NET Institute*

www.NETinst.org

Working Paper #06-07

November 2006

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

Anirban Sengupta and Stevcn Wiggins
Texas A&M University

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

Airline Pricing, Price Dispersion, and Ticket Characteristics On and Off the Internet [1]

Anirban Sengupta[2]
Department of Economics
Texas A&M University

Steven N. Wiggins
Department of Economics
Texas A&M University

Current Draft: November 8, 2006

Abstract

This paper uses a unique individual transactions data set to investigate the effects of internet purchase on the prices paid for airline tickets. The analysis also investigates the effects of changes in the percentage of online transactions on both online and offline prices and on price dispersion. The analysis also uses these unique data to provide a more complete analysis of the factors affecting airline price levels and price dispersion, contributing more generally to our understanding of airline pricing. Our novel data set includes detailed transaction level data that includes ticket characteristics and restrictions, carrier, estimated flight level load factors, date of issue, departure date, other hedonic factors affecting prices, whether the ticket was purchased online or offline, and the share of online purchases for the city-pair. Controlling for numerous observed ticket characteristics, as well as carrier and route effects, the results show that online prices average about 13 percent less than offline prices. The analysis also shows that a ten percent increase in the online share of tickets sold on a route decreases average prices by an additional 5 percent, with more of this effect coming in the form of lower *offline* prices. The paper also finds evidence that an increase in online shares decrease price dispersion.

The paper also uses these unique data to investigate the effects of hub dominance and high route shares on pricing. Due to data limitations previous investigations of these issues could not control for important ticket characteristics, load factors, and time of purchase in measuring the effects of concentration on price levels and dispersion. Our analysis controls for these factors while investigating the impact of market concentration on price levels and dispersion.

---

[1] Preliminary draft. Please do not cite or quote.
[2] Please send comments and suggestions to asengupta@econmail.tamu.edu or swiggins@tamu.edu.
JEL Code: L9, Industrial Organization
Keywords: Internet, search cost, online and offline

1. **Introduction**

This paper uses a unique, individual transaction level data set 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 effect of increased route shares of internet purchases on 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 affects 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 a wide range of products. Consumers now buy online numerous products and services, including books and CDs, electronic products through shopbots, and travel and entertainment. Internet penetration is particularly substantial in airlines where in particular city-pairs the internet share exceeds fifty 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 different markets. Early studies found mixed evidence regarding prices and price dispersion.[3] Morton, Zettlemeyer, and Silva-Rissio (2002) use micro level, transaction data and they find that consumers who searched for automobiles online paid about two percent less than consumers who did not. Brown and Goolsbee (2002) investigate the

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

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 lead 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 geographical settings.[4]

These studies provide important information regarding varied aspects of internet pricing, but due to data limitations they do not provide a comprehensive analysis of the impact of the internet on internet prices, overall price levels, and price dispersion for an entire 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.[5] 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 data set from the airline industry.[6] The central goal of the paper is to investigate the effects of internet sales on prices paid for airline tickets.

---

[4] Chen (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 et al. (2002).

[5] Note that much of the existing literature on internet pricing relies on posted prices as compared to transaction data (Shankar et all (2003)).

[6] Note 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 important, their analyses do not consider route-level variation in internet share from a given

The data set consists of actual individual transactions purchased through a large Computer Reservation System (CRS).[7] 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 peak times, market structure, and route characteristics.[8] The analysis shows that these factors 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 more than .8.

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

---

origin, and they also do not have available other ticket characteristics of the type considered here. The analysis below shows that internet savings will be overestimated if one does not control for ticket characteristics.

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

[8] The 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 stay-over.

The results show that when controlling for ticket characteristics, tickets purchased on the internet cost about 13 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 more than 40 percent.[9] The results also show that high levels of internet purchases drive down both offline and online prices by about ten percent, and these effects are actually larger in the offline market. Hence internet shopping in this industry generates a general market benefit similar to that underlying the results of Brown and Goolsbee. The results also show that greater internet penetration reduces price dispersion in the market as a whole by driving offline prices down toward internet prices and by reducing dispersion within the online and offline segments.

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 pay substantially lower fares than offline purchasers, regardless of the number of days in advance when 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 eight largest city-pairs for tickets with various characteristics. The characteristics considered include class of service, refundability, advance purchase requirements, Saturday stay-over, 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.

---

[9] Note 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.

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 shows that hub dominance and high route shares lead to higher average prices. Borenstein and Rose show that price dispersion decreases as routes become more monopolistic. Due to data limitations neither study controls 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.[10] 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 generally confirm those of Borenstein and Rose.

2. 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 stay-over.[11] The received wisdom is that airlines limit the quantity 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

[10] Note 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, however, uses certain posted prices and only a subset of ticket characteristics. Using these posted prices, Stavins 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 the consumers.
[11] See Smith (2001).

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 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.[12] 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, which implicitly place 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 80 percent of the variation in ticket prices.

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 most important component of this search, in terms of its impact on the ultimate price, is to find an open "bucket" with acceptable characteristics that has a low price. An empirically smaller effect is found by identifying low priced tickets within a given bucket. The analysis below separately identifies internet price reductions that occur due to finding lower priced buckets, and from finding lower prices within a given bucket. It also identifies the externality of increased internet purchases in terms of driving down overall fares.

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.

[12] See Smith (2001).

3. Search Theory, Pricing, and the Internet

The analysis here takes the same approach to the theory of internet pricing as that found in Brown and Goolsbee (2002). The analysis assumes, 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 assumes that a certain exogenous share of customers are 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 in the Nash equilibrium stores choose prices 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 price seller.

This logic generates three well-known results noted by Brown and Goolsbee. 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 there is a degenerate distribution at the monopoly price. Dispersion then rises as some consumers become informed. Dispersion eventually reaches a peak before descending back to zero when all consumers are fully informed and prices converge to the competitive level.

We will assume that online customers are better informed than offline customers, although ultimately we will investigate this assumption in the empirical analysis by determining if they pay lower prices holding ticket characteristics constant. Further, the analysis uses transactions prices rather than posted prices. The use of transactions prices means that we measure the distribution of actual transactions rather than the posted prices of sellers, some of which might net few if any sales.

It is also important to note that the data we use includes only fares available both online and offline. The data we used incorporated observations where transaction fares were matched to those found offline in an offline CRS. Further, the available information indicates that in general the same prices and fare combinations are available online and offline.[13] 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.

4. Empirical Analysis

4. A. Data

This study uses a unique data set consisting of contemporaneous online and offline transaction data of airline tickets for the last quarter of 2004.[14] These data were provided by a leading Computer Reservation System (CRS) vendor and include all transactions for a large number of domestic routes handled by the CRS during that quarter. The CRS offers services to airlines, travel agents, and numerous online sites, so that the data include transactions for all three outlets, though we believe the share for airline sites is small. 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 thirty

[13] Note that 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 may also 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 data set. See appendix.

[14] The data and construction of variables are discussed at length in Appendix A.

percent of domestic tickets sold. 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 separate data set from a different CRS containing both fares offered and purchased for travel in particular city-pairs, by departure dates on particular airlines. These data included the ticket characteristics not available from the first data set.[15] The data set from the second CRS was incomplete in that certain fares had been deleted from the archive, and so we were only able to match the fares imperfectly.[16] The criterion used was to keep transactions if we were able to match the fares within 2 percent; for multiple matches within two percent we kept the closest. The resulting data set contains individual ticket transactions that include ticket characteristics and restrictions, together with carrier, flight information, and dates of purchase, departure, and return.[17]

This procedure matches roughly 35 percent of the observations from the first data set. For both the online and the offline transactions our match rate is somewhat lower for the lowest priced tickets. Match rates for different city pairs are illustrated in Figures 2-4. For example, Figure 2 shows the matches for Chicago to Newark; Panel A shows matches for all airlines, and Panel B for Continental, the market leader. Figures 3-4 show similar kernel densities for two other large city-pairs. The kernel densities show an under-representation 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

[15] We have been informed that fares offered on the various CRS's 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 data set with the detailed ticket characteristics found in the second data set. The details are provided in the appendix.

[16] The data in the second archive are kept for unknown intervals of time. Individual fares are then deleted in an unknown pattern of time.

[17] Since the CRS de-regulation in 2004, the airlines are free to provide different fares to any distribution channel including the major CRS's, their own CRS, own web-site and online travel agencies like Expedia. This necessitated the adoption of the matching rule(s) as discussed in the paper. Please refer to the appendix for discussion on the matching procedure.

offline fares, and the under-representation of matches is comparable in both data sets. More specifically, Figure 5 compares the kernel densities for matched and unmatched transactions broken down by online versus offline transactions. 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 differences 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 an underestimation of low fares on the internet, although these effects will be minimized in the regressions because we control for ticket characteristics. We return to this issue below.

This study uses data for 150 U.S. domestic city pairs (routes), including a mix of both business and tourist routes, and routes with varying groups of customers. A complete list of routes is contained in the Appendix, Table A2. We define a route as a city pair regardless of direction. Following Borenstein (1989) and Borenstein and Rose (1994), we include itineraries with at most one stop-over in either direction. The prices used are for roundtrip fares, doubling the fares for one-way tickets to obtain comparability. We exclude itineraries with open-jaws and circular trip tickets. This study includes tickets for flights operated by American Airlines, Continental, Delta, Northwest, US Airways, United Airlines, Frontier, Air Tran, Spirit, Alaska, American Mid-west, Sun Country, Frontier Airlines, Hawaiian Airlines and American Trans Air.[18]

Observations consist of individual tickets, their fares and characteristics, and other data described above and more fully in the 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 later in this section.

4.B. Model Specification

The basic model to be estimated regards the relationship between internet purchase and prices, which is estimated using the following specification:

$$\mathbf{ln(Price_{ijtk}) = \beta_0 + \beta_1 (R_{ijtk}) + \beta_2 (Mktstructure_j) + \beta_3 (Mktstructure_{jk}) + \beta_4 (Online_{ijtk}) + \beta_5 (Loadfactor_{ijtk}) + \beta_6 C_{ijtk} + \beta_7 (DEP_{ijtk}) + \beta_8 (RET_{ijtk}) + \varepsilon_{ijtk} \quad (1)}$$

where i refers to the i^{th} ticket, j represents route j, t represents day t (departure date) while k stands for the k-th carrier.

R_{ijtk} represents the vector of the ticket restrictions (or characteristics) associated with the ith ticket. The set of characteristics include sets of dummy variables including refundability, advance purchase requirements, Saturday night stay-over, direct routes, round trip travel, first class, business class, travel restrictions (example, if the ticket is valid only if you travel on a MTF), and stay restrictions (minimum or maximum stay requirement). These variables also include the number of days in advance the ticket was purchased. These variables also include peak times of day, such as flights departing on weekdays between 7:00 and 10:00 a.m. or

[18] We 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 data bases.

between 3:00 and 7:00 p.m.[19] These are periods of peak travel, generally with high load factors and there may be network cost/capacity effects resulting in higher prices during these periods. 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 the omitted day.

The variables $Mktstructure_j$ represent a vector of route specific variables that have been widely used to study airline pricing (see, e.g., Borenstein). The Herfindahl index is measured using passenger shares in the fourth quarter of 2004. We also use a low cost route dummy indicating the presence of a low cost carrier and a separate Southwest dummy indicating Southwest's presence. We also include the logs of statutory distance, average population for the city pair, average per capita income, and the temperature differential between the origin and destination, with the latter potentially measuring tourist effects (see Brueckner and Spiller (1992, 1994)). The second market structure variable $Mktstructure_{jk}$ consists of the market share of carrier k on route j and a hub dummy if the carrier has a hub at the origin or destination. $Online_{ijtk}$ is an indicator variable corresponding to online purchases; C_{ijtk} represents carrier-specific fixed effects.

We employ both OLS and Instrument variable (IV) regression techniques to estimate equation (1). The IV regression is required to address potential endogeneity issues of market share and the Herfindahl index. In the IV approach, we use the same instruments as Borenstein (1989) and Borenstein and Rose (1994). 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.[20] To the extent that market share is endogenous, the route Herfindahl

[19] We also used the time window of 8-10am as a peak time window. The results were qualitatively unaffected.
[20] The Herfindahl index is subject to endogeneity issues because it is composed of market shares.

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 all other carriers' shares on the route.[21]

To investigate the effects of increased internet usage on fares 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 use an estimate of people who have looked at prices on the internet for a given group, while our data permit direct measurement of the share of transactions completed on the internet.[22] We also include an interaction variable to measure whether the savings from internet purchase vary across markets with the level of internet usage.

4.B. <u>Data Overview</u>

Table 2 presents the descriptive statistics of the variables used. The final data set consists of 523,618 observations from 150 major routes in US. Our data consists of two measures of internet usage. The first variable, online, is a dummy variable that takes a value of 1 if the transaction was completed over the internet. The second 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 58 percent. As expected, a higher share of purchases on the internet is found on leisure routes such as those to and from Orlando and Las Vegas.

The average load factor calculated at the time of purchase has a mean of 0.10. It needs to be emphasized that the average load factor is defined as the average load factor of the individual

[21] See Borenstein (1989) and Borenstein and Rose (1994) for further discussion.
[22] Recall from above that in our matched data all fares are available both online and offline.

segments for an itinerary at the time a ticket was purchased. Further, we calculate the load factor using both matched and unmatched observations in our transactions data. Since data of all transactions on a specific flight is not available our measured load factors only account for about 30 percent of transactions, corresponding to the share of transactions contained in our transactions data base. This means our load factors will be less than one. First class tickets on average comprise 8 percent of our data while business class and/or full coach class tickets account for 12 percent of our final data.[23]

Table 3 presents a simple hedonic regression of individual ticket prices on ticket restrictions and characteristics. The regressors include the ticket characteristics described above, the scarcity measures described, a hub variable, and route fixed effects. The results show that these variables explain about 80 percent of the variation in ticket prices.

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. Refundable tickets on average cost about 44 percent more than non-refundable ones. Advance purchase requirements generally reduce ticket prices, sometimes by more than 50 percent. These large advance purchase requirement effects 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. First class tickets cost about 70 percent more than ordinary tickets, and business class tickets cost about 32 percent more. The results also show that travel restrictions have a major effect on ticket prices, reducing

[23] These tickets do not include First class itineraries. The transaction data allows us to differentiate between first class and business or full *coach* class tickets. See Appendix for a detailed discussion on the same.

ticket prices by about 28 percent. Minimum and maximum stay requirements are also associated with slight reductions in ticket prices. Direct trips involving no change of planes cost roughly 6 percent more than itineraries involving a plane change while roundtrip tickets are 11 percent less expensive. An itinerary involving a Saturday night stay-over on average costs roughly 14 percent less than an itinerary with similar characteristics that does not.

The peaking variables are all highly statistically significant and of the right sign, but their economic magnitudes are small. Leaving at a peak time only increases average fares by less than one percent. Returns are most expensive on Sunday (not a peak day), and again the effects are small. Itineraries that involve a departure and/or return during peak times of day increase fares by .8 and 1.9 percent respectively.[24] The results also show that a 10 percent increase in the average load factor at the time of ticket purchase increase the fares on average by 3.0 percent. If one multiplies the mean of the variable times the coefficient, the results indicate that an increase in load factors at the time of purchase by one standard deviation increases fares by about two and a half percent. Overall, these peaking variable coefficients are economically small.

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

Turning to the effects of the internet, Table 4 presents the results testing our central hypothesis that internet purchases lead to lower prices. These regressions also introduce market structure variables for additional control. These variables and other route-specific variables could not be included in Table 3 because of the route fixed effects. Columns 1 and 2 of Table 4 present OLS and instrumental variable (IV) regressions of individual ticket prices on ticket characteristics, market structure variables and carrier fixed effects.[25] These columns test the

[24] These variables reflect peak travel times between 7:00-10:00 a.m. and 3:00-7:00 p.m., when there are generally high load factors on routes and throughout airline networks.
[25] IV regression involves instrumenting market share and HHI variables by their respective instruments.

hypothesis that online consumers pay lower prices on average for tickets, controlling for ticket characteristics, route, and carrier.

The coefficients of the remaining control variables are similar to those in Table 3. The coefficients on market structure variables are consistent with the literature. Our results show that an increase of 10 percent in the market share of a carrier increases prices by 2 percent on average. The coefficient on the Herfindahl Index is negative and statistically significant as found in several previous airline studies.[26] Fares on routes that involve a carriers' hub at either or both endpoints average 13 percent more than fares on routes not involving a carrier's hub.[27] The presence of an airport with restricted slots at either endpoint increases fares by about 14 percent.[28] On the other hand, the presence of a low cost carrier, other than Southwest, decreases average fares by roughly 10 percent, while Southwest's presence decreases average fares by 16 to19 percent. The results also suggest that an increase in the temperature difference between the two endpoints on a route by 1 percent decreases fares by about 0.2 percent, which according to the literature captures a tourist effect. Distance between endpoints and the average per capita income at the endpoints increase average fares while increases in average population decrease average fares.[29]

Turning to the internet effects, the results shows that after controlling for ticket characteristics, market structure, and carrier fixed effects online customers pay about 13 percent less than offline customers. Hence the results show very substantial effects of internet purchase

[26] This results, although not in keeping with common expectation, is consistent with the literature for empirical studies examining the determinants of profitability where both market share and market concentration are included as explanatory variables. In airlines see Borenstein (1989) and Stavins (2001). For a more general discussion on negative relationship between market concentration and prices see Ravenscraft (1983), Mueller (1986) and GAO (1991). Also see Appendix A.

[27] We included separate dummy variables to control for origin and destination hub airports for the operating carrier individually. Results show that the presence of a hub at the origin airport for the operating carrier increases prices by 11 percent while the presence of a hub at the destination airport increase prices by 9 percent on average.

[28] JFK (New York), La-Guardia (New York) and DCA (Washington D.C) are the three airports which still have restricted slots. Chicago O' Hare (ORD) was included in this list until 2002.

[29] See appendix for further discussion on this variable.

on ticket prices. These results, which control for most sources of variation in ticket characteristics and market structure, are both economically and statistically significant.[30]

The regressions in Table 4 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.[31] The Appendix, however, contains a similar estimation with a full set of route effects. Such estimation requires omission of the market structure variables and other variables that only vary across markets, including these route fixed effects results in a direct effect of internet purchase is to reduce fares by 11.4 percent.

The apparent reason for these observed price differences regards 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 methodology ensures that all of the online fares were available both online and offline. The regressions control for characteristics accounting for roughly 80 percent of the observed fare variation. Yet even with such control, internet purchasers are able to find fares discounted about 13 percent compared to those purchased through travel agents. Individual customers searching on the internet are able to and do make tradeoffs in ticket characteristics that enable them to pay lower prices. Hence customers buying online find better deals within the various buckets.

---

[30] We have also re-run 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.

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

The likely source of the higher offline fares consists either of communication or agency problems. While it appears that 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 tradeoff between price and convenience.

Columns (3) and (4) of Table 4 analyze the effect of the share of internet purchases on the average level of fares. In addition to the variables used in columns (1) and (2), as discussed above, we include an internet share variable representing the share of internet purchases on a given route. The internet share variable is large and highly significant. The results show that a ten percent increase in the share of tickets purchased on the internet for a route, which is approximately one standard deviation of that variable, decreases average fares on the route by about five percent. These results capture a spillover of internet shopping onto prices that is consistent with the empirical findings of Brown and Goolsbee (2002). Hence this specification shows that those who shop online receive direct benefits in terms of lower prices and also lower the prices of others who shop less.

This finding raises the issue of whether the returns to shopping are greater or smaller in markets where there are many shoppers. In Stahl's model, of course, shoppers obtain the competitive price because they know all prices. More generally, one can think of shoppers obtaining better prices than non-shoppers, but the precise returns would vary with the conditions of the market, the percentage of shoppers and the level of price dispersion.

Columns (5) and (6) of Table 4 present a partial analysis of this issue. The regressions in these columns investigate the relative magnitude of the savings of internet purchase as influenced by the share of total purchases made on the internet. More specifically, these regressions interact internet purchase with the online share variables. The results indicate that there are small but significant differences in the level of internet savings when there are more internet purchases. The results indicate that as the share of tickets purchased on the internet rises by 10 percent (roughly one standard deviation) online savings fall by about 0.3 percent. This result suggests that as the share of internet rises, the difference between the fares paid by the online and offline consumers falls modestly.[32] The effects of the other control variables remain similar to the earlier estimates. While these results are modest, they do suggest a link between price dispersion and internet usage. In particular, since internet fares are lower and since the fare differential shrinks with increased online purchases, dispersion should fall. More generally, as the percentage of shoppers rise, dispersion may either rise, fall, or stay the same. The issue of price dispersion is more fully investigated below.

To summarize, Table 4 provides three important results. First, controlling for ticket characteristics, capacity, route and carrier specific effects, online consumers on average pay roughly 13 percent less than the offline consumers. Second, equilibrium prices fall as the share of internet shoppers on a route rises. Finally, the results suggest that the potential savings of internet purchase in a lower internet usage market is marginally higher than the savings in higher internet usage markets.

[32] Though this effect may be surprising, it tends to provide a rationale for the decreased price dispersion. Increase in the share of internet usage implies more people buying on the internet such that a larger fraction faces similar prices such that the difference in the price paid on average will decrease. This is consistent with the theoretical prediction that as more informed consumers crowd a market, price dispersion will fall.

6. **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 non-monotonic relationship between online search and offline price dispersion for term life insurance. They attribute much of the non-linearity to the gradual maturation of internet markets.[33] 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 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.

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:

$$\mathbf{ln(Price_{ijtk})=\beta_0 + \beta_1 (R_{ijtk}) + \beta_2 (Loadfactor_{ijtk}) + \beta_3 (Route_j) + \beta_4 C_{ijtk} + \beta_5 (DEP_{ijtk}) + \beta_6 (RET_{ijtk}) + \varepsilon_{ijtk}} \qquad \mathbf{(2)}$$

where inclusion of $Route_j$ means we now include route fixed effects.[34]

[33] See Baliey (1998) and Brynjolfsson and Smith (1999) for implications of matured and immature internet markets.
[34] Refer to Table A1 for the estimated results of this equation.

From (2) we obtain the residuals and calculate different measures of dispersion (standard deviation, range and coefficient of variation) of the residuals by route-carrier for each departure date. These residuals parallel those of Brown and Goolsbee who measure residuals in life insurance after taking into account policy characteristics. These residuals represent unexplained price dispersion within a carrier-route after taking into account the observable characteristics of tickets. We then regress these various measures of dispersion measures on the share of tickets purchased on the internet and on market structure variables (Table 4).

Figure 6 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.

Table 5 presents a regression analysis of the relationship between price dispersion, internet share, and market structure variables. In column (1) of Table 5, we regress the standard deviation of the residuals on the share of internet usage, market share of carrier and Herfindahl index of the route, without correcting for the endogenity issues of the latter two variables.[35] In column (2) we run the same specification as in column (1) but we now correct for the potential endogeneity of market share and the Herfindahl index. The results show that price dispersion falls as the share of internet usage increases.

The relationship between price dispersion, market concentration and a carrier's share on a route parallels the results of Borenstein and Rose (1994). Using Gini coefficient as the measure of dispersion and without controlling for ticket characteristics, they find a 14 percent increase in dispersion from a one standard deviation (about 43 percent) decrease in the Herfindahl index. Their lack of control for ticket characteristics means that they measure dispersion with error,

[35] We 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.

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 role of ticket characteristics on prices.

We can use the results from Table 5 to investigate if the results of Borenstein and Rose (1994) vary when one controls for ticket characteristics. Using Table 5, a one standard deviation decrease in the market Herfindahl (about 53 percent) from the mean, leads to an increase in the standard deviation in prices about 7 percent. These results fully control for observed ticket characteristics. Hence our results for the Herfindahl are consistent with the findings of Borenstein and Rose (1994) though the effects are diminished, owing either to our use of the standard deviation as a measure of dispersion or our control for observable ticket characteristics.[36]

Turning to market share, Table 5 shows that a one standard deviation of a carrier's market share (about 54 percent) from its mean increases the standard deviation in prices by about 6 percent. This effect is much stronger when compared to Borenstein and Rose (1994) who find a positive but insignificant effect of increased carrier's share on dispersion. Hence our results are broadly consistent with those of Borenstein and Rose.

The results suggest that as the share of internet usage rises, price dispersion falls. To check for non-linearity between dispersion and the share of internet usage we include the square and the cube of the internet usage share. Column (3) does support a non-linear relationship, consistent with the theoretical predictions of Stahl (1989) and the empirical results of Brown and Goolsbee (2002). In column (4) we include the carrier specific effects and the carrier-departure date fixed effects respectively in addition to the variables used in column (2), correcting for the

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

potential endogeneity in the market structure variables. This too suggests a negative relationship between dispersion and internet share usage. More importantly, our result suggesting an inverse relationship between price dispersion and share of internet usage is robust to these alternative specifications. Thus, we can conclude that rise in the share of internet purchases on a route decreases the overall dispersion in the fares.

7. Implications for Consumer Surplus

In this section we provide some estimates of the overall magnitude of the savings the internet generates for customers. The analysis in Table 4 holds ticket characteristics constant. The internet, however, also helps customers identify the restrictions they are willing to accept by providing better information about the price/restriction tradeoff. Hence one can exclude the ticket characteristic regressors to identify the overall savings experienced by customers who buy on the internet. This approach likely overstates internet savings because it ignores potential selection bias in that internet customers may have found cheaper fares offline if the internet were unavailable. 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 dramatic impact of controlling for ticket characteristics when investigating price savings.

Column (1) of Table 6 presents a simple regression of the ticket prices and the online dummy variable along with return and departure day of the week, carrier and route fixed effects. In column (2) of Table 6 we run a similar specification without the departure and return day of the week. Both specifications suggest that online consumers on average save about 44 percent on ticket prices as compared to offline consumers.[37] This suggests that an offline consumer looking

[37] Note 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 online dummy variable as our dependent variable and the ticket characteristics as the explanatory variables. We did find a strong evidence that ticket characteristics associated with lower prices, namely Saturday

for a low price ticket and willing to accept restrictions saves very substantial sums compared to customers buying through offline outlets.

Table 6 also provides additional analysis to investigate whether our matching protocol has influenced estimated savings on the internet. Columns 3 and 4 of Table 6 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 from about forty-seven percent to about fifty-one percent. Hence the results suggest that omission of the unmatched observations may lead to a modest underestimation of the price savings from the internet.

8. **Conclusion**

This paper has provided 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 data set ever used to analyze airline pricing. This novel data set 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 only 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, carrier and route effects the results show that online prices are about 13 percent less than the offline prices. The analysis also shows that a ten percent increase in the share of tickets sold online decreases average prices

stay-over, travel restriction, travel during non-peak times, advance purchase requirements are more likely to be bough online than offline. However, tickets associated with minimum and maximum stay restrictions were found to be less likely to be bought on the internet.

by 5 percent, with more of its effect coming in the forms of lower offline prices. The paper also finds evidence 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 largely confirm the finding of Borenstein and Rose. The results show that even when one controls for ticket characteristics and peaking variables, the impact of market share and market concentration on the level and dispersion of ticket prices is qualitatively similar to that found by Borenstein and Rose.

References:

Bailey, Joseph (1998), "Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce," Ph.D. Dissertation, Technology, Management and Policy, Massachusetts Institute of Technology, Cambridge, MA.

Bakos, Yannis (1997), "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science,* 43 (12): 1676-1692.

Baye, Michael R. and John Morgan (2001), "Information gatekeepers on the Internet and the competitiveness of homogeneous product markets," The American Economic Review, 91 (3): 454-474.

Baye, Michael R. and John Morgan (2004), "Price Dispersion in the Lab and on the Internet: Theory and Evidence," Rand Journal of Economics, 35(3): 448-466

Baye, Michael R., John Morgan, and Patrick Scholten (2004), "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics*, 52(4):463-496

Baye, Michael R., John Morgan, and Patrick Scholten (2003), "The Value of Information in an Online Consumer Market," Journal of Public Policy and Marketing, 22 (Spring): 17-25.

Baye, Michael R., John Morgan, and Patrick Scholten (2004), "Persistent Price Dispersion in Online Markets," In: D. Jansen (Eds.), The New Economy. University of Chicago Press.

Borenstein. S (1989), "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry", Rand Journal of Economics, 20 (3):344-365

Borenstein. S and Rose.N (1994), "Competition and Price Dispersion in the U.S. Airline Industry", Journal of Political Economy, 102(4):653-683

Brown, Jeffery 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 Smith (2000), "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science,* 46 (4): 563-585.

Burdett, Kenneth and Kenneth Judd (1983), "Equilibrium Price Dispersion," *Econometrica,* 51 (July): 955-969.

Chevalier, Judith and Austan Goolsbee. 2003. "Measuring Prices and Price Competition Online: Amazon vs. Barnes and Noble." Quantitative Marketing and Economics. 1, pp203-222.

Chen, Jihui (2006), " Differences in Average Prices on the Internet: Evidence from the Online Market for Air Travel", Economic Inquiry, 44(4): 656-670.

Clay, Karen, Ramayya Krishnan, and Eric Wolff (2001), “Prices and Price Dispersion on the Web: Evidence from the Online Book Industry,” *The Journal of Industrial Economics*, 49 (4): 521-539.

Clemons, Eric, Il-Horn Hann, and Lorin Hitt (2002), “Price Dispersion and Differentiation in On-Line Travel: An Empirical Investigation,” Management Science, 48(4), 534-549.

Ellison, Glenn and Sara Fisher Ellison (2004), “Search, Obfuscation, and Price Elasticities on the Internet,” *Working Paper*, Sloan School of Management, MIT.

Lee, Darin and Prado, Maria (2005): “The Impact of Passenger Mix on Reported “Hub Premiums” in the US Airline Industry”, Southern Economic Journal, 72 (2): 372-394.

Lee, Ho Geun. 1998. “Do Electronic Marketplaces Lower the Price of Goods?.” Communications of the ACM. 41:1, pp. 73-80.

Lee, Zoonky and Sanjay Gosain (2002), “A Longitudinal price Comparison for Music CDs in Electronics and Brick-and-Mortar Markets: Pricing Strategies in Emergent Electronic Commerce,” *Journal of Business Strategies,* 19 (1): 55-71.

Morton, Fiona Scott, Florian Zettelmeyer, and Jorge Silva-Risso (2001), “Internet Car Retailing,” *Journal of Industrial Economics*, 49 (4): 501-519.

Morton, Fiona Scott, Florian Zettelmeyer, and Jorge Silva Risso. 2005. “Cowboys or Cowards: Why are Internet Car Prices Lower?” mimeo, University of California-Berkeley.

Morton, Fiona Scott, Florian Zettelmeyer, and Jorge Silva Risso. 2003. “Consumer Information and Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities?” Quantitative Marketing and Economics. 1:1, pp.65-92.

Pan, Xing, Brian T. Ratchford, Venkatesh Shankar (2002), “Can Price Dispersion in Online Markets be Explained by Differences in e-tailer Service Quality?” *Journal of the Academy of Marketing Science,* 30(4): 443-456.

Pan, Xing, Brian T. Ratchford, and Venkatesh Shankar (2003a), “Why Aren’t the Prices of the Same Item the Same at Me.com and You.com?: Drivers of Price Dispersion Among E-Tailers,” Working Paper, University of Maryland, College Park, MD 20742. July 2003.

Pan, Xing, Brian T. Ratchford and Venkatesh Shankar (2003b), “A Model of Retail Competition in Service and Price: Pure Play Internet vs. Bricks-and-Mortar Retailer.” Working Paper, Indiana University, Bloomington, IN 47405.

Pan, Xing, Brian T. Ratchford and Venkatesh Shankar (2004), "Price Dispersion on the Internet: A Review and Directions for Future Research." *Journal of Interactive Marketing,* 18 (4): 116-135

Salop, Steven and Joseph Stiglitz (1977), "Bargains and Rip-offs: A Model of Monopolistically Competitive Price Dispersion," *The Review of Economic Studies,* 44 (3): 493-510.

Salop, Steven and Joseph Stiglitz (1982), "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents," *The American Economic Review,* 72 (December): 1121-1130.

Scholten, Patrick A. and Adam Smith (2002), "Price Dispersion then and Now: Evidence from Retail and E-Tail Markets," *Advances in Microeconomics*: *Economics of the Internet and e-Commerce*, Vol. 11, 63-88.

Smith. B, Rao. B and Ratliff. R (2001), "E-Commerce and Operations Research in Airline Planning, Marketing, and Distribution", *Interfaces,* 31(2): 37-55

Smith, Michael and Erik Brynjolfsson (2001), "Customer Decision-Making at an Internet Shopbot: Brand Still Matters," *Quarterly Journal of Economics*, 49(4): 541-558.

Smith, Michael, Joseph Bailey, and Erik Brynjolfsson (2000), "Understanding Digital Markets: Review and Assessment," In Erik Brynjolfsson and Brian Kahin, editors, *Understanding the Digital Economy*. MIT Press, Cambridge, MA.

Sorensen, Alan (2000), "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs,"*Journal of Political Economy,* 108(4): 833-850.

Stahl, D (1989), "Oligopolistic Pricing with Sequential Consumer Search", American Economic Review, 79, September 1989: 700-712.

Stahl, D (1996), "Oligopolistic Pricing with Sequential Consumer Search and Heterogeneous Search Costs," October, International Journal of Industrial Organization, 14, 1996: 243–268.

Varian, Hal R. (1980), "A Model of Sales," *American Economic Review,* 70 (4): 651-659.

Varian, Hal R. (2000), "Market Structure in the Network Age," In Erik Brynjolfsson and Brian Kahin, editors, *Understanding the Digital Economy*. MIT Press, Cambridge, MA.

Appendix A:

Data description, construction of variables and expected effects:

We use a census of transactions by one of the major computer reservations systems (CRSs) for the fourth quarter of 2004. This CRS handles transactions for three of the major channels of ticket purchase – airlines, travel agents and several online sites. These data comprise approximately 30 percent of all transactions. For each ticket's itinerary sold through this CRS, we have information on the origin and destination, airline carrier, the flight number of each leg of the trip, the date of purchase, date of departure and return, the booking class, and whether the ticket was purchased online or offline.

Following Borenstein (1989) and Borenstein and Rose (1994), we include itineraries with at most one stop over in either direction. The prices used are for roundtrip fares, doubling the fares for one-way tickets to obtain comparability. We exclude itineraries with open-jaws and circular trip tickets. For the preparation of the data set used in this analysis, we also exclude tickets that included travel on more than one airline. Tickets indicating flights of which there is no record in the service segment data were also excluded from the analysis. Some flight numbers were not listed with the Official Airline Guide (OAG) and hence transactions involving travel on those flights had to be excluded from the analysis.[38] This study includes tickets which are operated by American Airlines, Continental, Delta, Northwest, US Airways, United Airlines, Frontier, Air Tran, Spirit, Alaska, American Mid-west, Hawaiian Airlines, Sun Country, Frontier Airlines and American Trans Air.

We analyze tickets for travel in the fourth quarter of 2004 with the exception of certain atypical travel times. Our period includes some of the peak travel period, particularly Thanksgiving, Christmas and New Year. To sidestep the problems of pricing during these peak travel periods, we dropped transactions for travel during the Thanksgiving week.[39] We also kept transaction which included departure and return within the 22nd of December, 2004. Thus we do not include itineraries involving travel during the last week of the year, since we believe that pricing could be different for these periods.

[38] OAG data consisted of operating flights only. There could be segments in the transaction data which were booked on flights which were later canceled and the flight numbers were changed for which there is no record.
[39] This includes the Wednesday prior to Thanksgiving till the following Monday.

Our transactions data do not include ticketing restrictions that may impact travelers' utility of a particular ticket[40]. In particular, the transaction data do not include information on ticket refundability, advance purchase restrictions, valid travel days or stay restrictions. Theoretical and empirical work (Dana (1999), Gale and Holmes (1993), Stavins (2001)) has shown that these ticketing restrictions may allow the airlines to separate travelers into different customer classes and charge different prices. However, we are able to merge our data to information on ticket-level travel restrictions. Travel agents' computer systems can access historical data on posted prices for one year. We collected additional data on restrictions from a local travel agent's CRS. The historical archive contains a list of fares/restrictions that were ever offered for travel on a specified carrier-city-pair-departure date. For each archived fare, we collected information on carrier, origin and destination, departure date from origin, fare, booking class (e.g. first class or coach), advance purchase requirement, refundability, travel restriction (e.g. travel can only occur on Tuesday through Thursday), and minimum or maximum stay restriction. We merged these data to the transaction data by carrier, fare, booking class and a variety of ticket characteristics.

The merging of these data used the following matching procedure. In the first step, we matched a ticket from the transaction data to a posted fare using carrier, date of departure (but not return), booking and coach class, and price. The prices were required to match within a 2 percent range and also on the day of travel and the carrier[41].

After this first step of the matching procedure, the resulting dataset included multiple matching posted fares for some individual transactions. This primarily included multiple matching fares with different combinations of advance purchase requirements and travel restrictions. Because our transaction data include no additional information to facilitate matching, we were required to make additional matching assumptions. In the second step of the matching procedure, we eliminate multiple matches based on advance purchase requirement. We assume that the ticket was purchased with the most restrictive advance purchase requirement for which it qualified. For example, suppose a ticket was purchased 16

[40] For confidentiality reasons, the CRS did not provide us with the full fare basis code.
[41] The data collected from the local travel agent belongs to a different computer reservation system. However the airline distribution was de-regulated in 2004. Since then, the airlines are free to provide different fares to any distribution channel including the major CRS's, their own CRS, own web-site and online travel agencies like Expedia. This necessitated the adoption of the matching rule(s) as discussed in the paper.

days before departure. If the first step matched both a 14 day and a 3 day advance purchase requirement, we match the transaction with the posted fare that required a 14 day advance purchase.

For any transactions that still matched multiple posted fares, we adopted a third matching step. In this step, we match on travel restrictions that involve travel on specified days of the week. For example, some posted fares required travel on a Tuesday, Wednesday or Thursday. Using the ticket's date of departure, we eliminated any multiple matches that did not satisfy the posted travel restriction. For any additional transactions with multiple matches, we assumed that any ticket meeting a travel restriction had that travel restriction. For example, a ticket matching fares with and without a travel restriction was assumed to have that travel restriction.

The final step includes the verification of minimum and maximum stay restrictions included in the travel agent data with our transaction data set. For the minimum and maximum stay restrictions collected from the travel agent, some restrictions were explicitly given (namely 1 day, 2 days etc) or were indicated to include a travel restriction but not specifically given and required further access to probe into the exact days of stay requirements. Unfortunately, we were not able to collect data on the specific days of stay restrictions (particularly minimum days stay required) that were required. For the matches, where the minimum and maximum days of stay restriction were given, we verified, if the actual transactions met the specific requirements. In case of multiple matches (which was a very small fraction, less than 1%) we assumed that if two tickets, had the same characteristics but one required a 1 day minimum stay while the other did not, and the transaction involved a 2 day stay, we assumed the ticket that required a 1 day minimum stay to be a relevant match for this observation as compared to the ticket which did not require a minimum stay restriction.

As we state above, we matched transactions to posted fares if the posted prices fell within 2 percent above or below the transaction price. This rule was adopted to overcome the problem of a sub-set of posted prices that we could collect from the travel agent's CRS system. We are thereby assuming, that for a ticket priced at $150 will be qualitative similar to one priced at $147 or $153. We however, took full precaution that the other matching criteria as discussed above.

The final data includes 523618 observations which include 150 different city-pairs. The complete list of the variables used in the analysis is discussed as follows:

Non-refundable tickets: equals 1 if the ticket is non-refundable else 0 (sign: negative)

Advance purchase requirement: the advance purchase requirement required on a ticket. This is usually 1, 3, 5, 7, 10, 14, 21 and 30 (sign: negative)

Days prior to departure ticket is purchased: the number of days before departure the ticket was bought. Given our data, we noticed some really low fares very close to the departure date; even on the day of departure while some really high fares more than a month in advance. So the relationship between days prior to departure and price paid is indeterminate. However, Stavins (2001) shows a strictly negative relationship between the prices paid and the days before the ticket was purchased. (sign: ?)

Online: equal to 1 if the ticket was bought online and 0 if purchased offline (sign: negative)
Direct Flight: equal to 1 if the itinerary did not involve a change of plane or stop-over and 0 otherwise. This is an indicator for non-stop itinerary (sign: positive)

Roundtrip Ticket: equal to 1 if the itinerary was for a roundtrip travel and 0 otherwise (sign: negative)

Saturday Stay-over: equal to 1 if the itinerary involved a Saturday stay over and 0 otherwise. This was created by using the departure and the return date indicated in the transaction data (sign: negative)

First class: equal to 1 if any segment of the itinerary involved a travel in the first class coach and 0 otherwise (sign: positive)

Business class: equal to 1 if any segment of the itinerary involved a travel in full coach fare class or business class (not first class) and 0 otherwise. This variable is required to differentiate between the regular coach class tickets and business or full coach fare class tickets, apart from the first class tickets. The transaction data provides us information regarding the booking and cabin class of the transaction, which helps us to identify this distinction. (sign: positive)

Travel restriction requirement: equal to 1 if the ticket required a travel day restriction and 0 otherwise. This primarily requires, that the ticket bought (price) is valid only if the individual travels during certain days of the week, say, Tuesday or Thursday (sign: negative)

Minimum stay restriction: equal to 1 if the ticket required a minimum stay requirement and 0 otherwise (sign: negative)

Maximum stay restriction: equal to 1 if the ticket required a maximum stay requirement and 0 otherwise (sign: negative)

Distance: non-stop mileage between the two endpoint airports on a route (sign: positive)

Temperature Difference: the absolute difference in the average January temperature between the origin and destination of the individual itinerary (sign: negative)

Hub: equal to 1 if any endpoint airport of the route is a hub airport for the operating carrier (sign: positive)

Slots: equal to 1 if any endpoint airport has restricted slots. This includes La Guardia Airport in New York (LGA), Kennedy Airport in New York (JFK) and Washington National (DCA.).

Flight load factor at Purchase: this is the load factor averaged over each individual segment involved in the itinerary at the time when the ticket was purchased. We had information regarding the flight numbers for travel in each segment of an individual itinerary. We used this information along with data from Official Airline Guide (OAG) to calculate the total number of seats on each of these flights scheduled for departure on a particular date. From our transaction data, we calculated the total number of seats that were sold on that flight the day before an individual transaction for that flight took place. That is, for a ticket involving a travel on flight 66 on American Airlines from DFW-ORD on October 10, 2004, and being bought on October 9th, we calculate how many seats were sold of flight 66 departing on October 10th was sold till October 8th. Since we cannot, observe the order of transaction taking place on the same day (October 9th) we assume that all tickets being bought October 9th for October 10th flight will face the same load factor as of October 8th. This is the closest approximation possible to calculate the contemporaneous load factor facing an individual ticket at the time of transaction. For all the segments involved in the itinerary, we average these load factors to calculate the average load factor associated with each transaction. We expect this variable to have a positive coefficient; since individuals will face a higher price if he travels on flights which are full (have a high demand) (sign: positive)

Departure Peak: equals to 1 if the individual itinerary involves departure at a peak time (between 7-10am or 3-7pm). Given the flight numbers, we use information from OAG to determine the local departure time (sign: positive)

Return Peak: equals to 1 if the individual itinerary involves return at a peak time (between 7-10am or 3-7pm). For one way tickets, this is equal to 0 (sign: positive)

Low Cost Route: takes a value equal to 1 if a low cost carrier (other than Southwest) operates on that route and 0 otherwise. The presence of the low cost carrier is expected to induce competition among operating airlines driving the prices down (sign: negative)

Southwest: takes a value equal to 1 if Southwest airlines operate on that route and 0 otherwise. It is well accepted that presence of Southwest airlines induces significant price reductions on the route owing to cost advantage of Southwest (sign: negative)

Population: the average population at the two endpoints of the route (Source: US Census 2003). On one hand, higher population between the endpoint airports of a route can create more demand such that price increases. On the other hand, airlines can have more flights on routes where there is more demand such that prices can be lower (sign :?)

Per Capita Income: the average per capita income at the two endpoints of the route (Source: US Census 2003) (sign: positive)

Market Share: calculated as the proportion of passengers accounted by a carrier on a route. T-100 segment data is used to calculate this share. If there is not a complete umbrella effect from the market power of a dominant firm, then holding the market concentration constant, an increase in market share is expected to increase the prices (sign: positive).

Herfindahl index (HHI): sum of the square of the market shares of each of the carriers operating on a route. To the extent that the dominant firm's high prices create an umbrella that allows a few firms in a concentrated market to collude easily, then increases in the concentration will increase the prices. If however, a dominant firm on a route has a competitive advantage owing to cost structure, advertising, marketing or other means, then it could possibly reduce the profit maximizing prices of the other firms (sign: ?)

Internet Share: share of all online transactions to the total transactions on a route (sign: negative)

Internet Share*Online: Internet share interacted with the online dummy (sign: ?)

GEOSHARE: given by ($\sqrt{}ENP_{x1}$. ENP_{x2})/ ($\sqrt{}ENP_{y1}$. ENP_{y2}) where y indexes all airlines, x the observed airline and ENP_{y1} and ENP_{y2} are airline y's average enplanements at the two enpoints airports during the fourth quarter of 2004.

XTHERF: is the square of the fitted value of for market share (from its first stage regression) plus the rescaled sum of the squared of all other carriers' share. This is given by:

xtherf = (predicted market share)^2 + [(HHI-market share^2)/(1-market share)^2]*(1-predcited market share)^2 . See discussion in Borenstein (1989) and/or Borenstein and Rose (1994).

Departure Day of Week: indicates the departure day of the week. It takes a value of 0 to 6, where 0 represents a Sunday and 6 a Saturday. In all estimates Sunday is treated as the base group. (sign: positive)

Return day of Week: indicates the return day of the week. It takes a value of 0 to 6, where 0 represents a Sunday and 6 a Saturday. In all estimates Sunday is treated as the base group. Also, for one way tickets, this takes a value of 0. (sign: negative)

Table A1
The Effects of Online When
Route Fixed Effects are Included

	lfare
Non-refundable	-0.438
	(269.87)**
Days prior to departure ticket purchased	0.0004
	(15.67)**
Advance Purchase Requirement:	
One day advance	-0.242
	(66.39)**
Three day advance	-0.045
	(27.17)**
Five day advance	-0.538
	(38.21)**
Seven day advance	-0.170
	(121.21)**
Ten day advance	-0.206
	(66.99)**
Fourteen day advance	-0.258
	(162.43)**
Twenty-one day advance	-0.301
	(91.78)**
Thirty day advance	-0.129
	(16.50)**
Direct	0.054
	(13.18)**
Online	-0.114
	(82.06)**
Roundtrip	-0.106
	(48.71)**
Saturday stay-over	-0.117
	(62.20)**
First class	0.699
	(362.90)**
Business class	0.315
	(192.49)**
Travel restriction	-0.275
	(264.44)**
Minimum stay requirement	-0.008
	(5.62)**
Maximum stay requirement	-0.046
	(30.75)**
HUB	0.023
	(8.50)**
Average load factor at purchase	0.294
	(43.73)**
Peak time of Day:	
Departure at peak time	0.008

Table A1
The Effects of Online When
Route Fixed Effects are Included

	(8.33)**
Return at peak time	0.017
Airline Fixed Effects (American Airlines omitted):	(15.16)**
Continental	0.001
	(0.480)
Delta	0.001
	(0.320)
Northwest	-0.012
	(3.60)**
United Airways	0.100
	(65.59)**
US Air	-0.031
	(10.82)**
Frontier	-0.228
	(56.56)**
Alaska	-0.042
	(4.44)**
Hawaiian Airlines	0.005
	(0.340)
America Mid-west	-0.219
	(54.25)**
American Trans Air	-0.207
	(14.45)**
Midwest Express	-0.040
	(2.68)**
Air Tran	-0.366
	(34.07)**
Spirit	-0.213
	(35.67)**
Sun County	-0.470
	(55.31)**
Departure Day of Week (Sunday omitted):	
Monday	0.001
	(0.770)
Tuesday	0.001
	(0.740)
Wednesday	0.002
	(1.210)
Thursday	0.008
	(4.51)**
Friday	0.015
	(7.51)**
Saturday	-0.064
	(30.22)**

Table A1
The Effects of Online When
Route Fixed Effects are Included

Return day of week (Sunday omitted):	
Monday	-0.032
	(15.00)**
Tuesday	-0.026
	(11.49)**
Wednesday	-0.034
	(15.04)**
Thursday	-0.025
	(11.27)**
Friday	-0.035
	(15.45)**
Saturday	-0.064
	(23.55)**
Constant	6.121
	(854.84)**
Route Effects	Yes
Observations	523618
R-squared	0.8

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

Table A2 List of City Pairs Used	
Routes	**Routes**
Atlanta (ATL)-Boston (BOS)	Chicago (ORD) – Orange County (SNA)
Atlanta (ATL)-Cincinnati (CVG)	Chicago (MDW) – Detroit (DTW)
Atlanta (ATL)- Fort Lauderdale (FLL)	Cleveland (CLE) – Chicago (MDW)
Atlanta (ATL)-Dulles, DC (IAD)	Cleveland (CLE)– Chicago O' Hare (ORD)
Atlanta (ATL)-Houston (IAH)	Cincinnati (CVG)–O' Hare (ORD)
Atlanta (ATL)-Los Angeles (LAX)	Columbus (CMH) – La Guardia (LGA)
Atlanta (ATL)-La Guardia (LGA)	Dallas (DFW) – Atlanta (ATL)
Atlanta (ATL)- Orlando (MCO)	Dallas (DFW) – Denver (DEN)
Atlanta (ATL)- Memphis (MEM)	Dallas (DFW) – Washington (IAD)
Atlanta (ATL) – Miami (MIA)	Dallas (DFW)- Houston (IAH)
Atlanta (ATL)-New Orleans (MSY)	Dallas (DFW) – Los Angeles (LAX)
Atlanta (ATL) – Chicago O' Hare (ORD)	Dallas (DFW) – Long Beach (LGB)
Atlanta (ATL)- Philadelphia (PHL)	Dallas (DFW) – Kansas City (MCI)
Atlanta (ATL)-Tampa (TPA)	Dallas (DFW) – Chicago (ORD)
Baltimore (BWI) – Atlanta (ATL)	Dallas (DFW) – Phoenix (PHX)
Baltimore (BWI) – Cleveland (CLE)	Denver (DEN) – Atlanta (ATL)
Baltimore (BWI) – Dallas (DFW)	Denver (DEN) – Boston (BOS)
Baltimore (BWI)- Fort Lauderdale (FLL)	Denver (DEN) – Washington (DCA)
Baltimore (BWI) – Los Angeles (LAX)	Denver (DEN) – Newark (EWR)
Baltimore (BWI)- Orlando (MCO)	Denver (DEN) – Houston (IAH)
Boston (BOS) – Baltimore (BWI)	Denver (DEN) – New York (LGA)
Boston (BOS)- Charlotte (CLT)	Denver (DEN) – Kansas City (MCI)
Boston (BOS)- Washington (DCA)	Denver (DEN) – Orlando (MCO)
Boston (BOS) – Dallas (DFW)	Denver (DEN) – Portland (PDX)
Boston (BOS) – Detroit (DTW)	Denver (DEN) – Philadelphia (PHL)
Boston (BOS) – Los Angeles (LAX)	Denver (DEN) – Phoenix (PHX)
Boston (BOS) – Philadelphia (PHL)	Denver (DEN) – St. Louis (STL)
Boston (BOS) – Pittsburgh (PIT)	Denver (DEN) – Tampa (TPA)
Boston (BOS) – Fort Myers (RSW)	Detroit (DTW) – Atlanta (ATL)
Boston (BOS) – Tampa (TPA)	Detroit (DTW) – Baltimore (BWI)
Charlotte (CLT) – Orlando (MCO)	Detroit (DTW) – Dallas (DFW)
Chicago (ORD) – Boston (BOS)	Detroit (DTW) – Newark (EWR)
Chicago (ORD) – Baltimore (BWI)	Detroit (DTW) – Fort Lauderdale (FLL)
Chicago (ORD) – Charlotte (CLT)	Detroit (DTW) – Las Vegas (LAS)
Chicago (ORD) – Denver (DEN)	Detroit (DTW) – Orlando (MCO)
Chicago (ORD) – Washington (IAD)	Detroit (DTW) – Chicago (ORD)
Chicago (ORD)- New York (LGA)	Fort Lauderdale (FLL) – Boston (BOS)
Chicago (ORD) – Miami (MIA)	Fort Lauderdale (FLL)- Chicago (ORD)
Chicago (ORD) – Minneapolis (MSP)	Hartford (BDL) – Washington (DCA)
Chicago (ORD) – New Orleans (MSY)	Hartford (BDL) – Chicago O' Hare (ORD)
Chicago (ORD) – Omaha (OMA)	Honolulu (HNL) – Los Angeles (LAX)
Chicago (ORD) – Ft. Myers (RSW)	Houston (IAH) – New Orleans (MSY)

Table A2 List of City Pairs Used	
Routes	**Routes**
Chicago (ORD) – San Diego (SAN)	Houston (IAH) – Chicago (ORD)
Routes	**Routes**
Las Vegas (LAS) – Burbank (BUR)	New York (LGA) – Cincinnati (CVG)
Las Vegas (LAS) – Los Angeles (LAX)	New York (LGA)- Dallas (DFW)
Las Vegas (LAS) – Chicago (ORD)	New York (LGA) – Detroit (DTW)
Long Beach (LGB) – Dallas (DFW)	New York (LGA)- Houston (IAH)
Los Angeles (LAX) – Denver (DEN)	New York (LGA) – Palm Beach, FL (PBI)
Los Angeles (LAX) – Detroit (DTW)	Oakland (OAK) – Denver (DEN)
Los Angeles (LAX) – Houston (IAH)	Oakland (OAK) – Seattle (SEA)
Los Angeles (LAX)- Miami (MIA)	Ontario (ONT) – Denver (DEN)
Los Angeles (LAX)- Chicago (ORD)	Orlando (MCO) – Washington (DCA)
Los Angeles (LAX) – Philadelphia (PHL)	Orlando (MCO) – Dallas (DFW)
Los Angeles (LAX) – Reno (RNO)	Orlando (MCO)- New York (LGA)
Los Angeles (LAX) – Tampa (TPA)	Palm Beach (PBI) – Boston (BOS)
Miami (MIA) – New York (LGA)	Philadelphia (PHL) – Chicago (ORD)
Miami (MIA) – Boston (BOS)	Philadelphia (PHL) – Palm Beach (PBI)
Miami (MIA)- Newark (EWR)	Phoenix (PHX) – Minneapolis (MSP)
Milwaukee (MKE) – Minneapolis (MSP)	Phoenix (PHX) – Ontario (ONT)
Minneapolis (MSP) – Denver (DEN)	Pittsburgh (PIT) – New York (LGA)
Minneapolis (MSP) – Dallas (DFW)	Pittsburgh (PIT) – Chicago (ORD)
Minneapolis (MSP) – Detroit (DTW)	Portland (PDX) – Las Vegas (LAX)
Minneapolis (MSP) – Los Angeles (LAX)	Portland (PDX) – Los Angeles (LAX)
Minneapolis (MSP) – New York (LGA)	Portland (PDX) – Oakland (OAK)
Minneapolis (MSP) – Chicago (MDW)	St. Louis (STL) – Los Angeles (LAX)
Newark (EWR) – Minneapolis (MSP)	Sacramento (SMF) – Los Angeles (LAX)
Newark (EWR) – Chicago (ORD)	Salt Lake City (SLC) – Denver (DEN)
Newark (EWR) – Atlanta (ATL)	San Francisco (SFO) – Boston (BOS)
Newark (EWR) – Boston (BOS)	San Francisco (SFO) – Dallas (DFW)
Newark (EWR) – Los Angeles (LAX)	San Jose (SJC) – Denver (DEN)
New Orleans (MSY) – New York (LGA)	Tampa (TPA) – New York (LGA)
New York (JFK) – Los Angeles (LAX)	Washington (DCA) – Atlanta (ATL)
New York (LGA) – Boston (BOS)	Washington (DCA) – Dallas (DFW)
New York (LGA) – Cleveland (CLE)	Washington (DCA)-La Guardia (LGA)
New York (LGA) – Charlotte (CLT)	Washington (DCA)- Chicago (ORD)

Table 1
Comparison of Online and Offline Mean Fares
Conditional on Ticket Characteristics
Various Routes

	Delta Atlanta-La Guardia		US Airways Boston-Dulles		Continental Los Angeles-Newark		American Airlines Boston-Chicago		American Airlines Los Angeles-JFK	
	Offline	Online	Offline	Online	Offline	Online	Offline	Online	Offline	Online
Ticket Characteristics:										
3-day advance	351.68	240.73	691.42	682.00	492.13	370.00	396.48	308.20	1549.47	553.79
7-day advance	264.77	233.47	384.08	242.27	307.24	283.11	269.34	198.23	336.70	264.41
14-day advance	191.28	185.20	300.20	239.93	293.76	292.18	292.06	202.78	361.54	283.05
Refundable	1365.12	426.00	435.44	394.46	1869.45	NA	1029.91	NA	1866.56	1523.33
Non-refundable	276.87	216.11	275.28	211.91	490.72	331.13	384.24	230.20	444.39	302.50
Saturday stay-over	239.18	202.29	294.63	225.13	478.96	299.67	268.07	205.13	643.37	283.46
Business/Full coach class	1105.44	NA	680.67	652.16	782.68	386.33	916.78	NA	1448.76	NA
First Class	1360.04	437.99	736.10	304.50	2128.44	NA	737.69	NA	2340.96	NA
Minimum stay	181.49	174.99	235.12	207.73	360.93	309.94	276.98	204.30	349.72	274.11
Maximum stay	181.48	175.03	274.65	211.61	440.84	412.30	281.11	204.56	421.27	347.28

	American Airlines Chicago-DFW		Continental Chicago-Newark		American Airlines Chicago-Philadelphia		American Airlines La Guardia-Chicago	
	Offline	Online	Offline	Online	Offline	Online	Offline	Online
Ticket Characteristics:								
3-day advance	587.15	1337.99	699.83	364.00	268.44	178.00	636.10	291.29
7-day advance	523.88	306.78	375.97	240.06	213.87	186.31	355.50	176.48
14-day advance	239.41	229.56	262.30	208.51	170.44	155.16	266.33	228.74
Refundable	1281.08	1223.99	1029.14	223.00	248.31	227.42	1182.55	NA
Non-refundable	561.74	308.04	383.18	235.11	199.76	158.70	407.30	219.59
Saturday stay-over	325.06	215.70	257.7	217.07	208.56	181.85	271.08	203.33
Business/Full coach class	849.48	1086.23	971.00	332.14	715.24	NA	1127.86	NA
First Class	1231.31	NA	1655.07	NA	620.55	628.00	1247.98	NA
Minimum stay	243.82	220.88	292.53	222.63	224.08	175.66	304.06	215.51
Maximum stay	235.50	210.61	287.83	230.67	248.67	172.10	308.96	221.24

Table 2
Descriptive Statistics

Variable Description	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Round-trip fare	441.711	433.549	57.67	4323.72
Non-refundable	0.756	0.429	0.000	1.000
Advance purchase requirement	5.428	6.149	0.000	30.000
Days prior to departure ticket purchased	15.514	20.050	0.000	202.000
Saturday stay-over	0.173	0.378	0.000	1.000
Travel restriction	0.398	0.489	0.000	1.000
Minimum stay requirement	0.218	0.413	0.000	1.000
Maximum stay requirement	0.190	0.392	0.000	1.000
First class	0.080	0.272	0.000	1.000
Business or Full coach fare class	0.122	0.327	0.000	1.000
Online	0.123	0.329	0.000	1.000
Internet share	0.184	0.096	0.024	0.587
Direct flight	0.989	0.103	0.000	1.000
Roundtrip	0.728	0.444	0.000	1.000
Load factor at time of purchase	0.109	0.081	0.003	0.880
Absolute temperature difference	15.217	10.893	0.001	46.000
HUB	0.742	0.437	0.000	1.000
Slot constrained airport	0.263	0.440	0.000	1.000
Departure at peak time	0.296	0.456	0.000	1.000
Return at peak time	0.215	0.410	0.000	1.000
Low cost carrier on route	0.473	0.499	0.000	1.000
Southwest Airlines	0.061	0.240	0.000	1.000
Distance	957.551	632.932	185.000	2704.000
Average population*	1986562	1593612	233014.6	5974809
Average per capita income*	36530.760	3321.425	23808.000	45046.490
Market share	0.546	0.254	0.000	1.000
HHI	0.538	0.191	0.189	1.000

*In thousands

Source: Ticket characteristic and fare data comes from one of the major CRS.
Market level data including market share and HHI is calculated using Department of Transportation's T-100 database.
List of hub airports is gathered from Air Traveler's
(http://www.faqs.org/faqs/travel/air/handbook/part2/section-13.html)
For data descriptions please refer to the Appendix.

Table 3
Regression of Natural Logarithms of Fares on Ticket Characteristics

	Log(Fare)
Advance Purchase Requirement:	
One day advance	-0.249
	(67.92)**
Three day advance	-0.044
	(26.50)**
Five day advance	-0.554
	(39.11)**
Seven day advance	-0.170
	(120.47)**
Ten day advance	-0.207
	(66.94)**
Fourteen day advance	-0.260
	(162.27)**
Twenty-one day advance	-0.310
	(94.11)**
Thirty day advance	-0.128
	(16.24)**
Other Characteristics:	
Non-refundable ticket	-0.439
	(268.53)**
Days prior to departure ticket purchased	0.000
	(10.66)**
Saturday stay-over	-0.138
	(73.69)**
Travel restriction requirement	-0.280
	(268.34)**
Minimum stay requirement	-0.007
	(4.42)**
Maximum stay requirement	-0.047
	(30.63)**
First class	0.706
	(364.62)**
Business class	0.320
	(194.34)**
Direct flight	0.060

Table 3
Regression of Natural Logarithms of Fares
on Ticket Characteristics

	(17.90)**
Roundtrip ticket	-0.110
	(50.05)**
Flight load factor at purchase	0.299
	(44.08)**
HUB	0.025
	(9.38)**
Peak Time of Day:	
Departure at peak time	0.008
	(8.61)**
Return at peak time	0.019
	(16.37)**
Departure Day of the Week (Sunday omitted):	
Monday	0.005
	(3.22)**
Tuesday	0.003
	(1.70)
Wednesday	0.006
	(3.39)**
Thursday	0.011
	(6.01)**
Friday	0.016
	(7.98)**
Saturday	-0.073
	(34.43)**
Return Day of the Week (Sunday omitted):	
Monday	-0.035
	(16.27)**
Tuesday	-0.024
	(10.61)**
Wednesday	-0.029
	(12.95)**
Thursday	-0.021
	(9.06)**
Friday	-0.030
	(13.51)**
Saturday	-0.071
	(26.29)**

Table 3
Regression of Natural Logarithms of Fares on Ticket Characteristics

Carrier Fixed Effects (American Airlines omitted):

Continental	-0.003
	(0.910)
Delta	-0.006
	(2.50)*
Northwest	-0.018
	(5.19)**
United Airways	0.095
	(62.38)**
US Air	-0.029
	(10.14)**
Frontier Airlines	-0.254
	(62.99)**
Alaska	-0.046
	(4.91)**
Hawaiian Airlines	0.009
	(0.610)
America West	-0.229
	(56.48)**
American Trans Air	-0.229
	(15.85)**
Midwest Express Airlines	-0.049
	(3.27)**
Air Tran	-0.357
	(33.00)**
Spirit	-0.252
	(42.13)**
Sun County	-0.494
	(57.77)**
Route effects	Yes
Constant	6.098
	(846.86)**
Observations	523618
R-squared	0.790

Note: Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%
Source: See Table 1.

Table 4
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets

	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
Market Structure Variables:						
Market share	0.117	0.286	0.120	0.291	0.120	0.291
	(32.74)**	(27.42)**	(33.57)**	(28.02)**	(33.42)**	(27.93)**
HHI	-0.005	-0.251	-0.023	-0.256	-0.022	-0.256
	(1.030)	(22.20)**	(5.13)**	(22.71)**	(5.02)**	(22.62)**
HUB	0.137	0.122	0.131	0.115	0.131	0.115
	(85.76)**	(61.11)**	(82.61)**	(57.67)**	(82.60)**	(57.67)**
Slot constrained airport	0.139	0.142	0.138	0.139	0.138	0.139
	(79.95)**	(77.86)**	(79.85)**	(76.66)**	(79.80)**	(76.65)**
Internet Variables:						
Online	-0.133	-0.128	-0.129	-0.124	-0.141	-0.131
	(79.03)**	(75.68)**	(76.84)**	(73.35)**	(38.40)**	(35.13)**
Internet share			-0.454	-0.479	-0.463	-0.484
			(55.39)**	(57.91)**	(54.02)**	(56.01)**
Internet share*Online					0.055	0.029
					(3.78)**	(1.970)*
Advance Purchase Requirement:						
(No advance purchase required omitted)						
One day advance	-0.387	-0.382	-0.370	-0.366	-0.370	-0.366
	(93.25)**	(91.45)**	(89.32)**	(87.69)**	(89.28)**	(87.67)**
Three day advance	0.052	0.056	0.056	0.060	0.056	0.060
	(27.79)**	(29.69)**	(29.92)**	(31.71)**	(29.99)**	(31.74)**
Five day advance	-0.705	-0.706	-0.665	-0.665	-0.665	-0.665
	(41.98)**	(41.86)**	(39.67)**	(39.54)**	(39.70)**	(39.55)**
Seven day advance	-0.134	-0.133	-0.132	-0.132	-0.132	-0.132
	(81.34)**	(80.75)**	(80.76)**	(80.31)**	(80.67)**	(80.26)**
Ten day advance	-0.245	-0.241	-0.244	-0.241	-0.243	-0.241
	(67.50)**	(66.18)**	(67.29)**	(66.22)**	(67.21)**	(66.18)**
Fourteen day advance	-0.193	-0.189	-0.196	-0.193	-0.196	-0.193
	(104.34)**	(101.88)**	(106.24)**	(104.24)**	(106.12)**	(104.18)**
Twenty-one day advance	-0.245	-0.245	-0.247	-0.247	-0.246	-0.247
	(63.71)**	(63.54)**	(64.40)**	(64.37)**	(64.27)**	(64.29)**
Thirty day advance	0.060	0.058	0.059	0.057	0.059	0.058
	(6.51)**	(6.29)**	(6.37)**	(6.21)**	(6.44)**	(6.24)**
Other Ticket Characteristics:						
Non-refundable ticket	-0.352	-0.356	-0.350	-0.354	-0.350	-0.354
	(202.33)**	(203.51)**	(202.22)**	(203.06)**	(202.16)**	(203.00)**
Days prior to departure ticket purchased	-0.0002	-0.0002	-0.0001	-0.0001	-0.0001	-0.0001
	(5.96)**	(7.27)**	(1.910)***	(2.95)**	(2.21)*	(3.10)**
Saturday stay-over	-0.129	-0.130	-0.127	-0.127	-0.127	-0.127

Table 4
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets

	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
	(57.37)**	(57.26)**	(56.32)**	(56.19)**	(56.40)**	(56.23)**
Travel restriction requirement	-0.284	-0.285	-0.280	-0.281	-0.280	-0.280
	(237.34)**	(236.31)**	(234.43)**	(232.92)**	(234.41)**	(232.92)**
Minimum stay requirement	0.010	0.014	0.008	0.011	0.008	0.011
	(5.94)**	(8.16)**	(4.82)**	(6.79)**	(4.85)**	(6.80)**
Maximum stay requirement	-0.012	-0.009	-0.010	-0.007	-0.010	-0.006
	(6.96)**	(5.01)**	(5.61)**	(3.76)**	(5.58)**	(3.75)**
First class	0.787	0.784	0.779	0.776	0.779	0.776
	(353.22)**	(349.29)**	(350.02)**	(345.81)**	(349.83)**	(345.74)**
Business class	0.425	0.435	0.432	0.440	0.432	0.440
	(240.08)**	(240.27)**	(244.10)**	(243.37)**	(244.09)**	(243.35)**
Remaining Ticket Characteristics:						
Direct flight	0.030	-0.001	0.025	-0.007	0.025	-0.007
	(6.09)**	(0.270)	(5.00)**	(1.300)	(5.05)**	(1.270)
Roundtrip ticket	-0.109	-0.110	-0.112	-0.114	-0.112	-0.113
	(41.32)**	(41.54)**	(42.57)**	(43.01)**	(42.46)**	(42.94)**
Flight load factor at purchase	0.163	0.144	0.148	0.128	0.147	0.128
	(20.56)**	(17.95)**	(18.70)**	(16.01)**	(18.51)**	(15.91)**
Peak Time of Day:						
Departure at peak time	0.016	0.018	0.016	0.018	0.016	0.018
	(14.15)**	(15.11)**	(14.23)**	(15.18)**	(14.24)**	(15.18)**
Return at peak time	0.027	0.028	0.027	0.027	0.027	0.027
	(19.74)**	(20.42)**	(19.51)**	(20.17)**	(19.51)**	(20.16)**
Other Route Specific Characteristics:						
Low cost carrier on route	-0.102	-0.102	-0.083	-0.083	-0.083	-0.083
	(85.87)**	(85.76)**	(67.59)**	(66.95)**	(67.37)**	(66.80)**
Southwest Airlines	-0.193	-0.185	-0.170	-0.162	-0.170	-0.162
	(77.72)**	(74.01)**	(67.71)**	(63.83)**	(67.81)**	(63.84)**
Other Route Level Variables:						
Distance (log)	0.356	0.352	0.385	0.381	0.385	0.381
	(351.86)**	(312.54)**	(339.00)**	(309.18)**	(338.75)**	(308.47)**
Absolute Temperature	-0.002	-0.002	-0.001	-0.002	-0.001	-0.002
Difference (Log)	(5.19)**	(5.65)**	(3.27)**	(3.49)**	(3.27)**	(3.49)**
Average population (Log)	-0.029	-0.030	-0.038	-0.039	-0.038	-0.039
	(35.45)**	(35.69)**	(45.73)**	(44.95)**	(45.69)**	(44.93)**
Average per capita	0.138	0.142	0.058	0.058	0.057	0.058
Income (Log)	(20.72)**	(21.25)**	(8.47)**	(8.49)**	(8.41)**	(8.46)**
Departure Day of Week (Sunday omitted):						
Monday	0.012	0.012	0.012	0.012	0.012	0.012
	(6.12)**	(6.27)**	(6.22)**	(6.37)**	(6.22)**	(6.37)**
Tuesday	0.018	0.018	0.017	0.018	0.017	0.018

Table 4
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets

	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
	(8.76)**	(9.06)**	(8.62)**	(8.87)**	(8.61)**	(8.87)**
Wednesday	0.022	0.023	0.022	0.023	0.022	0.023
	(10.37)**	(10.80)**	(10.38)**	(10.75)**	(10.35)**	(10.74)**
Thursday	0.023	0.024	0.023	0.024	0.023	0.024
	(10.35)**	(10.95)**	(10.44)**	(10.98)**	(10.42)**	(10.97)**
Friday	0.028	0.029	0.028	0.029	0.028	0.029
	(12.01)**	(12.30)**	(11.85)**	(12.05)**	(11.89)**	(12.06)**
Saturday	-0.055	-0.054	-0.054	-0.054	-0.054	-0.054
	(21.28)**	(21.14)**	(21.24)**	(21.19)**	(21.25)**	(21.19)**
Return Day of the Week (Sunday omitted):						
Monday	-0.035	-0.035	-0.034	-0.034	-0.035	-0.034
	(13.51)**	(13.49)**	(13.38)**	(13.34)**	(13.50)**	(13.40)**
Tuesday	-0.022	-0.021	-0.021	-0.021	-0.021	-0.021
	(8.00)**	(7.89)**	(7.80)**	(7.66)**	(7.96)**	(7.74)**
Wednesday	-0.031	-0.031	-0.030	-0.030	-0.030	-0.030
	(11.54)**	(11.44)**	(11.08)**	(10.93)**	(11.23)**	(11.00)**
Thursday	-0.022	-0.023	-0.021	-0.021	-0.021	-0.021
	(8.04)**	(8.22)**	(7.61)**	(7.72)**	(7.75)**	(7.79)**
Friday	-0.037	-0.038	-0.036	-0.036	-0.036	-0.036
	(13.79)**	(14.05)**	(13.23)**	(13.40)**	(13.36)**	(13.47)**
Saturday	-0.075	-0.076	-0.074	-0.074	-0.074	-0.074
	(23.10)**	(23.15)**	(22.71)**	(22.73)**	(22.82)**	(22.78)**
Carrier Fixed Effects (American Airlines omitted):						
Continental	-0.024	-0.022	-0.031	-0.029	-0.031	-0.029
	(11.97)**	(10.98)**	(15.25)**	(14.50)**	(15.33)**	(14.54)**
Delta	-0.126	-0.135	-0.130	-0.139	-0.130	-0.139
	(71.10)**	(73.35)**	(73.71)**	(75.95)**	(73.72)**	(75.95)**
Northwest	0.057	0.046	0.092	0.083	0.092	0.083
	(24.28)**	(18.87)**	(38.04)**	(33.04)**	(38.01)**	(33.03)**
United Airways	0.126	0.119	0.148	0.143	0.148	0.143
	(79.27)**	(73.17)**	(90.57)**	(86.42)**	(90.64)**	(86.44)**
US Air	-0.115	-0.131	-0.099	-0.115	-0.099	-0.115
	(52.02)**	(52.29)**	(44.46)**	(45.86)**	(44.53)**	(45.92)**
Frontier Airlines	0.046	0.056	0.091	0.104	0.089	0.104
	(12.64)**	(14.90)**	(24.36)**	(27.18)**	(23.92)**	(26.84)**
Alaska	-0.186	-0.220	-0.141	-0.169	-0.140	-0.169
	(32.59)**	(37.21)**	(24.39)**	(28.56)**	(24.32)**	(28.51)**
Hawaiian Airlines	0.327	0.292	0.314	0.283	0.314	0.283
	(19.99)**	(17.79)**	(19.28)**	(17.28)**	(19.29)**	(17.29)**
America West	-0.123	-0.120	-0.096	-0.087	-0.096	-0.087
	(29.20)**	(27.02)**	(22.57)**	(19.55)**	(22.55)**	(19.54)**
American Trans Air	0.139	0.126	0.146	0.133	0.146	0.133
	(8.80)**	(7.93)**	(9.26)**	(8.40)**	(9.28)**	(8.41)**
Midwest Express Airlines	0.483	0.614	0.469	0.598	0.468	0.597

Table 4
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets

	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
	(31.16)**	(36.05)**	(30.32)**	(35.25)**	(30.28)**	(35.20)**
Air Tran	-0.310	-0.255	-0.347	-0.294	-0.347	-0.294
	(25.71)**	(20.47)**	(28.79)**	(23.69)**	(28.82)**	(23.70)**
Spirit	-0.171	-0.155	-0.123	-0.102	-0.125	-0.103
	(27.95)**	(24.65)**	(19.90)**	(16.18)**	(20.17)**	(16.26)**
Sun County	-0.200	-0.134	-0.159	-0.091	-0.161	-0.092
	(22.97)**	(14.16)**	(18.32)**	(9.63)**	(18.51)**	(9.70)**
Constant	2.776	2.873	3.631	3.753	3.635	3.755
	(38.03)**	(39.15)**	(48.80)**	(50.12)**	(48.85)**	(50.14)**
Observations	523618	523618	523618	523618	523618	523618
R-squared	0.70	0.70	0.70	0.70	0.70	0.70

Note: Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%;
Source: Please refer to Table 1

Table 5
Regression of the Standard Deviation of Residuals on Market Structure Variables and Internet Usage

	(1) Standard deviation of residuals	(2) Standard deviation of residuals	(3) Standard deviation of residuals	(4) Standard deviation of residuals
Market share	0.133	0.110	0.133	0.052
	(37.74)**	(16.56)**	(37.67)**	(6.90)**
HHI	-0.149	-0.097	-0.147	-0.028
	(29.20)**	(10.40)**	(28.74)**	(2.89)**
Internet Share	-0.109	-0.101	-0.166	-0.077
	(14.30)**	(13.06)**	(2.27)*	(9.10)**
$(\text{Internet share})^2$			0.600	
			(2.11)*	
$(\text{Internet share})^3$			-1.076	
			(3.23)**	
Carrier Fixed Effects (American Airlines omitted):				
Continental				-0.015
				(4.44)**
Delta				-0.048
				(17.50)**
Northwest				-0.024
				(6.86)**
United				0.005
				(2.03)*
US Air				-0.046
				(13.46)**
Frontier				-0.036
				(9.03)**
Alaska				-0.035
				(5.90)**
Hawaiian Airlines				-0.061
				(2.71)**
America Mid West				-0.030
				(6.03)**
American Trans Air				-0.027
				(2.04)*
Midwest Express				-0.053
				(3.47)**
Air Tran				-0.189
				(27.58)**
Spirit				-0.085
				(14.95)**
Sun County				-0.089
				(10.58)**

Table 5
Regression of the Standard Deviation of Residuals on Market Structure Variables and Internet Usage

	(1) Standard deviation of residuals	(2) Standard deviation of residuals	(3) Standard deviation of residuals	(4) Standard deviation of residuals
Other Controls:				
Departure Date Fixed effects	No	No	No	Yes
Constant	0.260	0.241	0.257	0.200
	(84.40)**	(58.28)**	(42.63)**	(22.36)**
Observations	21670	21670	21670	21670
R-squared	0.070	0.070	0.070	0.130

Note: Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%
Source: Please refer to Table 1.

Table 6
Benefit from Purchasing on the Internet

	(1) log(Fare)	(2) log(Fare)	(3) log(Fare)	(4) log(Fare)
	Columns 1 & 2 Include Only Matched Observations		Columns 3 & 4 Include Matched & Unmatched Observations	
Online	-0.414	-0.442	-0.473	-0.510
	(176.86)**	(183.56)**	(479.25)**	(458.72)**
Departure Day of Week (Sunday omitted):				
Monday	0.001		-0.013	
	(0.250)		(9.80)**	
Tuesday	-0.015		-0.045	
	(5.21)**		(31.78)**	
Wednesday	-0.035		-0.062	
	(12.06)**		(44.17)**	
Thursday	-0.080		-0.108	
	(26.74)**		(74.76)**	
Friday	-0.147		-0.172	
	(47.22)**		(115.17)**	
Saturday	-0.116		-0.182	
	(31.12)**		(106.59)**	
Return Day of the Week (Sunday omitted):				
Monday	-0.392		-0.360	
	(131.94)**		(271.45)**	
Tuesday	-0.356		-0.334	
	(130.92)**		(253.59)**	
Wednesday	-0.362		-0.331	
	(143.75)**		(264.20)**	
Thursday	-0.348		-0.297	
	(145.60)**		(248.23)**	
Friday	-0.356		-0.294	
	(155.85)**		(255.36)**	
Saturday	-0.411		-0.353	
	(113.81)**		(221.78)**	
Carrier fixed effects	Yes	Yes	Yes	Yes
Route fixed effects	Yes	Yes	Yes	Yes
Constant	6.037	5.797	5.884	5.611
	(637.21)**	(600.95)**	(1184.78)**	(1120.81)**
Observations	523618	523618	2389605	2389605
R-squared	0.380	0.320	0.380	0.340

Note: Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%
Source: Please refer to Table 1.

Table 7
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets for Observations Matched within Five percent Range

	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
Market Structure Variables:						
Market share	0.078	0.294	0.084	0.297	0.083	0.296
	(26.10)**	(32.62)**	(28.26)**	(33.19)**	(27.77)**	(32.97)**
HHI	0.040	-0.262	0.011	-0.260	0.013	-0.258
	(10.84)**	(26.91)**	(3.03)**	(26.89)**	(3.44)**	(26.62)**
HUB	0.144	0.125	0.133	0.112	0.133	0.112
	(110.17)**	(76.65)**	(102.21)**	(68.88)**	(102.16)**	(68.87)**
Slot constrained airport	0.119	0.120	0.117	0.115	0.117	0.114
	(83.66)**	(79.81)**	(82.63)**	(76.57)**	(82.48)**	(76.47)**
Internet Variables:						
Online	-0.154	-0.149	-0.146	-0.141	-0.179	-0.166
	(114.26)**	(109.25)**	(108.84)**	(103.58)**	(59.83)**	(54.56)**
Internet share			-0.603	-0.635	-0.631	-0.656
			(90.63)**	(93.99)**	(89.76)**	(92.28)**
Internet share*Online					0.142	0.108
					(12.35)**	(9.30)**
Advance Purchase Requirement: **(No advance purchase required omitted)**						
One day advance	-0.439	-0.437	-0.405	-0.404	-0.405	-0.403
	(133.53)**	(132.17)**	(123.31)**	(122.01)**	(123.12)**	(121.87)**
Three day advance	-0.004	0.001	0.003	0.007	0.003	0.008
	(2.70)**	(0.470)	(1.600)	(4.42)**	(1.900)	(4.64)**
Five day advance	-0.630	-0.626	-0.586	-0.584	-0.587	-0.584
	(40.53)**	(40.09)**	(37.90)**	(37.57)**	(37.95)**	(37.62)**
Seven day advance	-0.172	-0.171	-0.169	-0.168	-0.168	-0.168
	(127.03)**	(126.04)**	(125.48)**	(124.70)**	(125.22)**	(124.50)**
Ten day advance	-0.236	-0.229	-0.238	-0.233	-0.237	-0.233
	(80.93)**	(78.19)**	(81.94)**	(79.96)**	(81.66)**	(79.75)**
Fourteen day advance	-0.235	-0.231	-0.239	-0.236	-0.239	-0.235
	(154.63)**	(150.30)**	(157.82)**	(154.48)**	(157.44)**	(154.22)**
Twenty-one day advance	-0.257	-0.256	-0.255	-0.255	-0.255	-0.254
	(80.43)**	(79.72)**	(80.19)**	(79.76)**	(79.96)**	(79.58)**
Thirty day advance	0.031	0.033	0.024	0.025	0.025	0.027
	(4.29)**	(4.60)**	(3.31)**	(3.56)**	(3.55)**	(3.74)**
Other Ticket Characteristics:						
Non-refundable ticket	-0.403	-0.408	-0.402	-0.406	-0.402	-0.406

Table 7
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets for Observations Matched within Five percent Range

	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
	(268.49)**	(269.74)**	(269.39)**	(269.90)**	(269.20)**	(269.72)**
Days prior to departure ticket purchased	-0.0004	-0.0004	-0.0002	-0.0002	-0.0002	-0.0002
	(15.34)**	(17.38)**	(7.37)**	(8.81)**	(8.18)**	(9.41)**
Saturday stay-over	-0.118	-0.119	-0.114	-0.115	-0.115	-0.115
	(64.86)**	(64.75)**	(62.97)**	(62.85)**	(63.26)**	(63.07)**
Travel restriction requirement	-0.305	-0.307	-0.301	-0.302	-0.301	-0.302
	(314.10)**	(313.16)**	(311.07)**	(309.25)**	(311.09)**	(309.27)**
Minimum stay requirement	0.013	0.017	0.010	0.014	0.010	0.014
	(10.02)**	(12.81)**	(7.75)**	(10.17)**	(7.73)**	(10.14)**
Maximum stay requirement	0.027	0.031	0.031	0.034	0.031	0.034
	(19.65)**	(22.37)**	(22.07)**	(24.44)**	(22.25)**	(24.57)**
First class	0.716	0.714	0.705	0.702	0.704	0.702
	(377.99)**	(374.91)**	(373.13)**	(369.80)**	(372.69)**	(369.51)**
Business class	0.369	0.377	0.375	0.382	0.375	0.382
	(247.99)**	(249.72)**	(253.46)**	(253.83)**	(253.40)**	(253.75)**
Remaining Ticket Characteristics:						
Direct flight	0.023	-0.017	0.020	-0.018	0.020	-0.018
	(6.13)**	(4.08)**	(5.26)**	(4.49)**	(5.35)**	(4.40)**
Roundtrip ticket	-0.088	-0.090	-0.093	-0.096	-0.093	-0.096
	(41.30)**	(41.92)**	(43.76)**	(44.78)**	(43.45)**	(44.53)**
Flight load factor at purchase	0.241	0.216	0.216	0.190	0.212	0.188
	(35.42)**	(31.30)**	(31.89)**	(27.76)**	(31.33)**	(27.36)**
Peak Time of Day:						
Departure at peak time	0.017	0.018	0.017	0.018	0.017	0.018
	(17.26)**	(18.66)**	(17.41)**	(18.75)**	(17.43)**	(18.76)**
Return at peak time	0.028	0.028	0.027	0.028	0.027	0.028
	(24.77)**	(25.29)**	(24.23)**	(24.75)**	(24.19)**	(24.72)**
Other Route Specific Characteristics:						
Low cost carrier on route	-0.115	-0.117	-0.090	-0.090	-0.089	-0.090
	(116.40)**	(116.97)**	(87.96)**	(87.91)**	(87.37)**	(87.42)**
Southwest Airlines	-0.198	-0.189	-0.169	-0.161	-0.170	-0.161
	(99.48)**	(94.28)**	(84.46)**	(79.23)**	(84.90)**	(79.52)**
Other Route Level Variables:						
Distance (log)	0.384	0.377	0.422	0.416	0.423	0.417
	(451.27)**	(395.39)**	(446.22)**	(403.63)**	(446.23)**	(402.79)**
Absolute Temperature	-0.007	-0.007	-0.005	-0.006	-0.005	-0.006

Table 7
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets for Observations Matched within Five percent Range

	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
Difference (Log)	(18.09)**	(19.80)**	(13.89)**	(15.03)**	(13.99)**	(15.09)**
Average population (Log)	-0.026	-0.027	-0.038	-0.038	-0.037	-0.037
	(38.34)**	(37.86)**	(54.74)**	(52.03)**	(54.64)**	(51.94)**
Average per capita	0.204	0.206	0.096	0.095	0.095	0.094
Income (Log)	(36.48)**	(36.69)**	(16.95)**	(16.59)**	(16.66)**	(16.38)**
Departure Day of Week (Sunday omitted):						
Monday	0.003	0.003	0.002	0.003	0.002	0.003
	(1.680)	(1.820)	(1.510)	(1.660)	(1.450)	(1.620)
Tuesday	0.005	0.006	0.004	0.005	0.004	0.005
	(3.03)**	(3.33)**	(2.55)*	(2.82)**	(2.52)*	(2.79)**
Wednesday	0.004	0.005	0.004	0.005	0.004	0.005
	(2.47)*	(3.02)**	(2.45)*	(2.93)**	(2.33)*	(2.84)**
Thursday	0.011	0.012	0.011	0.012	0.011	0.012
	(6.00)**	(6.70)**	(6.09)**	(6.71)**	(6.01)**	(6.65)**
Friday	0.014	0.015	0.014	0.014	0.014	0.014
	(7.20)**	(7.42)**	(7.07)**	(7.20)**	(7.18)**	(7.28)**
Saturday	-0.067	-0.067	-0.066	-0.066	-0.066	-0.066
	(31.90)**	(31.64)**	(31.57)**	(31.44)**	(31.57)**	(31.43)**
Return Day of the Week (Sunday omitted):						
Monday	-0.048	-0.048	-0.046	-0.046	-0.047	-0.047
	(23.49)**	(23.47)**	(22.89)**	(22.81)**	(23.31)**	(23.12)**
Tuesday	-0.038	-0.038	-0.036	-0.036	-0.038	-0.037
	(17.47)**	(17.37)**	(16.93)**	(16.76)**	(17.46)**	(17.15)**
Wednesday	-0.047	-0.047	-0.045	-0.044	-0.046	-0.045
	(21.29)**	(21.24)**	(20.50)**	(20.35)**	(20.99)**	(20.71)**
Thursday	-0.034	-0.035	-0.032	-0.033	-0.033	-0.034
	(15.68)**	(15.91)**	(14.85)**	(14.96)**	(15.32)**	(15.32)**
Friday	-0.041	-0.042	-0.038	-0.039	-0.039	-0.040
	(18.83)**	(19.14)**	(17.84)**	(18.02)**	(18.29)**	(18.35)**
Saturday	-0.075	-0.076	-0.072	-0.073	-0.073	-0.074
	(28.89)**	(29.09)**	(28.07)**	(28.19)**	(28.42)**	(28.45)**
Carrier Fixed Effects (American Airlines omitted):						
Continental	-0.044	-0.041	-0.052	-0.049	-0.053	-0.050
	(26.16)**	(23.91)**	(30.99)**	(29.12)**	(31.23)**	(29.30)**
Delta	-0.110	-0.121	-0.115	-0.127	-0.115	-0.126
	(75.11)**	(79.39)**	(79.05)**	(83.07)**	(79.00)**	(83.01)**

Table 7
Internet Purchase, Internet Share and Potential Savings from Internet Purchase in High-Low Internet Usage Markets for Observations Matched within Five percent Range

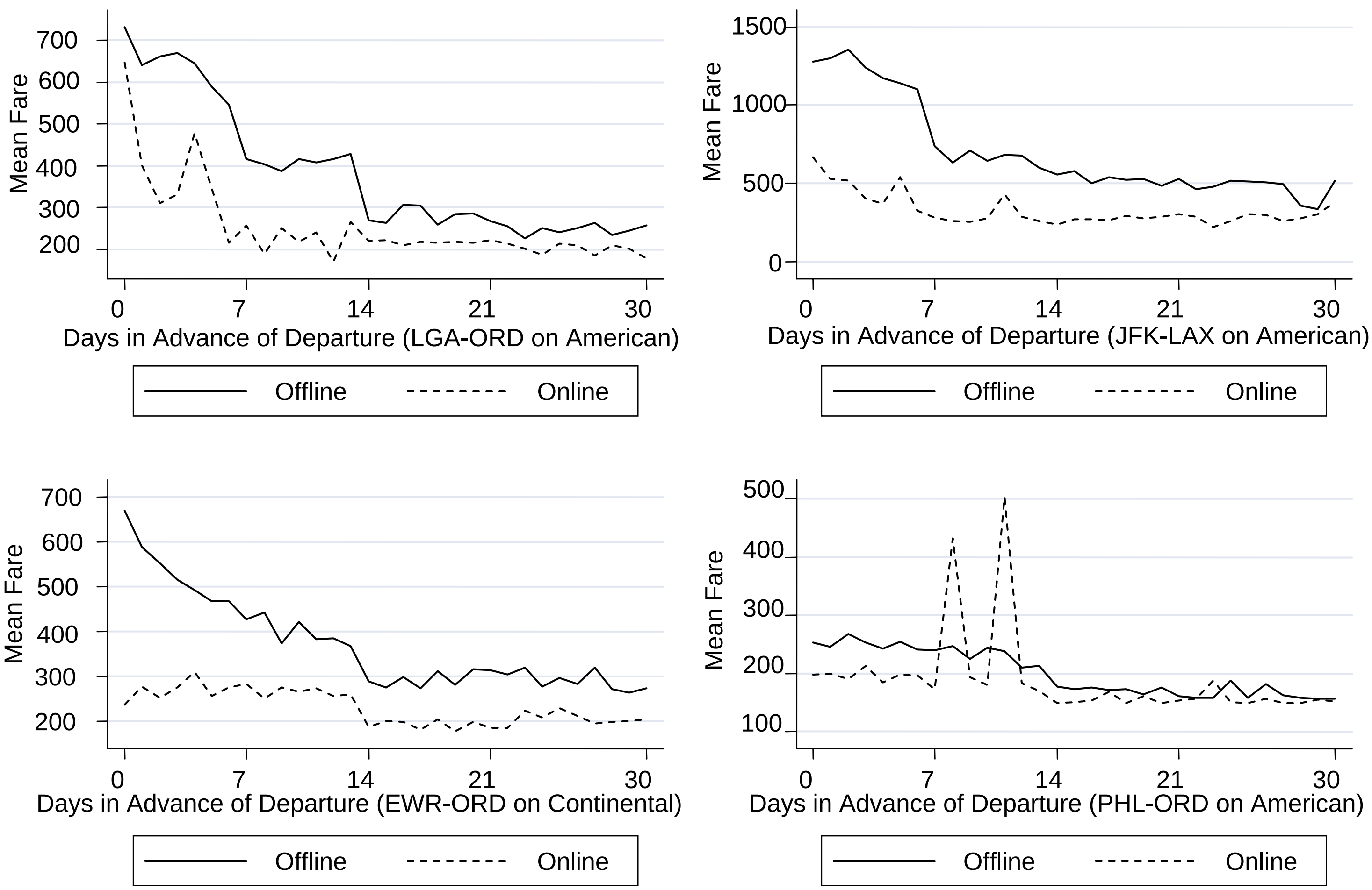
	(1) Log(Fare) (OLS)	(2) Log(Fare) (IV)	(3) Log(Fare) (OLS)	(4) Log(Fare) (IV)	(5) Log(Fare) (OLS)	(6) Log(Fare) (IV)
Northwest	0.076	0.064	0.123	0.113	0.124	0.113
	(38.71)**	(31.25)**	(60.91)**	(54.09)**	(60.95)**	(54.12)**
United Airways	0.109	0.101	0.136	0.131	0.137	0.132
	(82.49)**	(74.52)**	(100.75)**	(95.74)**	(101.09)**	(95.99)**
US Air	-0.111	-0.131	-0.092	-0.111	-0.092	-0.112
	(60.38)**	(62.99)**	(49.72)**	(53.86)**	(49.97)**	(54.07)**
Frontier Airlines	0.071	0.086	0.126	0.145	0.123	0.143
	(22.88)**	(26.39)**	(39.87)**	(44.07)**	(38.91)**	(43.25)**
Alaska	-0.096	-0.133	-0.036	-0.065	-0.035	-0.065
	(21.39)**	(28.57)**	(7.87)**	(14.05)**	(7.72)**	(13.92)**
Hawaiian Airlines	0.329	0.291	0.306	0.273	0.306	0.274
	(22.25)**	(19.52)**	(20.75)**	(18.46)**	(20.79)**	(18.50)**
America West	-0.057	-0.049	-0.021	-0.006	-0.021	-0.006
	(16.23)**	(13.18)**	(6.07)**	(1.610)	(5.93)**	(1.490)
American Trans Air	0.033	0.018	0.039	0.023	0.041	0.024
	(2.92)**	(1.550)	(3.45)**	(2.01)*	(3.63)**	(2.15)*
Midwest Express Airlines	0.500	0.666	0.483	0.642	0.482	0.640
	(36.03)**	(43.86)**	(35.00)**	(42.53)**	(34.86)**	(42.40)**
Air Tran	-0.330	-0.275	-0.388	-0.338	-0.389	-0.339
	(41.43)**	(33.36)**	(48.81)**	(41.26)**	(48.97)**	(41.37)**
Spirit	-0.225	-0.201	-0.168	-0.140	-0.173	-0.144
	(47.08)**	(40.75)**	(34.97)**	(28.19)**	(35.91)**	(28.83)**
Sun County	-0.109	-0.031	-0.057	0.022	-0.063	0.018
	(16.00)**	(4.10)**	(8.37)**	(2.99)**	(9.23)**	(2.34)*
Constant	1.917	2.059	3.058	3.210	3.075	3.222
	(31.32)**	(33.39)**	(49.19)**	(51.19)**	(49.45)**	(51.38)**
Observations	794134	794134	794134	794134	794134	794134
R-squared	0.690	0.680	0.690	0.690	0.690	0.690

Note: Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%;
Source: Please refer to Table 1

Note: The above set of regressions replicates the analysis in Table 4 in the paper but now it includes observations that were matched within 5 percent range as compared to the 2 percent matches included in the main analysis. The results are qualitatively similar, but now the direct and indirect effects of the internet are much stronger and significant. Also, note that the inclusion of the 5 percent matched observations also increases the sample size by almost 50 percent.

Figure 1

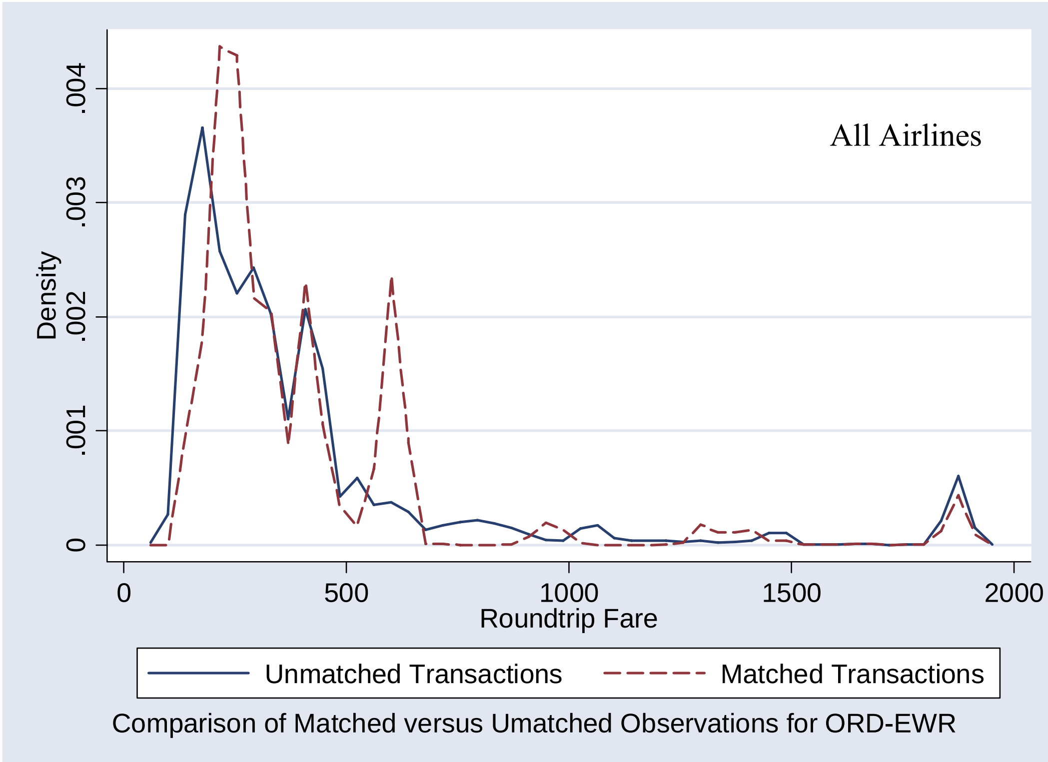
Comparison Between Online and Offline Daily Average Fares



Mean Fares Offline and Online for Purchases in Last 30 Days Before Departure

Figure 2

Comparing the Kernel Densities of Matched and Unmatched Transactions for flights between Chicago O' Hare – Newark Liberty



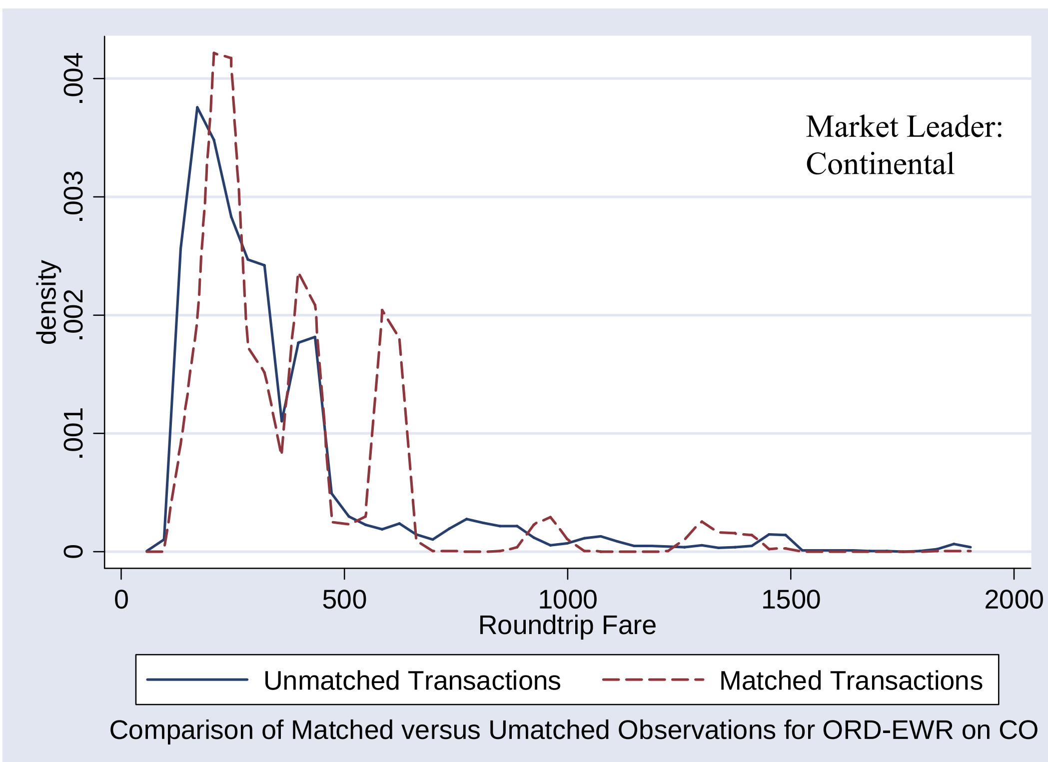
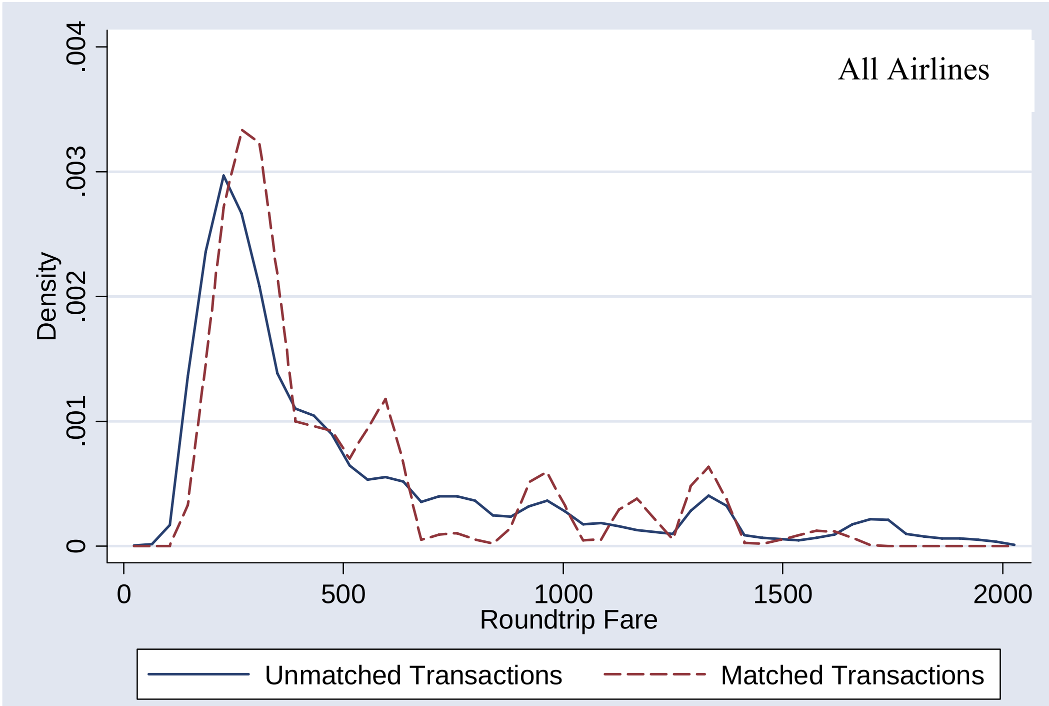
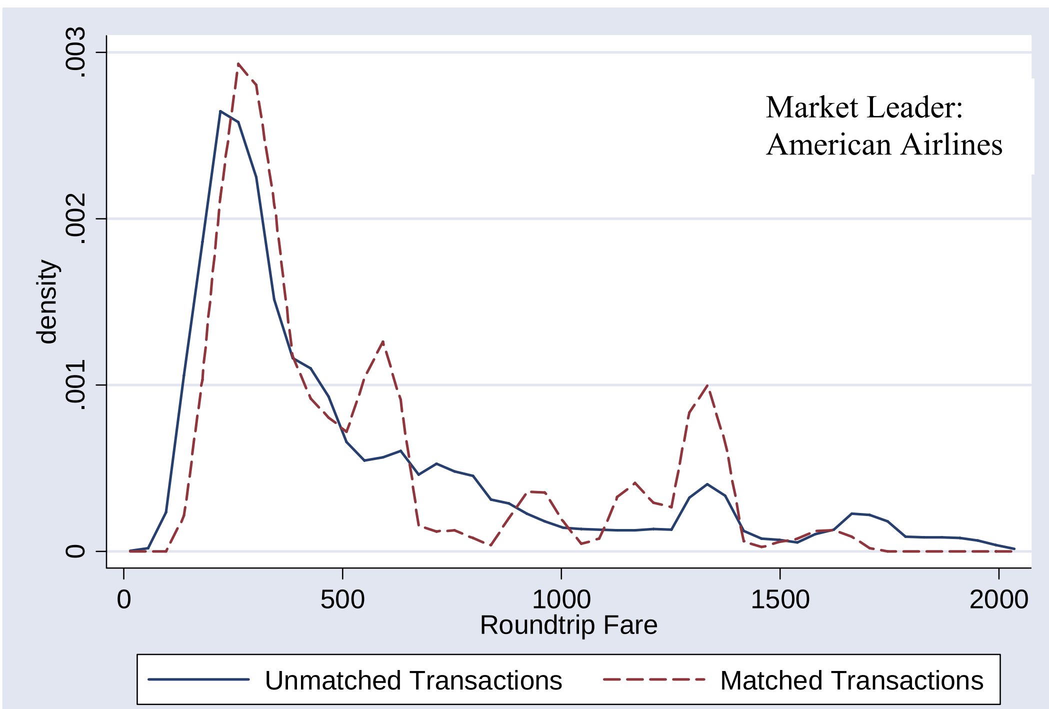


Figure 3
Comparing the Kernel Densities of Matched and Unmatched Transactions for flights between Kennedy, New York – Los Angeles, California

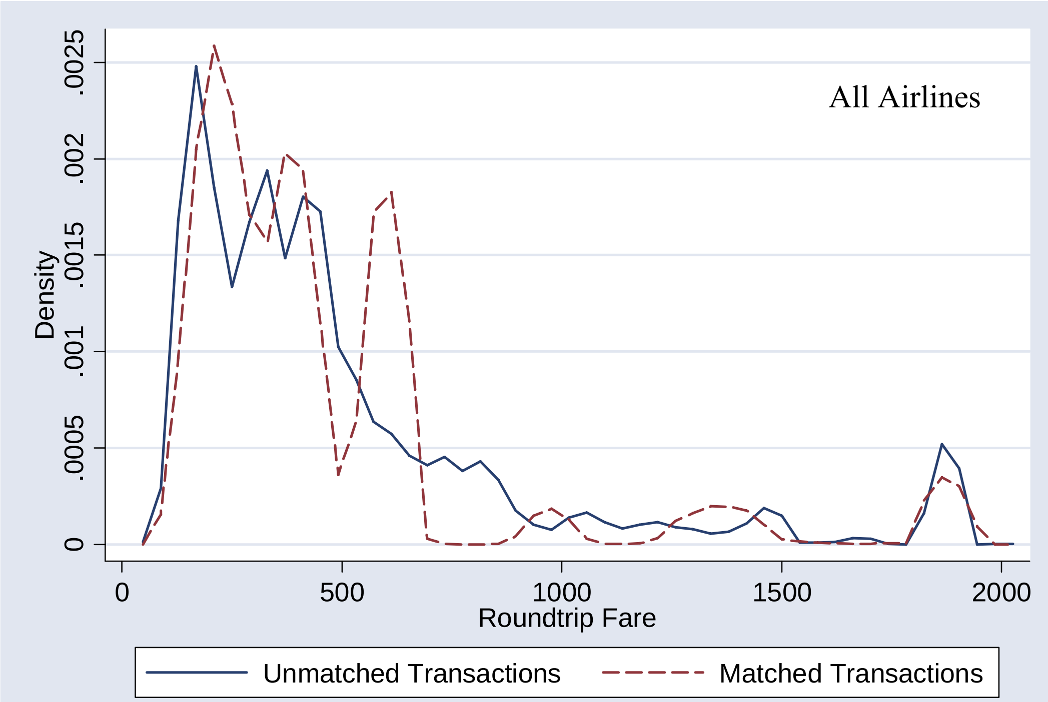


Comparison of Matched versus Umatched Observations for JFK-LAX

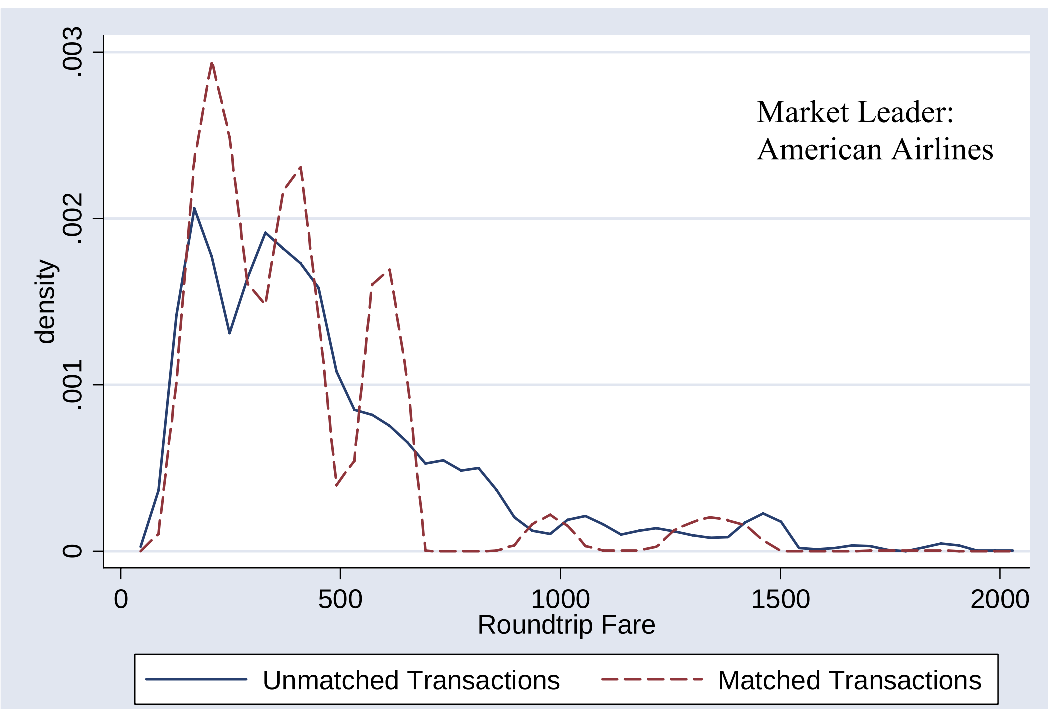


Comparison of Matched versus Umatched Observations for JFK-LAX on AA

Figure 4
Comparing the Kernel Densities of Matched and Unmatched Transactions for flights between
La Guardia, New York – Chicago O' Hare

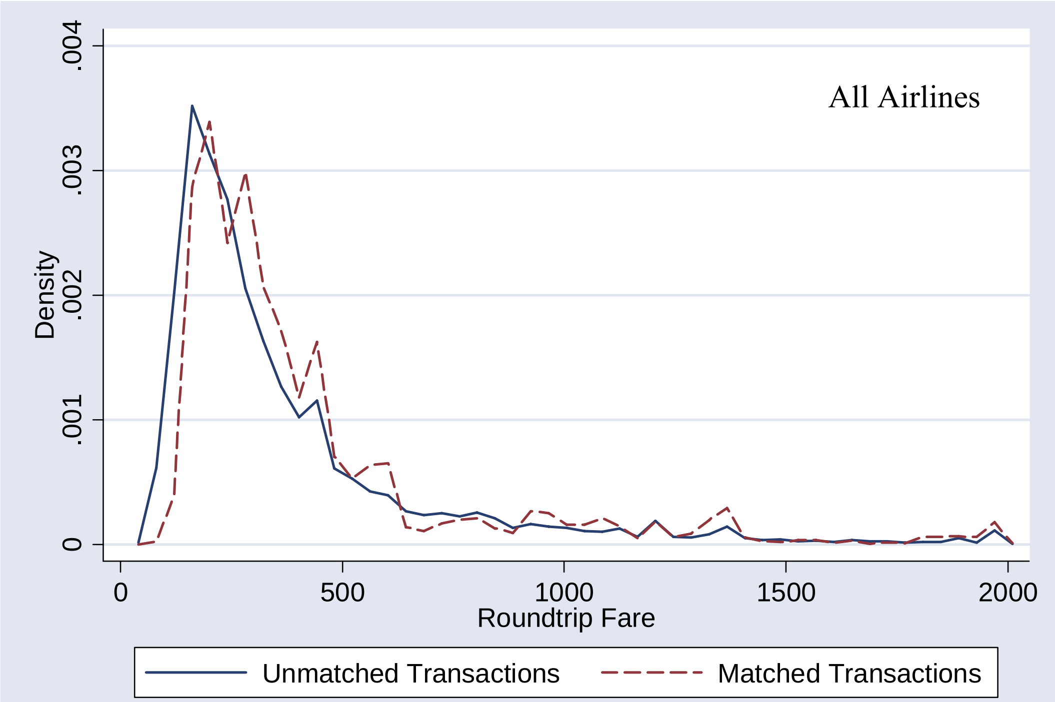


Comparison of Matched versus Umatched Observations for LGA-ORD

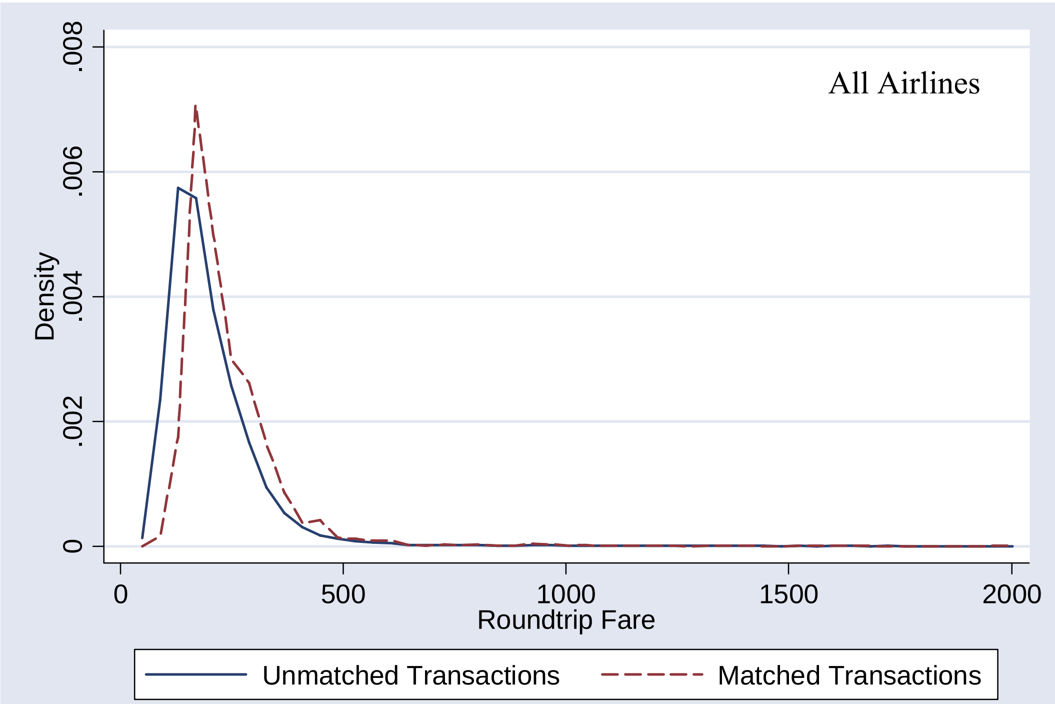


Comparison of Matched versus Umatched Observations for LGA-ORD on AA

Figure 5
Comparing the Kernel Densities of Matched and Unmatched Transactions for Overall Online and Offline Transactions

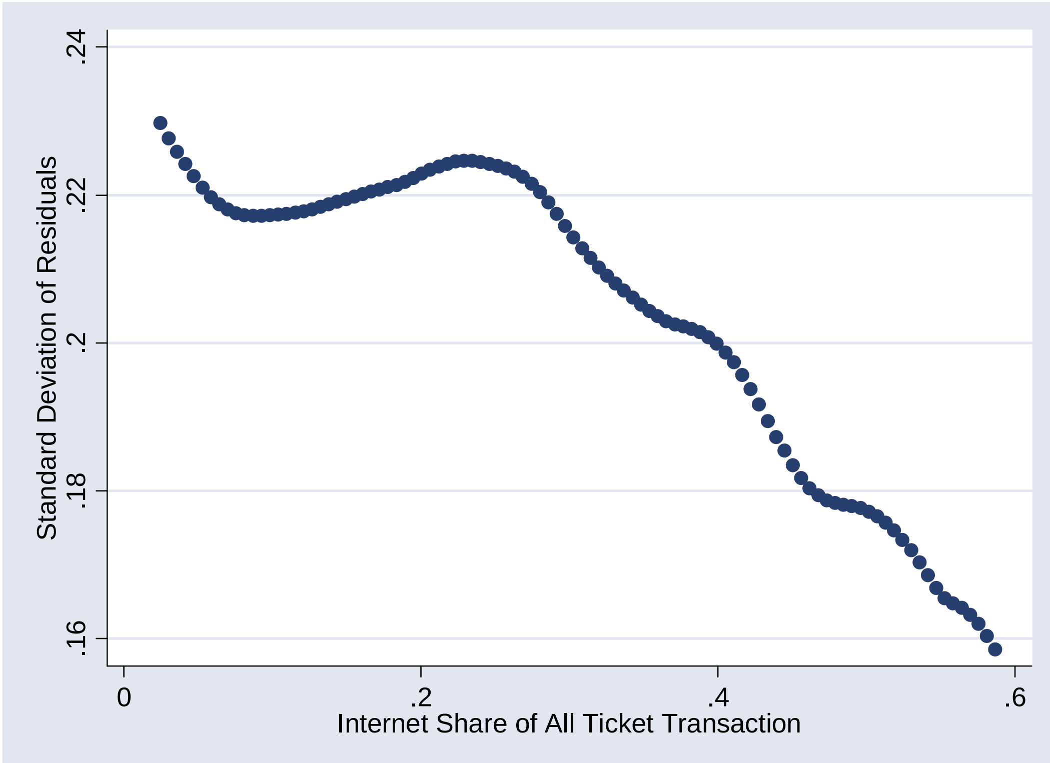


Comparing Matched versus Umatched Observations: All Offline Transactions



Comparing Matched versus Umatched Observations: All Online Transactions

Figure 6



Price Dispersion and Share of Internet Purchases