of this issue is beyond the scope of this paper.

For the present analysis, it is simply important to note that airlines price using these ticket characteristics, placing tickets in particular bins that feature different prices. The analysis below shows that variation in ticket prices is driven largely by variation in ticket characteristics in that a simple regression of price on ticket characteristics, carrier, and route dummies explains roughly 70 percent of the variation in ticket prices.

¹⁰See Smith et al. (2001).

TABLE 1—COMPARISON OF ONLINE AND OFFLINE MEAN FARES CONDITIONAL ON TICKET CHARACTERISTICS
(Continued)

	Los Angeles–JFK AA		Chicago O’Hare–Dallas AA	
	Offline	Online	Offline	Online
3-day advance	1,549.5	553.8	587.1	1,338.0
7-day advance	336.7	264.4	523.9	306.8
14-day advance	361.5	283.1	239.4	229.6
Refundable	1,866.6	1,523.3	1,281.1	1,224.0
Nonrefundable	444.4	302.5	561.7	308.0
Saturday stayover	648.7	283.6	325.6	216.5
First class	2,341.0	—	1,231.3	—
Business class	1,448.8	—	849.5	1,086.2
Minimum stay restriction	349.7	274.1	243.8	220.9
Maximum stay restriction	421.3	347.3	235.5	210.6
	Chicago O’Hare–Newark CO		Philadelphia–Chicago O’Hare AA	
	Offline	Online	Offline	Online
3-day advance	699.8	364.0	268.4	178.0
7-day advance	376.0	240.1	213.9	186.3
14-day advance	262.3	208.5	170.4	155.2
Refundable	1,029.1	223.0	248.3	227.4
Nonrefundable	383.2	235.1	199.8	158.7
Saturday stayover	257.6	217.0	208.5	181.8
First class	1,655.1	—	620.6	628.0
Business class	971.0	332.1	715.2	—
Minimum stay restriction	292.5	222.6	224.1	175.7
Maximum stay restriction	287.8	230.7	248.7	172.1

(Continued)

Airline customers and travel agents search for airline tickets by attempting to find sets of characteristics the customer is willing to accept at the lowest possible price. The search for low price tickets may take place either online where the customer directly investigates the fares offered by one or more online sites, or it may take place offline where the ultimate customer uses a travel agent.

II. Search Theory, Pricing, and the Internet

The analysis here takes the same theoretical approach to Internet pricing as Brown and Goolsbee (2002). We assume, as is implicit or explicit in much of the literature on the Internet, that the Internet lowers search costs. This assumption is consistent with the evidence presented below.

More formally, the analysis builds on the search model of Stahl (1989). Stahl (1989) assumes that a certain exogenous share of customers is fully informed regarding all prices available in the market, and that another group of customers must pay a search cost for each price quote received. Because customers search sequentially, stores choose prices in the Nash equilibrium from a price distribution rather than using a pure strategy. Searchers with positive search costs stop searching endogenously whenever the price they observe is at or below their endogenously determined reservation price. Fully informed customers have no search costs and search exhaustively, buying from the lowest priced seller.

TABLE 1—COMPARISON OF ONLINE AND OFFLINE MEAN FARES CONDITIONAL ON TICKET CHARACTERISTICS
(Continued)

	LaGuardia–Chicago O’Hare AA	
	Offline	Online
3-day advance	636.1	291.3
7-day advance	355.5	176.5
14-day advance	266.3	228.7
Refundable	1,182.5	—
Nonrefundable	407.3	219.6
Saturday stayover	272.3	203.6
First class	1,248.0	—
Business class	1,127.9	—
Minimum stay restriction	304.1	215.5
Maximum stay restriction	309.0	221.2

This logic generates three well-known results. First, in equilibrium firms draw prices from a distribution, generating price dispersion. Second, as the share of informed customers rises, the price distribution shifts monotonically downward leading to decreases in average prices. Third, the degree of price dispersion is not monotonic with respect to the percentage of informed consumers. Instead, when there are no searchers the price distribution degenerates to the monopoly price. Dispersion then rises as some consumers become informed, eventually reaching a peak, but then collapsing back to zero when all consumers are fully informed and prices converge to the competitive level.

We assume that online customers are better informed than offline customers, but we ultimately test this assumption by determining if they pay lower prices. We use transactions rather than posted prices, which means that we measure the actual price distribution rather than the distribution of posted prices, some of which might net few if any sales.

The methodology also ensures that the data used only includes fares available both online and offline. The data we use incorporates observations where transaction fares were matched to those found through a CRS accessed offline.¹¹ This matching procedure means that the fares considered were available both online and offline. The analysis below investigates this issue by examining how the percentage of Internet purchases affects the level of fares both on and off the Internet.

¹¹ Note that the available information indicates that, in general, the same prices and fare combinations are available online and offline. In the early days of Orbitz, the prices listed there included pricing specials offered “directly” by the airlines, falling outside the travel agent contracts. It also may be possible that some offered prices on particular airline sites are lower than fares offered elsewhere. However, this paper does not include web special fares by virtue of the construction of the dataset. For a detailed discussion on the construction of the dataset, please refer to the online Appendix.

III. Empirical Analysis

A. Data

The data consists of contemporaneous online and offline transaction data provided by a leading CRS for the last quarter of 2004.¹² The data includes all transactions completed through the CRS during that quarter, including transactions at airline sites, travel agents, and numerous online sites. As noted above, the data from the CRS includes airline and flight number, origin and destination, fare, booking class, a fare code, and dates of purchase, departure and return. Overall, these data represent roughly 30 percent of all domestic tickets sold in the United States during the fourth quarter of 2004. These data do not include refundability, advance purchase requirements, and travel and stay restrictions.

To obtain these variables, we electronically matched the data with a dataset from another CRS containing both fares offered and purchased for travel in particular city-pairs organized by departure date, airline, and city-pair. This second dataset includes ticket characteristics not available in the first dataset.¹³ The dataset from this second CRS was, however, incomplete in that certain fares had been deleted from the archive, and so we were only able to match the fares imperfectly.¹⁴ The criterion used was to keep transactions if we were able to match the fares within two percent; for multiple matches within 2 percent we kept the closest.¹⁵ The matching procedure also ensured that tickets matched with respect to carrier and booking class, and that travel and stay restrictions were met by the observed ticket. Since the first dataset contained purchase days in advance, and the second contained advance purchase restrictions, we also required consistency between the days in advance and advance purchase restrictions. The resulting matched dataset contains individual ticket transactions that include ticket characteristics and restrictions, together with carrier, flight information, and dates of purchase, departure, and return.¹⁶

This procedure matched roughly 35 percent of the observations from the first dataset. For both the online and offline transactions, our match rate is somewhat lower for the lowest priced tickets. In our analysis we focus primarily on the matched dataset. Before turning to that dataset, however, it is important to compare the matched

¹²The data and construction of variables are discussed at length in the online Appendix.

¹³We have been informed that fares offered on the various CRSs are normally the same, but that at times a fare will only be offered on some CRSs. This permits the use of departure dates to match the route, carrier, fares, and fare classes in the first dataset with the detailed ticket characteristics found in the second dataset. The details are provided in the online Appendix.

¹⁴The data in the second archive are kept for unknown intervals of time. Individual fares are then deleted in an unknown pattern over time. The online Appendix provides more details on the matching protocol and results.

¹⁵In the first matching step, we only require fares to match within 10 percent. In a later step, we require fares to match more closely. In addition, we match a transaction's date of departure to a seven-day window of departure dates in the posted fare data, and later use the match in which the dates of departure are closest. The online Appendix Table A5 reports results using a 5 percent matching criterion. Those results are qualitatively similar to the results reported below.

¹⁶Since the CRS deregulation in 2004, the airlines are free to provide different fares to any distribution channel including the major CRSs, their own CRS, and website and online travel agencies like Expedia. This necessitated the adoption of the matching rule(s) as discussed in the paper. Please refer to the online Appendix for discussion on the matching procedure.

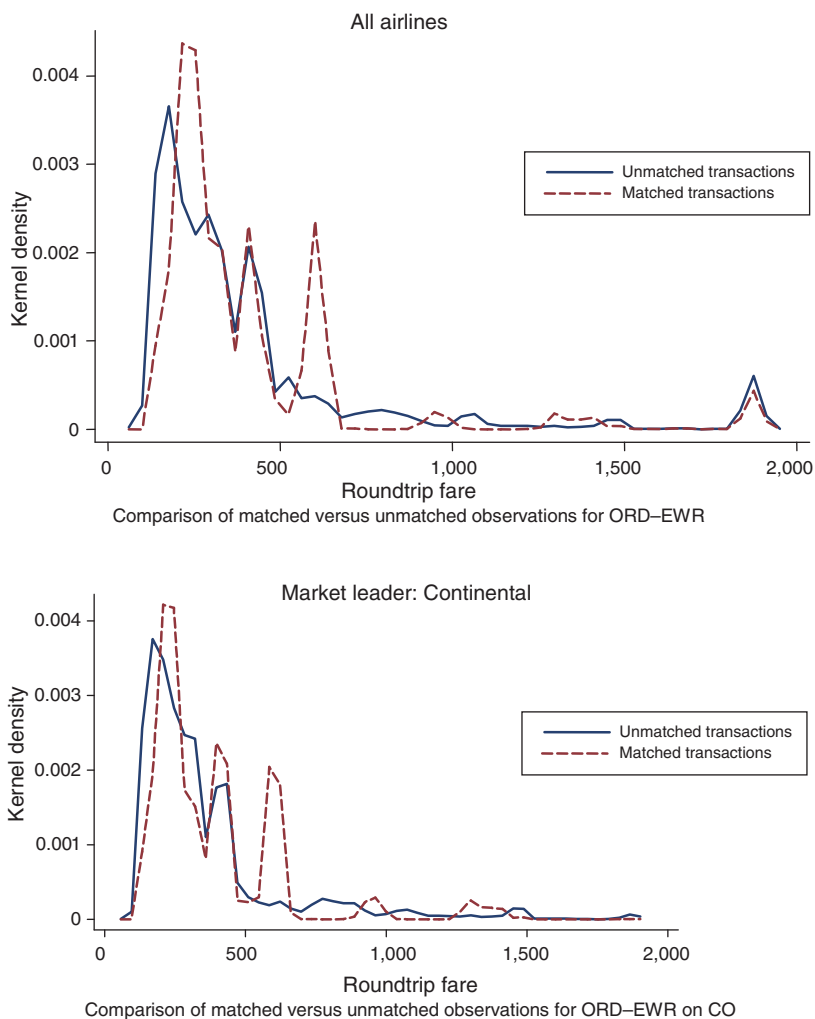


FIGURE 2. COMPARING THE KERNEL DENSITIES OF MATCHED AND UNMATCHED TRANSACTIONS FOR FLIGHTS BETWEEN CHICAGO O'HARE AND NEWARK LIBERTY

and unmatched datasets using information available in both datasets to ensure the representativeness of the matched data.

Kernel densities of fares for the matched and unmatched observations for different city-pairs are shown in Figures 2–4. For example, Figure 2 shows the data for Chicago to Newark; the top panel compares matched and unmatched data for all airlines, and the bottom panel for Continental, the market leader. Figures 3–4 show similar kernel densities for two other large city-pairs. These kernel densities show an underrepresentation of the very lowest fares for both all airlines and for the largest airlines on a route. Our analysis of online and offline fares, however, does not appear to be affected because we only consider matches for online and offline fares, and the underrepresentation of matches is comparable in both datasets. More specifically, Figure 5 compares the kernel densities for matched and unmatched transactions broken down by online versus offline transactions.

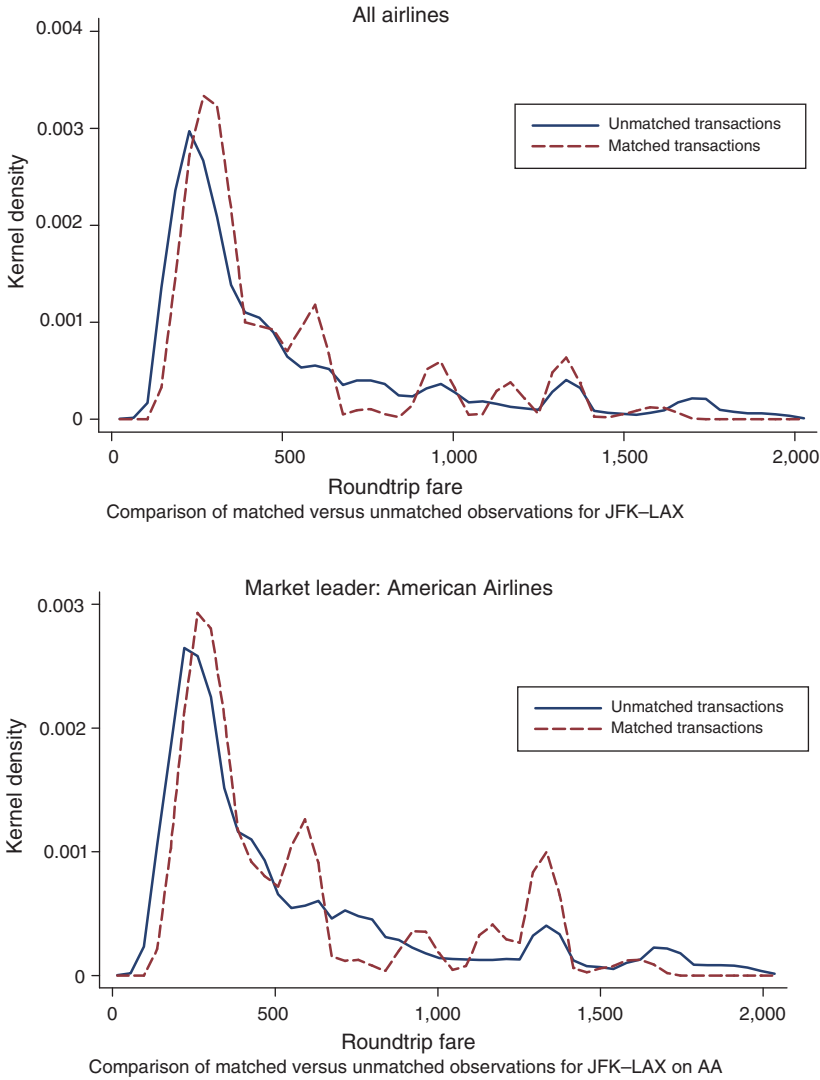


FIGURE 3. COMPARING THE KERNEL DENSITIES OF MATCHED AND UNMATCHED TRANSACTIONS FOR FLIGHTS BETWEEN JFK, NEW YORK, AND LOS ANGELES, CALIFORNIA

Both online and offline transactions have fewer matches for very low fares, but there do not appear to be significant differences in the match rate for online versus offline fares.

The small difference in online versus offline matches is also illustrated by examining the match rate in the left-hand tail of the distributions in Figure 5. This tail consists of price observations below \$221, the price at which the matched versus unmatched kernel densities cross. Below \$221, the match rate for offline tickets is about 22.6 percent while the match rate for online tickets is about 19.1 percent. These match rates are very comparable, and the differences point toward a small oversampling of offline fares. These effects point toward a comparative underrepresentation of low fares on the Internet, although these effects will be

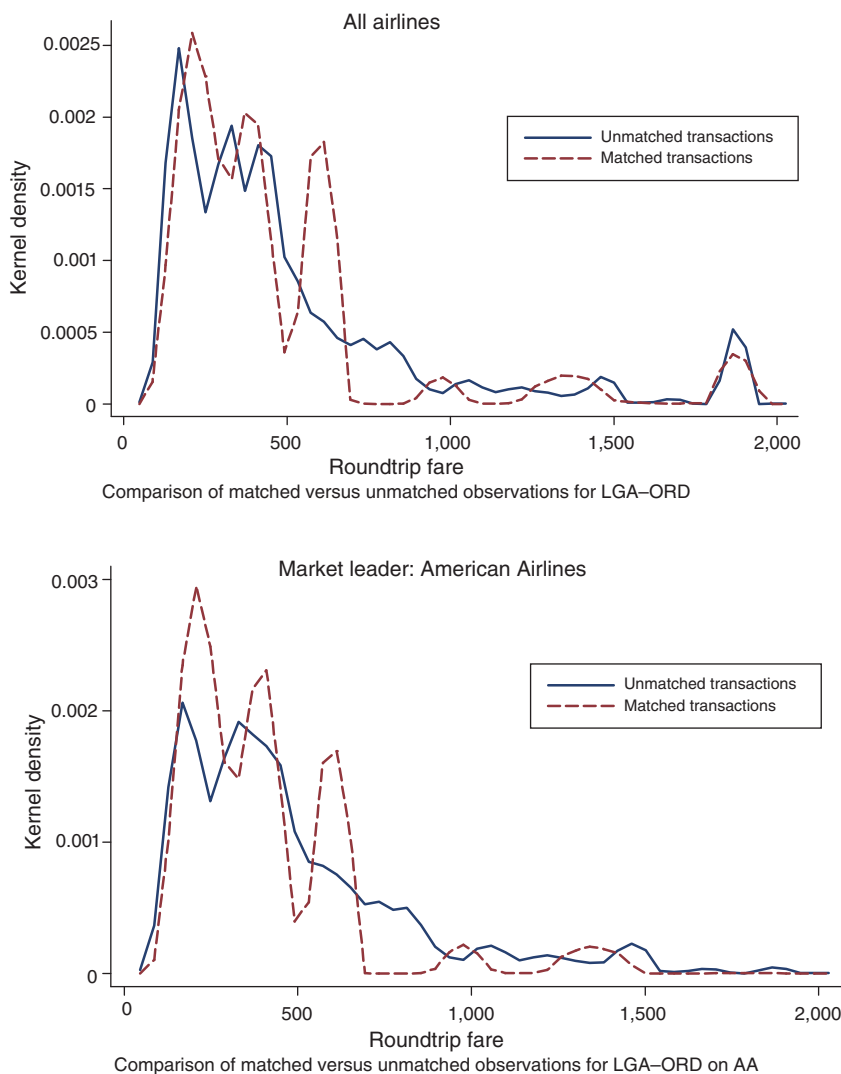


FIGURE 4. COMPARING THE KERNEL DENSITIES OF MATCHED AND UNMATCHED TRANSACTIONS FOR FLIGHTS BETWEEN LA GUARDIA, NEW YORK, AND CHICAGO O’HARE

minimized in the regressions because we control for ticket characteristics.¹⁷ We return to this issue below.

This study uses data for 150 large US domestic city-pairs, including a mix of both business and tourist routes, and routes with varying groups of customers. The routes we use are the 150 routes with the largest number of transactions in the transaction database from the first CRS. It should be noted that because we have chosen large routes where there was more direct service, the resulting sample is disproportionately

¹⁷The online data Appendix provides comparisons of summary statistics for the matched and unmatched data. This table shows that the matched data has a mean fare of \$376 while the unmatched data has a mean fare of \$341. In all other respects, however, the matched and unmatched data are quite similar.

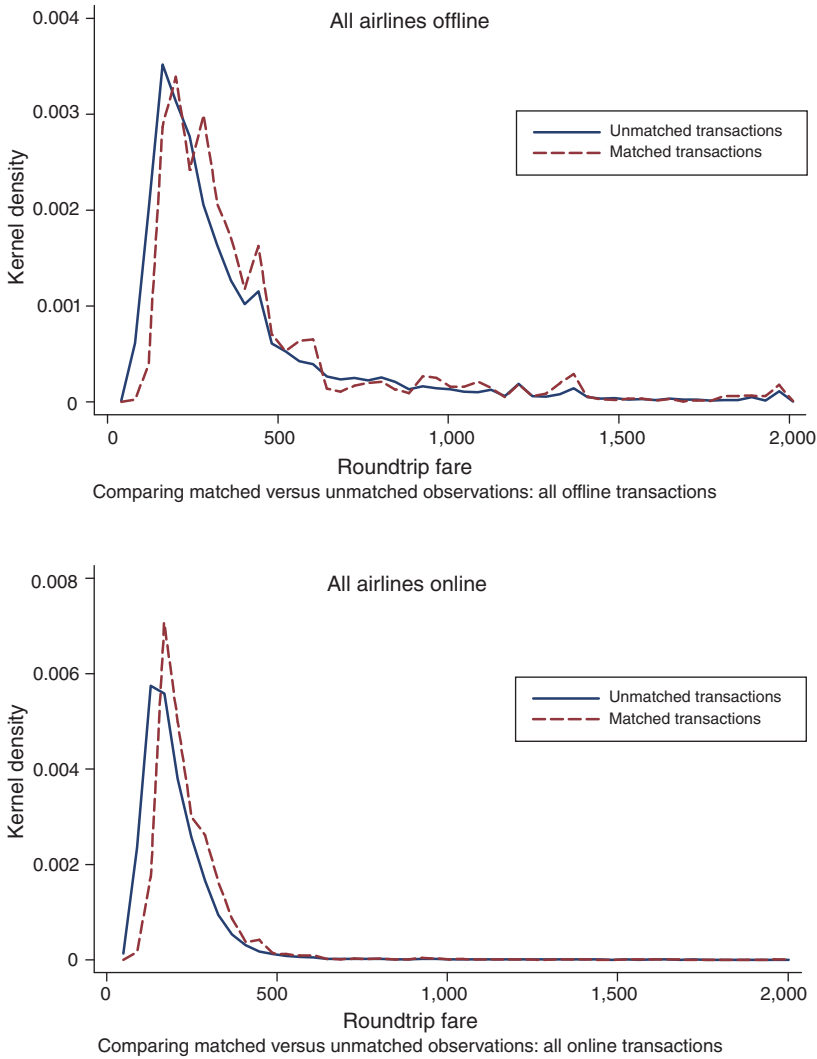


FIGURE 5. COMPARING THE KERNEL DENSITIES OF MATCHED AND UNMATCHED TRANSACTIONS FOR OVERALL ONLINE AND OFFLINE TRANSACTIONS

from hubs and slot-controlled airports. While we control for these variables, the results should be interpreted recognizing that the sample is drawn from these large routes. A complete list of routes is contained in the online Appendix, Table A2. We define a route as a city-pair regardless of direction. We restrict our analysis to direct (nonstop) itineraries only.¹⁸ The prices used are for roundtrip fares, doubling the fares for one-way tickets to obtain comparability. Our analysis also excludes first and business class tickets, but does include full coach fare class tickets. This study

¹⁸The results do not change if nondirect itineraries are included in the analysis. Given that approximately 98 percent of the matched itineraries in our data involved direct itineraries, we excluded the small number of nondirect itineraries.

includes tickets for flights operated by American, Continental, Delta, Northwest, US Airways, United, Frontier, Air Tran, Spirit, Alaska, America West, Sun Country, Hawaiian, and American Trans Air.¹⁹

Observations consist of individual tickets, their fares and characteristics, and other data described above and more fully in the online Appendix. The data also includes control variables for carrier and route effects, route market shares, HHI, hubs, and other standard variables measuring tourism, income, and population. We also include variables indicating the presence of discount carriers on routes, and a separate variable for Southwest. The complete set of variables is discussed below.

B. Model Specification

$$(1) \quad \ln(\text{Price}_{ijkt}) = \alpha + \beta_1(\mathbf{R}_{ijkt}) + \beta_2(\mathbf{Travel}_{ijkt}) + \beta_3(\mathbf{Online}_{ijkt}) \\ + \beta_4(\mathbf{Market\ Structure}_j) + \beta_5(\mathbf{Market}_{jk}) \\ + \beta_6(\mathbf{Carrier}_{ijkt}) + \varepsilon_{ijkt}.$$

\mathbf{R}_{ijkt} represents a vector of ticket characteristics. These ticket characteristics include refundability, advance purchase requirements, Saturday night stayover, round trip travel, travel restrictions (for example, if the ticket is valid only if you travel on a MTF), and stay restrictions (minimum or maximum stay requirement).

\mathbf{Travel}_{ijkt} represents the vector of advance purchases and travel times associated with individual transactions. One component of this vector is the number of days in advance the ticket was purchased. Ticket prices are also affected by the time of day when travel occurs. To control for the time of travel, we first identify whether the departure or return takes place during a peak time of day, where peak time corresponds to flights departing on weekdays between 7:00 AM and 10:00 AM or between 3:00 PM and 7:00 PM.²⁰ These periods of peak travel generally have high load factors and there may also be network cost/capacity effects resulting in higher prices during these periods. Since departure and return occurs at different times of the day, we construct three categories: both departure and return occur at peak times; both departure and return occur at off-peak hours of the day; and either departure or return occur at peak times but not both. For similar reasons, certain days of the week are busier than others so we include a full set of dummy variables representing each day of the week, with separate variables included for the departure day of the week and the return day of the week. Sunday is treated as the base group. Finally, the vector also includes flight level load factor, which equals the deviation of the flight's load factor at the time of purchase from the average load factor at the carrier-route level corresponding to that particular number of days prior to departure (see below).

¹⁹We can identify routes served by Southwest, but we do not have data regarding Southwest's ticket characteristics because they are not included in one of the datasets.

²⁰We also used the time window of 8:00 to 10:00 AM as a peak time window. The results were qualitatively unaffected.

$Online_{ijkt}$ is an indicator variable representing online transactions. Furthermore, **Market Structure_j** includes a vector of route-specific variables that have been widely used to study airline pricing (see, e.g., Borenstein 1989). The vector includes the Herfindahl index, which is measured using the passenger shares in the fourth quarter of 2004 from DB1B. We also include a low-cost carrier dummy variable indicating the presence of a low-cost carrier and a separate Southwest dummy variable indicating Southwest's presence. Similar to earlier studies, we also include the logs of statutory distance, average population of the cities in the city-pair, and average per capita income.

The **Market Structure_j** vector also contains a tourist variable, which relies on the business share index developed by Borenstein (2010). He reports for each airport code the share of passengers traveling for business both inbound and outbound. We use these data to calculate business shares by direction. Consider, for example, Chicago and Atlanta. The outbound business share from Chicago (ORD) is 0.41 while the inbound business share is 0.60. The outbound share for Atlanta (ATL) is 0.44 while the inbound share is 0.54. For our analysis, an itinerary with ORD as origin airport and ATL as the destination airport will be assigned a business share of 0.475 (averaging the 0.41 outbound share from Chicago with the 0.54 inbound share into Atlanta), while an itinerary that originates at ATL with ORD as its destination will be assigned a business share of 0.52.²¹ This method accounts for differences in tourism by direction of travel, which could affect "shopping" behavior in the purchase of tickets. The tourist share equals one minus the business share.

The vector **Market_{jk}** represents two variables at the carrier-route level. The first variable is the market share of a carrier on a specific route. The other variable is a hub dummy which takes a value of one if the carrier has a hub at either of the endpoints on a route. Finally, **Carrier_{ijkt}** is a vector representing carrier fixed effects.

We employ both standard OLS and Instrumental Variable (IV) regression techniques to estimate equation (1). The IV regression is required in part to address potential endogeneity of market share and the Herfindahl index. The carrier's market share is endogenous because it is influenced by prices and is instrumented using the carrier's enplanement share at the two endpoints on the route.²² To the extent that market share is endogenous; the route Herfindahl index is also endogenous since the square of market share is one component of the route Herfindahl. The instrument for the Herfindahl index is the square of the fitted value for market share (from its first stage regression) plus the "rescaled" sum of the squares of the remaining carriers' shares on the route.²³ We also discuss below the potential endogeneity of the share of Internet purchases.

To investigate the effects of increased Internet usage on fares more generally we include an Internet share variable, representing the ratio of online transactions to total transactions on a route. This measure is conceptually similar to that used by Brown and Goolsbee (2002) except that they *estimate* people who have searched for

²¹ For robustness, we also replicated the analyses using alternative tourist definitions. These included separate controls for the outbound and inbound measures of tourist share at the origin and destination airports respectively.

²² The instrument parallels that of Borenstein (1989) and Borenstein and Rose (1994).

²³ See Borenstein (1989) and Borenstein and Rose (1994) for further discussion.

prices on the Internet, while we use a direct measurement of the *share of transactions actually completed* on the Internet.²⁴ We also include an interaction variable to measure whether the savings from direct purchase on the Internet vary across markets with the share of tickets bought online.

Including the share of Internet purchases on a given route as an explanatory variable also creates potential endogeneity issues. Consumer's purchase decisions, including whether to shop online or offline, can be influenced by prices and price distributions. To address this potential simultaneity bias, we use two instruments. The first instrument is the percentage of transactions three or more weeks prior to departure on a route as an instrument for the online share on a route. Potential online shoppers are likely to be price sensitive customers who are also more likely to buy their tickets well in advance. Further, the share of transactions three or more weeks in advance is highly correlated with the share of online transactions on a route but is unlikely to be correlated with prices. The second instrument we use is the share of Internet purchases on other routes out of the endpoint airports.²⁵ This measure is constructed by calculating the share of online purchases originating at the endpoint airports of a route excluding the route in question.

C. Data Overview

Table 2 presents the descriptive statistics of the variables used. The final dataset consists of 453,347 observations from 150 major routes in the US. Our data contain two measures of Internet usage. The first Internet variable, *online*, is a dummy variable that takes a value of one if the transaction was completed over the Internet. The second Internet variable, *Internet share*, is the ratio of all tickets purchased on the Internet to all transactions on a given route. The share of Internet purchases varies significantly across routes, varying from a meager 2 percent to as high as 59 percent. As expected, a higher share of Internet purchases is found on leisure routes such as those to and from Orlando and Las Vegas.

The deviation of the load factor calculated at the time of purchase is calculated as follows. The average load factor is computed as the average of load factor for a carrier and route corresponding to a particular number of days in advance prior to departure for a given airline, city-pair combination. This average measures the typical rate at which flights sell out on a particular carrier/city-pair combination. The deviation is calculated as the difference between the actual load factor at purchase for a particular ticket and the average load factor for that carrier-route and that number of days prior to departure. Note that the average load factor is defined as the average load factor for individual segments of an itinerary. Load factors were calculated using tickets sold for both matched and unmatched observations in our transactions data, and using through tickets where the given segment only consists of a portion of the trip. We also include open-jaws and circular trips in calculating load factors. Since data for all transactions on a specific flight are not available, our measured load factors account for roughly 30 percent of transactions, corresponding

²⁴ Recall from above that in our matched data all fares are available both online and offline.

²⁵ We also report the primary results using the two instruments separately in the online Appendix.

TABLE 2—SUMMARY STATISTICS

Variable description	Mean	Standard deviation	Minimum	Maximum
<i>Market structure variables</i>				
Market share	0.539	0.252	0	1
HHI	0.532	0.189	0.189	1
Hub	0.732	0.443	0	1
Slot restricted airport	0.254	0.435	0	1
<i>Internet variables</i>				
Online	0.133	0.340	0	1
Internet share	0.186	0.098	0.024	0.587
<i>Ticket characteristics</i>				
No advance purchase requirement	0.400	0.490	0	1
1-day advance purchase requirement	0.017	0.131	0	1
3-day advance purchase requirement	0.085	0.279	0	1
5-day advance purchase requirement	0.001	0.031	0	1
7-day advance purchase requirement	0.201	0.401	0	1
10-day advance purchase requirement	0.024	0.153	0	1
14-day advance purchase requirement	0.244	0.429	0	1
21-day advance purchase requirement	0.024	0.153	0	1
30-day advance purchase requirement	0.003	0.059	0	1
Roundtrip fare	376.498	294.487	61.995	3860.002
Nonrefundable	0.816	0.387	0	1
Days prior to departure ticket purchased	15.866	20.722	0	202.000
Saturday stayover	0.195	0.397	0	1
Travel restriction	0.441	0.497	0	1
Minimum stay requirement	0.241	0.428	0	1
Maximum stay requirement	0.211	0.408	0	1
Full coach fare class	0.089	0.285	0	1
<i>Remaining ticket characteristics</i>				
Roundtrip itinerary	0.766	0.423	0	1
Deviation in load factor	0.001	0.063	-0.322	0.737
<i>Time of day</i>				
Departure and return at peak time	0.055	0.229	0	1
Departure and return during off-peak time	0.533	0.499	0	1
Either departure or return at peak time, but not both	0.412	0.492	0	1
<i>Other route specific characteristics</i>				
Low cost carrier on route	0.467	0.499	0	1
Southwest airlines	0.064	0.246	0	1
Distance (log)	6.642	0.686	5.220	7.902
Tourist share	0.561	0.069	0.406	0.761
Population (log)	14.158	0.855	12.359	15.603
Per capita income (log)	10.502	0.095	10.078	10.715
<i>Departure day of week</i>				
Sunday	0.103	0.304	0	1
Monday	0.210	0.407	0	1
Tuesday	0.188	0.391	0	1
Wednesday	0.165	0.371	0	1
Thursday	0.149	0.356	0	1
Friday	0.126	0.332	0	1
Saturday	0.059	0.236	0	1
<i>Return day of week</i>				
Sunday	0.337	0.473	0	1
Monday	0.079	0.269	0	1
Tuesday	0.105	0.306	0	1
Wednesday	0.129	0.335	0	1
Thursday	0.145	0.352	0	1
Friday	0.156	0.363	0	1
Saturday	0.049	0.217	0	1

to the share of transactions contained in our transactions data base. These load factors will be less than one.

To address this issue, we used data from the Bureau of Transportation Statistics (BTS). While we do not know the number of tickets sold through other CRSs on a flight, we can use available data to construct an unbiased estimate of these unobserved tickets at the carrier-route level. The BTS reports monthly data on the total number of tickets sold for each route by airline. Using these data we can calculate the exact share of total tickets that we observe in our CRS data for a given carrier, route and month. We then scale up the observed transactions on a particular flight by the inverse of that observed share to obtain an unbiased estimate of realized load factor for a given flight.

Table 3 provides crude, preliminary estimates of the savings from Internet purchase. These estimates do not control for ticket characteristics and so the results overstate Internet savings. In particular, if the Internet were not available, customers purchasing online would likely have found cheap fares offline by purchasing highly restricted, nonrefundable tickets. These preliminary estimates do not control for this adverse selection problem because they do not control for ticket characteristics. Still, such an approach does measure the overall savings, providing an upper bound on the gains consumers obtain from the Internet. The results also show the importance of controlling for ticket characteristics when they are compared to the more complete results below.

Column 1 of Table 3 presents a simple regression of ticket prices and the online dummy variable along with departure and return day of the week, carrier, and route fixed effects. We run a similar specification without the departure and return day of the week in column 2 of Table 3. Both specifications suggest that online consumers on average save roughly 27 percent on ticket prices as compared to offline consumers.²⁶ These results indicate that online consumers looking for low price tickets and willing to accept restrictions save substantial amounts compared to customers buying through offline outlets.

Table 3 also provides additional analysis to investigate whether our matching protocol has influenced estimated savings on the Internet. Columns 3 and 4 of Table 3 present regressions estimating Internet savings using both matched and unmatched observations so that these results can be compared to the results of columns 1 and 2, which include only matched observations. Inspection of the results shows that when one includes the unmatched observations, estimated Internet savings rise about 10 percent from about 27 percent to between 31 and 33 percent. Hence the results suggest that omission of the unmatched observations may lead to a modest underestimation of the price savings from the Internet.

²⁶Note that this result is driven in part by the adverse selection problem where customers who want lower fares are willing to accept various restrictions to get them and are more prone to shop on the Internet. We ran a simple logit model with an online dummy variable as our dependent variable and the ticket characteristics as the explanatory variables. We did find strong evidence that ticket characteristics associated with lower prices, namely Saturday stayover, travel restriction, travel during nonpeak times, and advance purchase requirements are more likely to be bought online than offline. However, tickets associated with minimum and maximum stay restrictions were found to be less likely to be purchased on the Internet.

TABLE 3—BENEFIT FROM PURCHASING ON THE INTERNET

	log (roundtrip fare)	log (roundtrip fare)	log (roundtrip fare)	log (roundtrip fare)
	Matched transactions only		Matched and unmatched transactions	
Online	-0.297029 (0.012171)**	-0.319548 (0.013830)**	-0.367019 (0.014302)**	-0.406488 (0.016717)**
<i>Departure day of week [Omitted: Sunday]</i>				
Monday	0.00753 (0.007259)		-0.002537 (0.004946)	
Tuesday	-0.004714 (0.008598)		-0.022804 (0.006065)**	
Wednesday	-0.01142 (0.009239)		-0.03638 (0.008065)**	
Thursday	-0.043601 (0.011310)**		-0.071233 (0.010456)**	
Friday	-0.100444 (0.013039)**		-0.131035 (0.012072)**	
Saturday	-0.150466 (0.015151)**		-0.214286 (0.015593)**	
<i>Return day of week [Omitted: Sunday]</i>				
Monday	-0.237293 (0.012470)**		-0.227638 (0.009373)**	
Tuesday	-0.189137 (0.012568)**		-0.173215 (0.008727)**	
Wednesday	-0.187893 (0.012250)**		-0.16933 (0.009177)**	
Thursday	-0.174843 (0.012215)**		-0.131665 (0.009465)**	
Friday	-0.181784 (0.012533)**		-0.122816 (0.010118)**	
Saturday	-0.245828 (0.011482)**		-0.211923 (0.010257)**	
Carrier fixed effects	Yes	Yes	Yes	Yes
Route fixed effects	Yes	Yes	Yes	Yes
Constant	5.573233 (0.047855)**	5.425069 (0.047016)**	5.516633 (0.034126)**	5.360861 (0.031223)**
Observations	453,347	453,347	1,553,608	1,553,608
R ²	0.39	0.36	0.41	0.38

As mentioned above, this approach likely overestimates the gains from buying on the Internet, since much of the savings is generated from customers' willingness to accept highly restricted tickets with lower prices. To underscore the role of ticket characteristics on prices paid, Table 4 presents a simple hedonic regression of individual ticket prices on ticket restrictions and characteristics. To isolate the incremental effects of ticket-specific characteristics from carrier and route characteristics on individual prices paid, we use a stepwise regression approach. The first column in Table 4 only includes the regression of roundtrip fares on route and carrier fixed effects. The route and carrier fixed effects explain only 32 percent of the variation in the prices paid. Column 2 of Table 4 includes route and carrier fixed effects, as well as the ticket characteristics and scarcity measures described above, and a hub variable. The results show that these variables explain about 70 percent of the variation

TABLE 4—REGRESSION OF NATURAL LOGARITHM OF FARES ON TICKET CHARACTERISTICS

	log (roundtrip fare)	log (roundtrip fare)
<i>Advance purchase restriction [omitted: no advance purchase requirement]</i>		
1-day		-0.293551 (0.083308)**
3-day		-0.051236 (0.024733)*
5-day		-0.582979 (0.041268)**
7-day		-0.188988 (0.023901)**
10-day		-0.245013 (0.031068)**
14-day		-0.285223 (0.020716)**
21-day		-0.320397 (0.033361)**
30-day		-0.134444 (0.077049)
<i>Other ticket characteristics</i>		
Nonrefundable		-0.309812 (0.050447)**
Days in advance ticket purchased		-0.000566 (0.000160)**
Saturday stayover		-0.146626 (0.012254)**
Travel restriction		-0.276083 (0.016360)**
Minimum stay required		-0.022372 (0.020398)
Maximum stay restriction		-0.042493 (0.023926)
Full coach fare class		0.389817 (0.044540)**
<i>Remaining ticket characteristics</i>		
Roundtrip		-0.066355 (0.010889)**
Deviation in load factor		0.229491 (0.030558)**
Hub		0.032793 (0.037209)
<i>Time of day</i>		
Departure and return at peak time		0.042727 (0.006011)**
Either departure or return at peak time, but not both		0.017698 (0.003320)**
Return day of week fixed effects	No	Yes
Departure day of week fixed effects	No	Yes
Carrier fixed effects	Yes	Yes
Route fixed effects	Yes	Yes
Constant	5.40617 (0.051998)**	6.04765 (0.083907)**
Observations	453,347	453,347
R ²	0.32	0.7

**Significant at the 1 percent level.

*Significant at the 5 percent level.

in ticket prices. In sum, ticket characteristics along with peaking, load factor, and hubbing variables explain an additional 38 percent of the variation in ticket prices when included with route and carrier fixed effects.

The regression coefficients generally have the expected signs, and the results comport with conventional wisdom regarding airline pricing. The results show that the ticket characteristics with the largest effects on ticket prices are refundability, class of travel, stay restrictions, and advance purchase requirements. Nonrefundable tickets on average cost about 27 percent less than refundable ones (noting that the percentage reduction is modestly less than the coefficient in the logarithmic specification). Advance purchase requirements generally reduce ticket prices, typically by 20–30 percent but sometimes by more than 50 percent. These large effects from advance purchase requirements occur even though we separately control for the days in advance a ticket was bought. It is worth noting, however, that the pricing for advance purchase requirements is not monotonic. Longer periods of advance purchase do not necessarily correspond to lower prices. The results also show that travel restrictions have a major effect on ticket prices, reducing ticket prices by about 24 percent. An itinerary involving a Saturday night stayover on average costs roughly 14 percent less than an itinerary with similar characteristics that does not.

The time of day variables are all highly statistically significant and of the expected sign, but their economic magnitudes are modest. Departures and returns during peak times are just 4 percent higher than travel during off-peak hours.²⁷ The results also show that a 10 percent increase in the average load factor at the time of ticket purchase is associated with an increase in average fares of about 2 percent.

IV. Empirical Results: The Impact of the Internet on Price Levels

Turning to the effects of the Internet, Table 5 presents results testing whether customers find lower prices on the Internet. These regressions also introduce market structure variables for additional control. These variables and other route-specific variables could not be included in Table 4 because of the route fixed effects. Columns 1 and 2 of Table 5 present OLS and instrumental variable (IV) regressions of individual ticket prices on ticket characteristics, market structure variables, and carrier fixed effects.²⁸ These columns test the hypothesis that online consumers pay lower prices on average for tickets, controlling for ticket characteristics, market structure, purchase, departure and return time, and carrier.

The coefficients of the remaining control variables are similar to those in Table 4. The coefficients of the market structure variables are consistent with the literature but are not statistically significant.²⁹ Fares on routes that involve a carrier's hub at either or both endpoints average 8 percent more than fares on routes not involving a carrier's hub, but this coefficient is not statistically significant. The presence of an

²⁷ These variables reflect peak travel times between 7:00–10:00 AM and 3:00–7:00 PM, when there are generally high load factors on individual routes and throughout airline networks.

²⁸ IV regression involves instrumenting for the market share and HHI variables using their respective instruments.

²⁹ The variables lose their statistical significance once the standard errors are corrected by clustering at the carrier-route level. The error terms are likely to be correlated for transactions on the same carrier on a route.

TABLE 5—DIRECT EFFECTS OF INTERNET PURCHASES

	log (roundtrip fare) [OLS]	log (roundtrip fare) [OLS]
<i>Market structure variables</i>		
Market share	0.137452 (0.094218)	0.220174 (0.299660)
HHI	-0.048327 (0.113034)	-0.233344 (0.311476)
Hub	0.084259 (0.039784)*	0.078244 (0.053265)
Slot restricted airport	0.138823 (0.046903)**	0.147189 (0.051995)**
<i>Internet variable</i>		
Online	-0.122171 (0.008344)**	-0.120013 (0.010673)**
<i>Advance purchase restriction</i>		
1-day	-0.427676 (0.074626)**	-0.418618 (0.076816)**
3-day	0.041527 (0.034951)	0.044996 (0.034889)
5-day	-0.738733 (0.055559)**	-0.732138 (0.059516)**
7-day	-0.149978 (0.027976)**	-0.147961 (0.028177)**
10-day	-0.28948 (0.031201)**	-0.284522 (0.031422)**
14-day	-0.207232 (0.024959)**	-0.202646 (0.025153)**
21-day	-0.271809 (0.031045)**	-0.269982 (0.032276)**
30-day	0.061056 (0.076167)	0.059172 (0.074316)

(Continued)

airport with restricted slots at either endpoint increases fares by about 16 percent.³⁰ The presence of a low-cost carrier, other than Southwest, decreases average fares by roughly 8 percent, while Southwest's presence decreases average fares by roughly 15 percent. Distance between endpoints and the average population at the endpoints increase average fares while average per capita income does not statistically affect prices.³¹ The share of leisure passengers on a route does not significantly affect the average fare on a route.

Regarding Internet effects, the results show that after controlling for ticket characteristics, market structure, and carrier fixed effects, online customers pay about 11 percent less than offline customers. Hence the results show a substantial effect of Internet purchase on ticket prices. These results, which control for most sources of

³⁰ JFK (New York), LaGuardia (New York), and DCA (Washington, DC) are the three airports which still have restricted slots. Chicago O'Hare (ORD) was included in this list until 2002.

³¹ See the online Appendix for further discussion of this variable.

TABLE 5—DIRECT EFFECTS OF INTERNET PURCHASES (*Continued*)

	log (roundtrip fare) [OLS]	log (roundtrip fare) [OLS]
<i>Other ticket characteristics</i>		
Nonrefundable	-0.21544 (0.056845)**	-0.220343 (0.057224)**
Days in advance ticket purchased	-0.000991 (0.000227)**	-0.001009 (0.000231)**
Saturday stayover	-0.131459 (0.012625)**	-0.130204 (0.012535)**
Travel restriction	-0.284196 (0.023111)**	-0.287217 (0.023129)**
Minimum stay required	0.003667 (0.022975)	0.006911 (0.024119)
Maximum stay restriction	-0.012487 (0.023292)	-0.009167 (0.023193)
Full coach fare class	0.468984 (0.056248)**	0.478378 (0.055782)**
<i>Remaining ticket characteristics</i>		
Roundtrip	-0.062448 (0.012052)**	-0.06202 (0.013715)**
Deviation in load factor	0.208794 (0.035753)**	0.204473 (0.035767)**
<i>Time of day</i>		
Departure and return at peak time	0.054656 (0.012760)**	0.055344 (0.013073)**
Either departure or return at peak time, but not both	0.025443 (0.006321)**	0.025793 (0.006444)**
<i>Other route-specific characteristics</i>		
Low-cost carrier on route	-0.082439 (0.035308)*	-0.080337 (0.035272)*
Southwest	-0.173375 (0.052571)**	-0.16543 (0.052979)**
Distance (log)	0.343066 (0.029006)**	0.342998 (0.033867)**

(Continued)

variation in ticket characteristics and market structure, are both economically and statistically significant.³²

The regressions in Table 5 do not contain route fixed effects, but instead follow standard practice by including market structure variables. This standard procedure is required to allow estimation of the effect of market structure variables and other variables that do not vary at the route level, including the percentage of Internet purchases.³³ The online Appendix, however, contains a similar estimation with a full set of route fixed effects. Such estimation requires omission of the market structure

³² We have also rerun the regressions in Table 3 replicating the “bucket structure” described earlier in the paper; we included various combinations of restrictions to allow for variation in the pricing of different combinations of the restrictions described in Table 3. These regressions included the roughly 100 different restriction combinations that make up the large majority of tickets sold. The Internet effects were roughly the same as in Table 3.

³³ Estimation without these effects follows well-established procedures for analysis in this industry, see, for example, Borenstein (1989), Borenstein and Rose (1994), and Stavins (2001).

TABLE 5—DIRECT EFFECTS OF INTERNET PURCHASES (*Continued*)

	log (roundtrip fare) [OLS]	log (roundtrip fare) [OLS]
Tourist route	-0.276802 (0.189297)	-0.29956 (0.185686)
Population (log)	-0.036802 (0.021296)	-0.041709 (0.020948)*
Per capita income (log)	0.096673 (0.158348)	0.100451 (0.157350)
<i>Departure day of the week</i>		
Monday	-0.002598 (0.006827)	-0.002371 (0.006863)
Tuesday	0.007247 (0.008221)	0.007764 (0.008265)
Wednesday	0.01262 (0.008593)	0.013272 (0.008535)
Thursday	0.016519 (0.008875)	0.017338 (0.008720)*
Friday	0.01946 (0.008091)*	0.020079 (0.007990)*
Saturday	-0.082639 (0.009574)**	-0.081442 (0.009510)**
<i>Return day of the week</i>		
Monday	-0.037646 (0.006585)**	-0.037499 (0.006674)**
Tuesday	-0.026283 (0.007844)**	-0.026103 (0.008083)**
Wednesday	-0.034065 (0.007282)**	-0.034117 (0.007582)**
Thursday	-0.025419 (0.007891)**	-0.025947 (0.007939)**
Friday	-0.038151 (0.007337)**	-0.038797 (0.007392)**
Saturday	-0.07725 (0.008348)**	-0.076959 (0.008444)**
Carrier fixed effects	Yes	Yes
Constant	3.536961 (1.733246)*	3.637131 (1.707713)*
Observations	453,347	453,347
R ²	0.56	0.56

**Significant at the 1 percent level.

*Significant at the 5 percent level.

variables and other variables that only vary across markets. Including these route fixed effects results in a moderately reduced estimate of the direct effect of Internet purchase of 10 percent.

The apparent reason for these observed transaction price differences online stems from the superior search mechanism provided by the Internet, combined with potential agency and/or communication problems with travel agents. As noted, fares consist of the pricing of tickets with a series of restrictions and the matching methodology ensures that all of the online fares were available both online and offline. The regressions control for characteristics accounting for roughly 70 percent of the

observed fare variation. Yet even with such control, Internet purchasers are able to find fares that cost about 11 percent less than those purchased through travel agents. Individual customers searching on the Internet are able to and do make trade-offs in ticket characteristics that enable them to pay lower prices. Hence customers buying online find better deals within the various buckets.

The likely source of the higher transacted fares offline consists either of communication or agency problems. While travel agents have access to the same fares, the process of communicating the large array of restriction combinations and the willingness of customers to accept particular restrictions is problematic. Regarding agency issues, travel agents do not face the same incentives as consumers who spend their own money on a ticket, and it may also be true that a greater percentage of offline tickets are reimbursed, resulting in offline customers using a different trade-off between price and convenience.

Another, less restrictive way to test for online/offline differences is to investigate such differences at the city-pair level and within a specific category of tickets. We then measure online and offline differences within very specific categories of similar departure times where the tickets were purchased during comparable intervals prior to departure. We then consider tickets solely within particular city-pairs. Within this narrow pool we then estimate differences for online and offline fares.

More specifically, we categorize the time of day into four sub-categories: peak morning (between 7 AM and 10 AM), midday (after 10 AM and earlier than 3 PM), peak evening (3 PM to 7 PM), and late-evening (after 7 PM). We also categorize the number of days in advance a ticket was purchased into four categories of 0 to 6 days in advance, 7 to 13 days in advance, 14 to 21 days in advance, and more than three weeks in advance. As noted, we also limit the sample to include only nonrefundable tickets with advance purchase, travel, and/or stay restrictions.³⁴ We separately consider individual city-pairs to control for potential route heterogeneity that could affect the differences in fares. By restricting attention to the single group of nonrefundable, restricted tickets and individual city-pairs this regression has better control.

This specification also allows for heterogeneous Internet effects within defined groups of comparable days in advance and departure time of day. By measuring Internet effects within these narrow groups, we can test whether lower priced online transactions are found in generally in the sub-groups of these highly comparable groups of tickets, or whether they are limited to specific groups of tickets. Further, comparison of online and offline prices within these narrow groups provides enhanced control because it allows for heterogeneous Internet effects, by assessing potentially disparate effects across narrowly defined groups.

Table 6 presents results for three major city-pairs: LaGuardia (LGA)–Chicago O’Hare (ORD); Los Angeles (LAX)–Kennedy (JFK); and Chicago O’Hare (ORD)–Newark (EWR).³⁵ Similar results were found for other city-pairs.

³⁴ Another reason to constrain the analysis to the most restricted bucket was because of the insufficient number of online transactions in other groups.

³⁵ The regression results are reported in Table A4 in the online Appendix.

TABLE 6

Panel A. Chicago to LaGuardia

		7 AM–10 AM coefficients	10 AM–3 PM coefficients	3 PM–7 PM coefficients	After 7 PM coefficients	Weighted mean of row coefficients
0–6 days in advance	All tickets	5.991 (0.0127)**	5.952 (0.012)**	6.026 (0.001)**	5.965 (0.013)**	5.987 (0.006)**
	Tickets bought online	–0.316 (0.071)**	–0.182 (0.045)**	–0.202 (0.045)**	–0.374 (0.044)**	–0.262 (0.024)**
7–13 days in advance	All tickets	5.750 (0.014)**	5.735 (0.014)**	5.778 (0.011)**	5.796 (0.016)**	5.765 (0.007)**
	Tickets bought online	–0.356 (0.051)**	–0.366 (0.052)**	–0.297 (0.041)**	–0.371 (0.056)**	–0.34 (0.024)**
14–21 days in advance	All tickets	5.528 (0.013)**	5.484 (0.011)**	5.623 (0.011)**	5.501 (0.014)**	5.538 (0.006)**
	Tickets bought online	–0.175 (0.055)**	–0.111 (0.033)**	–0.236 (0.029)**	–0.146 (0.037)**	–0.173 (0.018)**
More than 21 days in advance	All tickets	5.437 (0.010)**	5.431 (0.009)**	5.534 (0.009)**	5.431 (0.015)**	5.467 (0.005)**
	Tickets bought online	–0.191 (0.024)**	–0.147 (0.022)**	–0.259 (0.018)**	–0.142 (0.033)**	–0.204 (0.011)**
Weighted mean of column coefficients	All tickets	5.630 (0.006)**	5.596 (0.006)**	5.716 (0.005)**	5.679 (0.007)**	
	Tickets bought online	–0.22 (0.019)**	–0.162 (0.016)**	–0.254 (0.016)**	–0.214 (0.020)**	

(Continued)

The results in Table 6 indicate that Internet purchase is strongly associated with lower prices for all groupings of tickets in these restricted samples. All of the Internet effects are significant and negative within these groups, with point estimates typically ranging between 10 and 35 percent. The right-hand column in Table 6 reports the weighted means of the row variables corresponding to the different days in advance the tickets are purchased, weighted by the number of transactions in each cell. For example, the weighted mean fare of all roundtrip tickets bought 0 to 6 days prior to departure across all times of the day on ORD–EWR route is 5.77 (in logs), about \$321. Corresponding to this specific sub-group an online customer on average pays roughly 18 percent less than the offline customer. Similar results hold true for the remaining sub-groups.

We can also calculate the weighted mean fares of all tickets and weighted online savings for travel at a particular time of day, where tickets were bought in different periods prior to actual departure. For ORD–EWR, the weighted mean fare of all tickets bought for travel in the peak morning hours (7 AM–10 AM) is 5.63 (in logs), equal to \$279 for a roundtrip ticket. The online customer on average pays 18 percent less for travel in the peak morning hours than customers who buy their tickets offline. The magnitude of online savings, on this route, is robust to travel at other times of the day. The results are similar for the other routes. These effects, for individual city-pairs, are higher than the aggregate effects in Table 5, indicating that Internet savings could be as high as 15 to 20 percent.

TABLE 6 (Continued)

<i>Panel B. JFK to Los Angeles</i>		7 AM–10 AM coefficients	10 AM–3 PM coefficients	3 PM–7 PM coefficients	After 7 PM coefficients	Weighted mean of row coefficients
0–6 days in advance	All tickets	6.165 (0.014)**	6.201 (0.013)**	6.196 (0.013)**	6.199 (0.016)**	6.190 (0.007)**
	Tickets bought online	−0.204 (0.041)**	−0.276 (0.042)**	−0.227 (0.045)**	−0.198 (0.046)**	−0.228 (0.022)**
7–13 days in advance	All tickets	5.734 (0.012)**	5.764 (0.013)**	5.771 (0.014)**	5.765 (0.017)**	5.756 (0.007)**
	Tickets bought online	−0.107 (0.032)**	−0.208 (0.036)**	−0.155 (0.042)**	−0.110 (0.040)**	−0.143 (0.018)**
14–21 days in advance	All tickets	5.820 (0.012)**	5.838 (0.012)**	5.833 (0.014)**	5.820 (0.018)**	5.828 (0.007)**
	Tickets bought online	−0.189 (0.032)**	−0.186 (0.035)**	−0.123 (0.036)**	−0.158 (0.034)**	−0.165 (0.017)**
More than 21 days in advance	All tickets	5.770 (0.011)**	5.779 (0.011)**	5.764 (0.013)**	5.755 (0.018)**	5.769 (0.007)**
	Tickets bought online	−0.078 (0.026)**	−0.095 (0.026)**	−0.039 (−0.030)	−0.006 (−0.033)	−0.086 (0.018)**
Weighted mean of column coefficients	All tickets	5.848 (0.006)**	5.878 (0.006)**	5.892 (0.007)**	5.887 (0.009)**	
	Tickets bought online	−0.129 (0.016)**	−0.166 (0.017)**	−0.160 (0.024)**	−0.152 (0.023)**	

(Continued)

The preferred approach to delineate the difference between online and offline fares would be to compare the matched pairs of online and offline tickets on the same flight, travel date and time, and ticket restrictions. This methodology, unfortunately, slices the data too thin, limiting the feasibility of the approach. The limiting factor is the combination of days at the city pair/carrier level. To override the thin data issue, we regressed the log fares on ticket characteristics, carrier and route characteristics, excluding the Internet variable—a slight variation of the approach in Table 5. This procedure then provides a set of residuals for both online and offline purchases. We then calculated the kernel densities for these two sets of residuals, one for offline and one for online for each day in advance of purchase.

Figures 6 and 7 show the distribution of online and offline residuals for tickets purchased five and ten days in advance.³⁶ The remaining graphs for each particular day are qualitatively similar and are available separately from the authors. Inspection of the densities shows clearly that the distribution of online residuals is to the left (lower) as compared to the distribution of offline residuals.

³⁶The distributions of online and offline residuals for tickets purchased corresponding to other days prior to departure can be made available on request.

TABLE 6 (Continued)

		7 AM–10 AM coefficients	10 AM–3 PM coefficients	3 PM–7 PM coefficients	After 7 PM coefficients	Weighted mean of row coefficients
<i>Panel C. Chicago to Newark</i>						
0–6 days in advance	All tickets	5.770 (0.012)**	5.739 (0.011)**	5.813 (0.008)**	5.716 (0.013)**	5.771 (0.006)**
	Tickets bought online	–0.198 (0.036)**	–0.174 (0.029)**	–0.240 (0.027)**	–0.168 (0.033)**	–0.198 (0.015)**
7–13 days in advance	All tickets	5.599 (0.019)**	5.596 (0.017)**	5.637 (0.013)**	5.612 (0.019)**	5.616 (0.009)**
	Tickets bought online	–0.164 (0.047)**	–0.171 (0.048)**	–0.106 (0.041)*	–0.135 (0.054)*	–0.143 (0.023)**
14–21 days in advance	All tickets	5.526 (0.016)**	5.526 (0.013)**	5.550 (0.010)**	5.533 (0.021)**	5.537 (0.007)**
	Tickets bought online	–0.245 (0.063)**	–0.217 (0.036)**	–0.207 (0.034)**	–0.217 (0.068)**	–0.216 (0.022)**
More than 21 days in advance	All tickets	5.508 (0.016)**	5.473 (0.012)**	5.514 (0.011)**	5.484 (0.020)**	5.497 (0.007)**
	Tickets bought online	–0.217 (0.047)**	–0.158 (0.028)**	–0.208 (0.024)**	–0.287 (0.045)**	–0.203 (0.016)**
Weighted mean of column coefficients	All tickets	5.633 (0.008)**	5.594 (0.007)**	5.658 (0.006)**	5.628 (0.010)**	
	Tickets bought online	–0.204 (0.023)**	–0.176 (0.017)**	–0.205 (0.015)**	–0.194 (0.022)**	

The Kolmogorov-Smirnov equality-of-distributions test was used, and it rejects the null that the online and offline distributions are the same.³⁷

In sum, the substantial price differential between online and offline prices reported in Table 5 is robust even when one conducts an analysis of nonrefundable, restricted tickets at the city-pair level and considers differences for tickets purchased the same number of days prior to departure, and for travel at similar times of day. The robustness of the results also suggests that this difference more likely stems from superior search on the Internet and is unlikely attributable to heterogeneity across routes or unobservable factors causing disparity in fares paid online versus offline.

V. Indirect Effects of Internet on Average Prices

Table 7 analyzes the effect of the share of Internet purchases on the average level of fares. In addition to the variables used in Table 5 we include an Internet share variable representing the share of Internet purchases on a given route. Columns 1 and 3 of Table 7 report the OLS estimates while columns 2 and 4 report the estimates

³⁷ Similar kernel densities for other days prior to departure have also been constructed, and the results are similar to those reported here. These additional results are available on request.

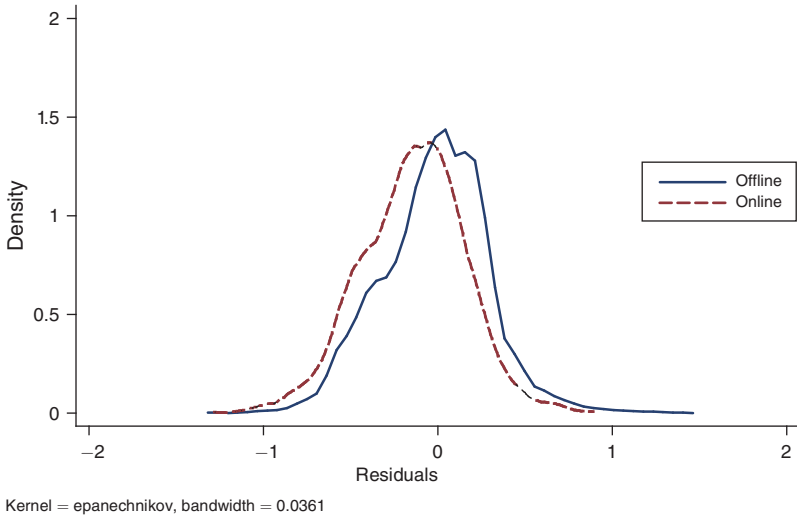


FIGURE 6. COMPARING THE KERNEL DENSITIES OF FARES NET OF TICKET, CARRIER, AND ROUTE CHARACTERISTICS FOR ONLINE AND OFFLINE TRANSACTIONS—FIVE DAYS PRIOR TO DEPARTURE DATE

from the instrumental variable (IV) regressions.³⁸ While the Internet share variable is large and highly significant in the OLS specifications, it is not significant in the IV specifications. Hence the results do not support the presence of spillovers from Internet shopping into reductions of fares more generally.³⁹ It should also be noted that we have run specifications with weaker control variables for tourist shares, particularly regressions using absolute temperature differentials. In these specifications, the online share coefficient was large and significant. The contrast between those results and the ones in Table 7 highlights the importance of using appropriate measures of tourism to control for the incremental “shopping” effects associated with tourists. The results indicate that without accurate tourism measures one can incorrectly attribute tourist effects to the Internet.⁴⁰

These results stand in contrast to those found by Brown and Goolsbee (2002) in the life insurance industry. Our results show that in the airline industry reduced prices appear to be captured by those shopping online without significant spillovers to other prices. A likely reason for the difference is the general complexity of airline fares in contrast to what is likely to be more standardized pricing for term life insurance. Still, regardless of the reasons, the evidence here does not support the

³⁸ Market share, HHI, and share of online transactions on a route are the instrumented variables. Recall that the share of online transactions on a route is instrumented using two variables—the proportion of transactions in the period three or more weeks prior to departure, and the share of Internet purchases on other routes out of the endpoint airports. The sensitivity of using each of these instruments individually is reported in Table A9 of the online Appendix. Those results indicate that the results are not robust with respect to the choice of instrument, raising the possibility that one or more of the instruments is not valid. It is important to note, however, that the magnitude and significance of the remaining primary variables of interest are largely unaffected by the choice of instrument.

³⁹ Note that the last two columns in Table 7 also report results for an interaction term between Internet share and the online variable. This variable allows for the possibility that a greater Internet share leads to reduced pricing in that channel, but this coefficient is not significant.

⁴⁰ A robustness check using separate controls for outbound and inbound tourist shares for the endpoint airports is presented in Table A8 of the online Appendix.

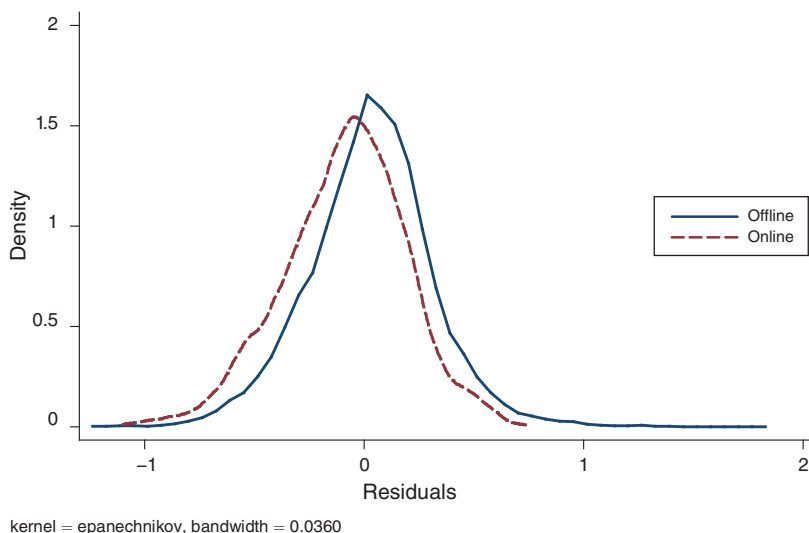


FIGURE 7. COMPARING THE KERNEL DENSITIES OF FARES NET OF TICKET, CARRIER, AND ROUTE CHARACTERISTICS FOR ONLINE AND OFFLINE TRANSACTIONS—TEN DAYS PRIOR TO DEPARTURE DATE

presence of pricing spillovers in airlines as the share of tickets purchased on the Internet rises.

The results do, however, offer marginal support for the proposition that greater tourist shares lead to reduced prices independent of the channel of purchase. Ticket prices systematically fall with the share of tourists on a route. It is worth noting, moreover, that the measure of tourism developed by Borenstein (2010) seems to be the superior measure for capturing these effects.⁴¹

To summarize, Tables 5 through 7 provide two important results. First, controlling for ticket characteristics, capacity, route and carrier fixed effects, online customers on average pay roughly 11 percent less than the offline consumers. Second, equilibrium prices do not appear to fall as the share of Internet shoppers on a route rises. The results do, however, offer marginally significant support for the proposition that a greater tourist share leads to reduced prices.

VI. Price Dispersion

Stahl (1989) shows that the degree of price dispersion is not monotonic with respect to the percentage of informed consumers or searchers. Instead, when there are no searchers there is a degenerate distribution of prices at the monopoly price. Dispersion then rises as some consumers become informed, eventually reaching a peak before descending back to zero at the competitive level when all consumers are informed. Brown and Goolsbee (2002) find evidence of such a nonmonotonic relationship between online search and offline price dispersion for term life insurance. They attribute much of the nonlinearity to the gradual maturation of Internet

⁴¹ We would like to thank the referee for suggesting the use of this measure.

TABLE 7—INTERNET PURCHASE, INTERNET SHARE, AND POTENTIAL SAVINGS FROM INTERNET PURCHASE IN HIGH-LOW INTERNET USAGE MARKET

	log (roundtrip fare) [OLS]	log (roundtrip fare) [IV]	log (roundtrip fare) [OLS]	log (roundtrip fare) [IV]
<i>Market structure variables</i>				
Market share	0.134 (0.094)	0.246 (0.301)	0.133 (0.094)	0.247 (0.301)
HHI	-0.060 (0.112)	-0.269 (0.312)	-0.059 (0.112)	-0.270 (0.310)
Hub	0.087 (0.039)*	0.076 (0.054)	0.087 (0.039)*	0.076 (0.054)
Slot restricted airport	0.139 (0.046)**	0.146 (0.052)**	0.139 (0.046)**	0.146 (0.052)**
<i>Internet variables</i>				
Online	-0.118 (0.008)**	-0.118 (0.011)**	-0.142 (0.026)**	-0.126 (0.029)**
Share of online transactions	-0.444 (0.171)**	-0.179 (0.251)	-0.464 (0.181)*	-0.187 (0.260)
Share of online purchases × online			0.106 (0.120)	0.035 (0.135)
<i>Advance purchase restriction</i>				
1-day	-0.414 (0.070)**	-0.413 (0.075)**	-0.413 (0.070)**	-0.413 (0.075)**
3-day	0.045 (0.035)	0.047 (0.035)	0.046 (0.035)	0.047 (0.035)
5-day	-0.699 (0.053)**	-0.717 (0.061)**	-0.700 (0.053)**	-0.717 (0.061)**
7-day	-0.149 (0.028)**	-0.147 (0.028)**	-0.148 (0.028)**	-0.147 (0.028)**
10-day	-0.286 (0.032)**	-0.283 (0.032)**	-0.285 (0.032)**	-0.283 (0.032)**
14-day	-0.209 (0.025)**	-0.203 (0.025)**	-0.208 (0.025)**	-0.203 (0.025)**
21-day	-0.272 (0.031)**	-0.270 (0.032)**	-0.271 (0.031)**	-0.270 (0.032)**
30-day	0.060 (0.073)	0.059 (0.073)	0.061 (0.072)	0.059 (0.073)

(Continued)

markets.⁴² They argue, based on Stahl (1989), that starting with a small enough share of Internet users, giving a small group of customers access to more information will initially cause dispersion to rise, but as more and more customers become informed, dispersion will eventually fall. Starting with a sufficiently high fraction of searchers (high Internet share usage), Stahl's (1989) model would predict a monotonic negative relationship between Internet purchases and price dispersion. We investigate similar issues but evaluate differences in price dispersion in the cross section, exploiting differences in Internet usage.

⁴²See Bailey (1998) and Brynjolfsson and Smith (1999) for implications of mature and immature Internet markets.

TABLE 7—INTERNET PURCHASE, INTERNET SHARE, AND POTENTIAL SAVINGS FROM INTERNET PURCHASE IN HIGH—LOW INTERNET USAGE MARKET (*Continued*)

	log (roundtrip fare) [OLS]	log (roundtrip fare) [IV]	log (roundtrip fare) [OLS]	log (roundtrip fare) [IV]
<i>Other ticket characteristics</i>				
Nonrefundable	-0.214 (0.057)**	-0.220 (0.057)**	-0.214 (0.057)**	-0.220 (0.057)**
Days in advance ticket purchased	-0.001 (0.0002)**	-0.001 (0.0002)**	-0.001 (0.0002)**	-0.001 (0.0002)**
Saturday stayover	-0.129 (0.013)**	-0.129 (0.012)**	-0.130 (0.013)**	-0.129 (0.013)**
Travel restriction	-0.281 (0.023)**	-0.286 (0.023)**	-0.281 (0.023)**	-0.286 (0.023)**
Minimum stay required	0.003 (0.023)	0.007 (0.024)	0.003 (0.023)	0.007 (0.024)
Maximum stay restriction	-0.010 (0.024)	-0.008 (0.023)	-0.010 (0.024)	-0.008 (0.023)
Full coach fare class	0.473 (0.056)**	0.480 (0.056)**	0.472 (0.056)**	0.480 (0.056)**
<i>Remaining ticket characteristics</i>				
Roundtrip	-0.065 (0.012)**	-0.064 (0.014)**	-0.065 (0.012)**	-0.064 (0.014)**
Deviation in load factor	0.210 (0.036)**	0.205 (0.036)**	0.209 (0.036)**	0.204 (0.036)**
<i>Time of day</i>				
Departure and return at peak time	0.055 (0.013)**	0.056 (0.013)**	0.055 (0.013)**	0.056 (0.013)**
Either departure or return at peak time, but not both	0.026 (0.006)**	0.026 (0.006)**	0.026 (0.006)**	0.026 (0.006)**
<i>Other route-specific characteristics</i>				
Low-cost carrier on route	-0.069 (0.035)	-0.075 (0.036)*	-0.068 (0.035)	-0.075 (0.036)*
Southwest	-0.154 (0.052)**	-0.157 (0.054)**	-0.155 (0.052)**	-0.157 (0.054)**
Distance (log)	0.369 (0.031)**	0.352 (0.036)**	0.370 (0.031)**	0.352 (0.036)**
Tourist route	-0.136 (0.190)	-0.248 (0.180)	-0.135 (0.190)	-0.247 (0.180)
Population (log)	-0.045 (0.022)*	-0.045 (0.022)*	-0.045 (0.022)*	-0.045 (0.022)*
Per capita income (log)	0.026 (0.160)	0.073 (0.161)	0.025 (0.161)	0.073 (0.161)

(Continued)

Following Brown and Goolsbee (2002) we adopt a two-step approach to investigate the relationship between price dispersion and Internet share. First, we estimate an hedonic pricing model, and second, we examine the residuals of that model to assess variation in price dispersion. The first equation is

$$(2) \quad \ln(\text{Price}_{ijkt}) = \alpha + \beta_1(\mathbf{R}_{ijkt}) + \beta_1(\mathbf{Travel}_{ijkt}) + \beta_2(\mathbf{Route}_j) \\ + \beta_3(\mathbf{Carrier}_{ijkt}) + \varepsilon_{ijkt},$$

TABLE 7—INTERNET PURCHASE, INTERNET SHARE, AND POTENTIAL SAVINGS FROM INTERNET PURCHASE IN HIGH—LOW INTERNET USAGE MARKET (*Continued*)

	log (roundtrip fare) [OLS]	log (roundtrip fare) [IV]	log (roundtrip fare) [OLS]	log (roundtrip fare) [IV]
<i>Departure day of the week</i>				
Monday	-0.002 (0.007)	-0.002 (0.007)	-0.002 (0.007)	-0.002 (0.007)
Tuesday	0.007 (0.008)	0.008 (0.008)	0.007 (0.008)	0.008 (0.008)
Wednesday	0.013 (0.009)	0.013 (0.009)	0.012 (0.009)	0.013 (0.009)
Thursday	0.017 (0.009)	0.017 (0.009)*	0.016 (0.009)	0.017 (0.009)*
Friday	0.019 (0.009)*	0.020 (0.008)*	0.019 (0.008)*	0.020 (0.008)*
Saturday	-0.083 (0.010)**	-0.081 (0.009)**	-0.083 (0.010)**	-0.082 (0.010)**
<i>Return day of the week</i>				
Monday	-0.038 (0.007)**	-0.038 (0.007)**	-0.038 (0.006)**	-0.038 (0.007)**
Tuesday	-0.026 (0.008)**	-0.026 (0.008)**	-0.027 (0.008)**	-0.026 (0.008)**
Wednesday	-0.033 (0.007)**	-0.034 (0.008)**	-0.034 (0.007)**	-0.034 (0.007)**
Thursday	-0.025 (0.008)**	-0.026 (0.008)**	-0.026 (0.008)**	-0.026 (0.008)**
Friday	-0.037 (0.007)**	-0.038 (0.007)**	-0.038 (0.007)**	-0.039 (0.007)**
Saturday	-0.077 (0.008)**	-0.077 (0.008)**	-0.077 (0.008)**	-0.077 (0.008)**
Carrier fixed effects	Yes	Yes	Yes	Yes
Constant	4.215 (1.758)*	3.909 (1.751)*	4.224 (1.757)*	3.914 (1.750)*
Observations	453,347	453,347	453,347	453,347
R ²	0.56	0.56	0.56	0.56

**Significant at the 1 percent level.

*Significant at the 5 percent level.

where inclusion of **Route_{*j*}** means we now include route fixed effects.⁴³

From (2) we obtain the residuals and calculate different measures of dispersion (standard deviation, range, and the ratio of standard deviation to mean fare) of the residuals by carrier-route for each departure date. These residuals parallel those of Brown and Goolsbee who measure residuals in life insurance prices after taking into account policy characteristics. Our residuals represent unexplained price dispersion within a carrier-route after accounting for the price effects of observable characteristics of tickets. To assess the factors affecting price dispersion, we regress various measures of dispersion on the share of tickets purchased on the Internet and on market structure variables.

⁴³Refer to Table A3 in the online Appendix for the estimated results of this equation.

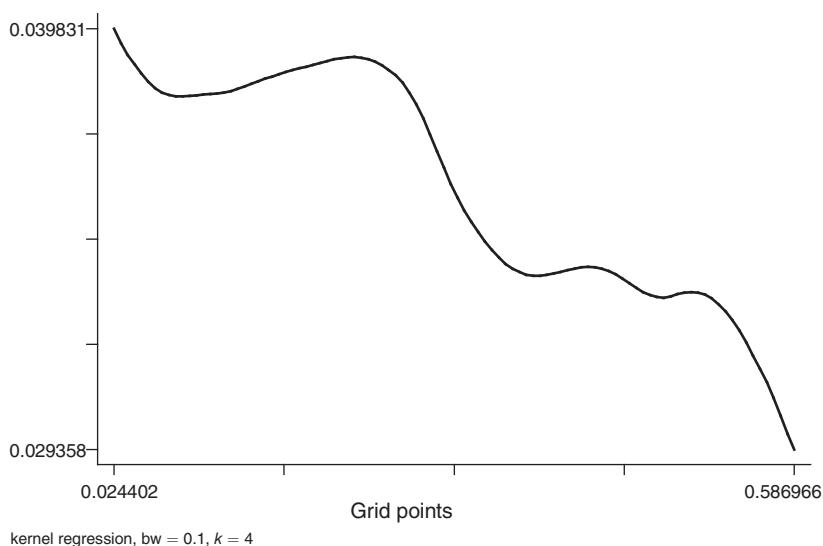


FIGURE 8. PRICE DISPERSION AND SHARE OF INTERNET PURCHASES

Figure 8 presents a kernel regression between the standard deviation of the residuals and the share of Internet usage. This regression indicates that initially there is a reasonably stable level of price dispersion, but then as online purchases rise, dispersion drops significantly. Hence an increase in Internet share appears to have a significant effect on dispersion.

Table 8 presents a regression analysis of the relationship between price dispersion, Internet share, and market structure variables. The measure of dispersion is, however, positively correlated with the average level of fares. To account for this, we scaled the measure of dispersion by average fares. In column 1 of Table 8, we regress the scaled standard deviation of the residuals on the share of Internet usage, carrier market share, and the Herfindahl index of the route, without correcting for the endogeneity issues of the latter two variables.⁴⁴ In column 2, we run the same specification as in column 1 but we now correct for the potential endogeneity of market share, Herfindahl index, and Internet share. We use the same instruments for these variables as described above. The OLS results show that price dispersion falls as the share of Internet usage increases but the results are not statistically significant in the instrumental variable specification.

The effects of market share and HHI parallels that of Borenstein and Rose (1994) but the effects are diminished, and are not generally significant. These differences may result either from our use of the scaled standard deviation as a measure of dispersion or our control for observable ticket characteristics.⁴⁵ Further, earlier studies were characterized by their lack of control for ticket characteristics so that they

⁴⁴We replicated these regressions with the range and coefficient of variation of the residuals as the dependent variables. The negative relationship between dispersion and share of Internet usage was robust in all specifications irrespective of the choice of the dependent variable.

⁴⁵Note, the use of standard deviation of residuals to measure dispersion is to parallel the analysis of Brown and Goolsbee (2002).

TABLE 8—REGRESSION OF SCALED STANDARD DEVIATION OF RESIDUALS ON MARKET STRUCTURE AND INTERNET USAGE VARIABLES

	Standard deviation of residuals (scaled by average log(fare)) [OLS]	Standard deviation of residuals (scaled by average log(fare)) [IV]	Standard deviation of residuals (scaled by average log(fare)) [OLS]	Standard deviation of residuals (scaled by average log(fare)) [IV]
Market share	0.009834 (0.002918)**	0.010419 (0.005002)*	0.009707 (0.002938)**	-0.009334 (0.006343)
HHI	-0.014545 (0.004056)**	-0.012421 (0.006879)	-0.014411 (0.004058)**	0.005914 (0.007122)
Internet share	-0.01954 (0.006010)**	-0.012209 (0.007756)	0.006003 (0.077039)	0.000695 (0.008367)
(Internet share) ²			-0.07711 (0.275054)	
(Internet share) ³			0.061804 (0.291298)	
Carrier fixed effects	No	No	No	Yes
Departure date fixed effects	No	No	No	Yes
Constant	0.044839 (0.002481)**	0.041844 (0.003278)**	0.042656 (0.006406)**	0.036137 (0.003552)**
Observations	19,338	19,338	19,338	19,338
R ²	0.02	0.02	0.02	0.08

**Significant at the 1 percent level.

*Significant at the 5 percent level.

measure dispersion with error, which could lead to poor estimation of the effect of market concentration on the dispersion of prices. This possibility is significant because our results show the important effect of ticket characteristics on prices.

The results provide, at best, weak evidence that price dispersion may fall as the share of Internet usage rises. To check for nonlinearity between dispersion and the share of Internet usage we include the square and the cube of the Internet usage share. Column 3 does not support a nonlinear relationship consistent with the theoretical predictions of Stahl (1989) or the empirical results of Brown and Goolsbee (2002). In column 4, we include carrier fixed effects and carrier-departure day-of-week fixed effects respectively in addition to the variables used in column 2, correcting for the potential endogeneity in the market structure variables and Internet share variable. In this specification however, none of the market structure or the Internet share variable is statistically significant.⁴⁶ More importantly, our result suggesting an inverse relationship between price dispersion and share of Internet usage is not robust to these alternative specifications. Thus, this evidence does not indicate that an increase in the share of Internet purchases on a route decreases overall dispersion in fares.

⁴⁶Table A6 in the online Appendix reports results where the measure of dispersion is not scaled by average fares. The results are qualitatively similar though in the final specification the sign of the internet share variable is reversed, but it is not statistically significant.

VII. Conclusions

This paper provides an in-depth empirical analysis of the effects of the Internet on the price of airline tickets. The research makes use of the most complete dataset ever used to analyze airline pricing. This novel dataset includes data on individual ticket transactions including ticket characteristics, carrier, flight level load factor at purchase, measures of peak departure and return times, the date of issue, other hedonic factors affecting prices, and an indicator to denote if the ticket was purchased online. This paper is one of only a few studies of Internet pricing using data for contemporaneous online and offline transactions in comparable geographic markets.

While controlling for numerous observed ticket characteristics, as well as carrier and route effects, the results show that consumers buying online pay about 11 percent less than those buying offline. The analysis, however, does not support the proposition that increases in the share of tickets purchased on the Internet leads to reduced fares more generally. Reduced prices associated with Internet purchases appear to be confined to those who purchase online. Regarding price dispersion, the results have some suggestive value, but do not ultimately support a conclusion that increased online shares decrease price dispersion.

The paper has also used these new data to present a more complete analysis regarding the impact of market structure on the level and dispersion of airline prices. The results for market share and HHI are not significant in the primary regressions presented above. When one controls for ticket characteristics and peaking variables, the coefficients for market share and market concentration become insignificant regarding the level and dispersion of fares.

REFERENCES

- Bailey, Joseph P.** 1998. "Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce." PhD diss. Massachusetts Institute of Technology Department of Electrical Engineering and Computer Science.
- Baye, Michael R., and John Morgan.** 2001. "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets." *American Economic Review* 91 (3): 454–74.
- Borenstein, Severin.** 1989. "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry." *RAND Journal of Economics* 20 (3): 344–65.
- Borenstein, Severin.** 2010. "An Index of Inter-City Business Travel for Use in Domestic Airline Competition Analysis." <http://www.nber.org/bti/BizTrvIdx.pdf>.
- Borenstein, Severin, and Nancy L. Rose.** 1994. "Competition and Price Dispersion in the U.S. Airline Industry." *Journal of Political Economy* 102 (4): 653–83.
- Brown, Jeffrey R., and Austan Goolsbee.** 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy* 110 (3): 481–507.
- Brynjolfsson, Erik, and Michael D. Smith.** 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46 (4): 563–85.
- Bureau of Transportation Statistics.** 2004. "T-100 Domestic Market (U.S. Carriers)." Research and Innovative Technology Administration. http://www.transtats.bts.gov/Fields.asp?Table_ID=308 (accessed February 2006).
- Chen, Jihui.** 2006. "Differences in Average Prices on the Internet: Evidence from the Online Market for Air Travel." *Economic Inquiry* 44 (4): 656–70.
- Clemons, Eric K., Il-Horn Hann, and Lorin M. Hitt.** 2002. "Price Dispersion and Differentiation in Online Travel: An Empirical Investigation." *Management Science* 48 (4): 534–49.
- Gerardi, Kristopher S., and Adam Hale Shapiro.** 2009. "Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry." *Journal of Political Economy* 117 (1): 1–37.

- Lee, Ho Geun.** 1998. "Do Electronic Marketplaces Lower the Price of Goods?" *Communications of the ACM* 41 (1): 73–80.
- Morton, Fiona Scott, Florian Zettelmeyer, and Jorge Silva-Risso.** 2001. "Internet Car Retailing." *Journal of Industrial Economics* 49 (4): 501–19.
- Pan, Xing, Brian T. Ratchford, and Venkatesh Shankar.** 2001. "Why Aren't the Prices of the Same Item the Same at Me.com and You.com?: Drivers of Price Dispersion Among E-Tailers." <http://bschool.nus.edu/departments/marketing/papers%20for%20seminars/peter%20pan.pdf>.
- Pan, Xing, Brian T. Ratchford, and Venkatesh Shankar.** 2004. "A Model of Retail Competition in Service and Price: Pure Play Internet vs. Bricks-and-Mortar Retailer." Unpublished.
- Sengupta, Anirban, and Steven N. Wiggins.** 2014. "Airline Pricing, Price Dispersion, and Ticket Characteristics On and Off the Internet: Dataset." *American Economic Journal: Economic Policy*. <http://dx.doi.org/10.1257/pol.6.1.272>.
- Scholten, Patrick, and S. Adam Smith.** 2002. "Price Dispersion Then and Now: Evidence from Retail and E-Tail Markets." In *Advances in Applied Microeconomics: The Economics of the Internet and E-Commerce*, Vol. 11, edited by Michael R. Baye and John Maxwell, 63–88. Emerald Group Publishing.
- Smith, Barry C., Dirk P. Günther, B. Venkateshwara Rao, and Richard M. Ratliff.** 2001. "E-Commerce and Operations Research in Airline Planning, Marketing, and Distribution." *Interfaces* 31 (2): 37–55.
- Stahl, Dale O., II.** 1989. "Oligopolistic Pricing with Sequential Consumer Search." *American Economic Review* 79 (4): 700–712.
- Stavins, Joanna.** 2001. "Price Discrimination in the Airline Market: The Effect of Market Concentration." *Review of Economics and Statistics* 83 (1): 200–202.

This article has been cited by:

1. James D. Dana, Jr., Eugene Orlov. 2014. Internet Penetration and Capacity Utilization in the US Airline Industry. *American Economic Journal: Microeconomics* 6:4, 106-137. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]