

## Tracing the Woes: An Empirical Analysis of the Airline Industry<sup>†</sup>

By STEVEN BERRY AND PANLE JIA\*

*The US airline industry went through tremendous turmoil in the early 2000s, with four major bankruptcies, two major mergers, and various changes in network structure. This paper presents a structural model of the industry, and estimates the impact of demand and supply changes on profitability. Compared with 1999, we find that, in 2006, air-travel demand was 8 percent more price sensitive, passengers displayed a stronger preference for nonstop flights, and changes in marginal cost significantly favored nonstop flights. Together with the expansion of low-cost carriers, they explain more than 80 percent of legacy carriers' variable profit reduction. (JEL L13, L25, L93)*

The airline industry went through tremendous turmoil in the early 2000s with four major bankruptcies and two mergers. In August 2002, US Airways filed for bankruptcy. A few months later, United Airlines followed suit. It stayed under Chapter 11 bankruptcy protection for more than three years, the largest and longest airline bankruptcy in history. In September 2005, Delta Airlines and Northwest Airlines went bankrupt on the same day. By then, four of the six legacy carriers were under bankruptcy reorganization.<sup>1</sup> Only American and Continental airlines managed to escape bankruptcy, but all legacy carriers reported a large reduction in profits.<sup>2</sup>

On the other hand, when measured by domestic revenue passenger miles,<sup>3</sup> by 2004 the industry's output had recovered from the sharp post-9/11 downturn and has been trending upward since (see Figure 1). The load factor,<sup>4</sup> another important measure of profitability, has increased steadily since 2001. According to Figure 2, the average load factor for US airlines rose from 71.2 percent in 1999

\* Berry: Department of Economics, Yale University, 37 Hillhouse Ave., New Haven, CT 06520 and National Bureau of Economics (e-mail: [steven.berry@yale.edu](mailto:steven.berry@yale.edu)); Jia: Department of Economics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 and National Bureau of Economics (e-mail: [pjia@mit.edu](mailto:pjia@mit.edu)). We would like to thank Severin Borenstein, Glenn Ellison, Michael Greenstone, Jerry Hausman, Richard Ratliff, Marc Rysman, the anonymous referees, and seminar participants at Vanderbilt University, MIT, Harvard University, the Department of Justice, University of Maryland, and University of Minnesota for their helpful comments. Special thanks go to Nancy Rose, who is extremely generous with her time and help. Michael Powell provided excellent research assistance. Comments are welcome.

<sup>†</sup> To comment on this article in the online discussion forum, or to view additional materials, visit the articles page at <http://www.aeaweb.org/articles.php?doi=10.1257/mic.2.3.1>.

<sup>1</sup> The legacy carriers are: American Airlines, Continental Airlines, Delta Airlines, Northwest Airlines, United Airlines, and US Airways.

<sup>2</sup> As documented by Severin Borenstein and Nancy L. Rose (2007), airline profits have always been volatile. However, developments in the early 2000s still seem severe.

<sup>3</sup> Revenue passenger miles is the product of the number of revenue-paying passengers aboard and the distance travelled (measured in miles).

<sup>4</sup> Load factor is the ratio of revenue passenger miles to available seat miles of a flight.

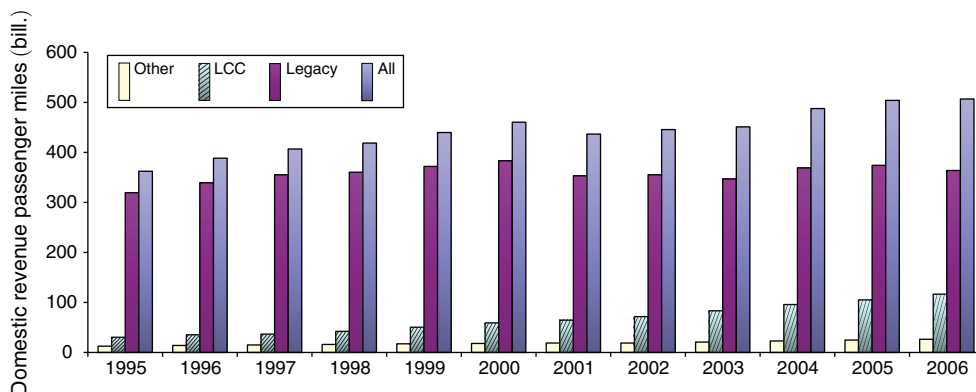


FIGURE 1. DOMESTIC REVENUE PASSENGER MILES (in billions)

Source: MIT Airline Data Project

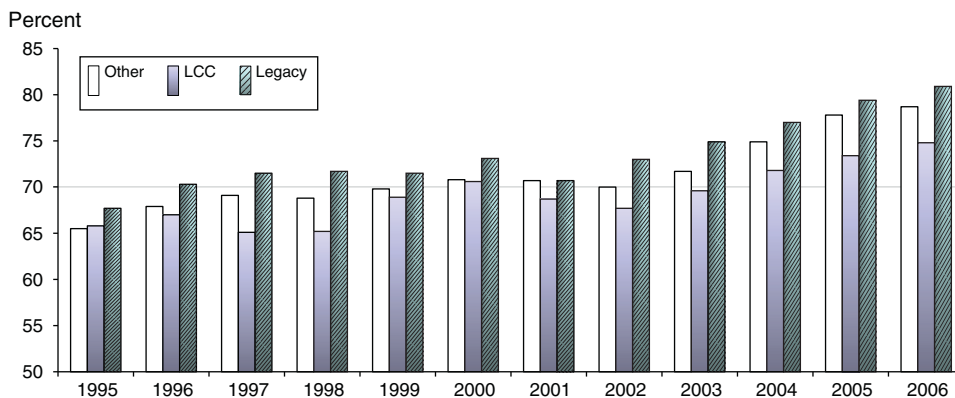


FIGURE 2. US AIRLINES' SYSTEM LOAD FACTORS

Source: MIT Airline Data Project

to 79.7 percent in 2006, and posted a record high of 80.5 percent in 2007. If more passengers traveled and planes were fuller, what caused the financial stress on most airlines?

Several recent developments provide potential explanations. One category of explanations is related to changes in air travel demand. Perhaps the bursting of the dot-com bubble and improvements in electronic communications have decreased the willingness-to-pay of business travelers. As the economy cooled down, many companies imposed maximum reimbursement limits, and even business travelers started to shop around for cheaper flights.

Another potential change in demand stems from the tightened security regulations after 9/11. Passengers have to go through a strict security check, and many items are no longer allowed in carry-on luggage. The extra luggage handling, combined with stricter security regulations, have lengthened the average travel time. In the meantime, with most flights full, it has become increasingly difficult for passengers to board a different plane in case of missed connections or flight cancellations.

Consequently, carriers find it harder to charge high fares for connecting flights as passengers start to search for alternatives.

A third important development is the option of purchasing airline tickets on the Internet. In 1996, most tickets were sold through airlines' reservation offices or traditional travel agencies, with less than 0.5 percent sold online.<sup>5</sup> By 2007, online sales accounted for 26 percent of global sales, and as high as 50–60 percent of sales in the US.<sup>6</sup> The proliferation of online sites that provide information previously limited to travel agents has increased consumers' awareness of fare availability and fare premiums across carriers and travel dates. The various search engines (travelocity.com, expedia.com, etc.) have dramatically reduced consumers' search costs, and allowed them to easily find the most desirable flights. All of these changes affect consumers' sensitivity to flights with different attributes (high- versus low-fare tickets, direct versus connecting flights, frequent versus less frequent departures, etc.).<sup>7</sup>

On the supply side, a variety of changes have affected the industry's market structure and profitability. The most cited transition is the expansion of the low-cost carriers (LCC), whose market share of domestic origin-destination passengers increased steadily over the past decade, from 22.6 percent in 1999 to 32.9 percent in 2006.<sup>8</sup> As a result, the legacy carriers may have been forced to lower fares and offer competing services. Some legacy carriers have shifted capacity to the more lucrative international markets.

Recent progress in aviation technology, in particular the advent of regional jets with different plane sizes, allows carriers to better match aircraft with market size, and hence enables carriers to offer direct flights to markets that formerly relied on connecting services. In addition, with lower labor costs than traditional jets, regional jets have become a popular choice for carriers under financial pressure.<sup>9</sup> On the other hand, the cost of jet fuel, which accounts for roughly 15 percent of the operation cost, more than doubled over the period of our data.<sup>10</sup>

In this paper, we estimate a structural model of the airline industry, and disentangle the impact of the various factors on the profitability of the legacy carriers. We find that, compared with the late 1990s, in 2006 the price elasticity of air travel demand increased by 8 percent. Passengers displayed a strong preference for non-stop flights. The connection semi-elasticity was 17 percent higher. On the supply side, changes in marginal cost significantly favored nonstop flights. A more elastic demand, a higher aversion toward connecting flights, and increasing cost disadvantages of connecting flights are the most robust findings of our study and are present in almost all specifications we have estimated. These factors, together with the expansion of low-cost carriers, explain more than 80 percent of the decrease in

<sup>5</sup> Source: Department of Transportation (DOT) report CR-2000-111.

<sup>6</sup> Source: SITA (2008).

<sup>7</sup> Technically, "direct" means that passengers do not change planes between origin and destination, while "non-stop" means that the flight does not stop between origin and destination. In this paper, we use both terms to refer to flights that do not stop between origin and destination.

<sup>8</sup> Data source: <http://www.darinlee.net/data/lccshare.html>.

<sup>9</sup> See Aleksandra L. Mozdzanowska (2004).

<sup>10</sup> See Borenstein and Rose (2007).

legacy carriers' variable profits, with changes in demand contributing to more than 50 percent of the reduction.

The remainder of the paper is structured as follows. Section I reviews the related literature. Section II presents the model. Section III describes the data sources. Section IV proposes the empirical strategy. Section V discusses the results. Section VI presents the conclusions.

## I. Literature Review

There have been many empirical papers that study the airline industry. Among the most recent ones, Borenstein (2005) reported that, adjusted for inflation, airline prices fell more than 20 percent from 1995 to 2004. He also found that premiums at hub airports declined, and that there was substantially less disparity between the cheaper and the more expensive airports than there had been a decade ago. Austan Goolsbee and Chad Syverson (2008) examined how incumbents responded to the threat of Southwest entry. Steven L. Puller, Anirban Sengupta, and Steven N. Wiggins (2007) tested theories of price dispersion and scarcity pricing in the airline industry. Federico Ciliberto and Elie Tamer (2009) used a partially identified entry model to investigate the heterogeneity in carrier profits. They found that repealing the Wright Amendment would increase the number of markets served out of Dallas Love airport. James D. Dana and Eugene Orlov (2008) studied the impact of Internet penetration on airlines' capacity utilization. Silke Forbes (2008) exploited a legislative change in takeoff and landing restrictions at LaGuardia Airport in 2000. She discovered that prices fell by \$1.42 on average for each additional minute of flight delay.

There are several recent discrete choice applications in the airline literature.<sup>11</sup> Craig Peters (2006) simulated post-merger prices for five airline mergers in the late 1980s, and found evidence that supply-side effects, such as changes in marginal costs and deviations from the assumed model of firm conduct, were important factors in post-merger price increases. Berry, Michael Carnall, and Spiller (hereafter BCS) (2006) focused on the evolution of the airline industry toward a hub-and-spoke system after deregulation in the 1970s. They found evidence of economies of density on longer routes. Olivier Armantier and Oliver Richard (2008) investigated the consumer welfare consequences of the code-share agreement between Continental Airlines and Northwest Airlines. The results suggested that the code-share agreement increased the average surplus of connecting passengers and decreased the average surplus of nonstop passengers, but did not impact consumers significantly on average. We contribute to the literature by examining recent developments in the airline industry and analyzing how they contribute to the drastic profit reductions witnessed in this industry.

<sup>11</sup> Earlier discrete choice demand studies of the airline industry include Steven A. Morrison et al. (1989), Peter C. Reiss and Pablo T. Spiller (1989), and Steven Berry (1990).

## II. Model

We consider a model of airline oligopoly “supply and demand” in the spirit of the recent literature on differentiated products following Berry, James Levinsohn, and Ariel Pakes (BLP) (1995). Our model is particularly close to BCS (2006).<sup>12</sup> The point of this paper is not to provide any methodological innovation, but to make use of the existing models to understand the recent evolution of the industry.

For now, we think of US airlines as offering a set of differentiated products in each of a large cross-section of “origin-and-destination” markets. Airline products are differentiated by price, the number of connections, airline brand, frequency of departures, and so forth. Ticket restrictions (such as advanced-purchase and length-of-stay requirements) are important elements of product differentiation that are not observed in our data. Neither do we observe certain flight-level details, such as the time of departure. Thus, it is particularly important to allow for product-unobservable characteristics that are correlated with price, as explained below.

### A. Demand

The demand model is a simple random-coefficient discrete-choice model in the spirit of Daniel McFadden (1981) and BLP (1995). Like BCS (2006), we use a “discrete-type” version of the random coefficient model. Suppose there are  $R$  types of consumers. For product  $j$  in market  $t$ , the utility of consumer  $i$ , who is of type  $r$ , is given by

$$(1) \quad u_{ijt} = x_{jt} \beta_r - \alpha_r p_{jt} + \xi_{jt} + \nu_{it}(\lambda) + \lambda \epsilon_{ijt},$$

where

- $x_{jt}$  is a vector of product characteristics,
- $\beta_r$  is a vector of “tastes for characteristics” for consumers of type  $r$ ,
- $\alpha_r$  is the marginal disutility of a price increase for consumers of type  $r$ ,
- $p_{jt}$  is the product price,
- $\xi_{jt}$  is the unobserved (to researchers) characteristic of product  $j$ ,
- $\nu_{it}$  is a “nested logit” random taste that is constant across airline products and differentiates “air travel” from the “outside” good,
- $\lambda$  is the nested logit parameter that varies between 0 and 1, and
- $\epsilon_{ijt}$  is an independently and identically distributed (across products and consumers) “logit error.”

<sup>12</sup> As will be clear in Section IVA, there are several differences between our model and the BCS (2006) model. On the demand side, we construct the number of departures (as a measure of flight frequency) using the minute-by-minute flight schedules for all flights operated within the US continent. We also instrument this variable (because we treat both prices and departure frequencies as endogenous) using end city characteristics. Carrier dummies are included in both demand and supply equations. While the reported specification might look similar to BCS (2006), we tried many other specifications (discussed in Section IVA) in our robustness analysis. On the cost side, we have a less explicit model of the source of airlines’ “hub advantage.” The simpler marginal cost model allows us to perform counterfactual analysis. Our marginal cost specification is discussed in Section IIB.

The utility of the outside good is given by

$$(2) \quad u_{i0t} = \epsilon_{i0t},$$

where  $\epsilon_{i0t}$  is another logit error. The error structure

$$v_{it}(\lambda) + \lambda \epsilon_{ijt}$$

is assumed to follow the distributional assumption necessary to generate the classic nested logit purchase probability for consumers of type  $r$ , where the two nests consist of: all the airline products, and the outside option of not flying. If  $\lambda = 1$ , then  $v_{it}(\lambda) \equiv 0$ , and the purchase probability of type  $r$  consumers takes the simple multinomial logit form. If  $\lambda = 0$ , then the independently and identically distributed  $\epsilon$ 's have no effect. Conditioning on flying, all type  $r$  consumers buy the product with the highest  $x_{jt}\beta_r - \alpha_r p_{jt} + \xi_{jt}$ . When  $\lambda \in (0, 1)$ , the product shares have the traditional nested logit form.

Specifically, conditional on purchasing some airline product, the percentage of type  $r$  consumers who purchase product  $j$  in market  $t$  is given by

$$\frac{e^{(x_{jt}\beta_r - \alpha_r p_{jt} + \xi_{jt})/\lambda}}{D_{rt}},$$

where the denominator is

$$(3) \quad D_{rt} = \sum_{k=1}^J e^{(x_{kt}\beta_r - \alpha_r p_{kt} + \xi_{kt})/\lambda}.$$

The share of type  $r$  consumers who make a purchase is

$$(4) \quad s_i^r(x_t, p_t, \xi_t, \theta_d) \equiv \frac{D_{rt}^\lambda}{1 + D_{rt}^\lambda}.$$

Let  $\gamma_r$  denote the percentage of type  $r$  consumers in the population. The overall market share of product  $j$  in market  $t$  is

$$(5) \quad s_{jt}(x_t, p_t, \xi_t, \theta_d) \equiv \sum_r \gamma_r \frac{e^{(x_{jt}\beta_r - \alpha_r p_{jt} + \xi_{jt})/\lambda}}{D_{rt}} s_i^r(x_t, p_t, \xi_t, \theta_d).$$

Notice that the vector of demand parameters to be estimated,  $\theta_d$ , includes the taste for product characteristics,  $\beta_r$ , the disutility of price,  $\alpha_r$ , the nested logit parameter,  $\lambda$  (which governs substitution to the outside good), and the consumer-type probability  $\gamma_r$ .

Following BLP (1995), we form moments that are expectations of the unobservable  $\xi$  interacted with exogenous instruments that are discussed in Section IVB. Further details of the estimation method are found in BLP (1995) and the related literature, but we provide a brief review here.

We first invert the market share equation (5) to solve for the vector of demand unobservables  $\xi_t$  as a function of the product characteristics, prices, the observed market shares, and parameters:

$$(6) \quad \xi_t = s^{-1}(x_t, p_t, s_t, \theta_d).$$

As in BCS (2006), the multiple-type nested logit model requires us to slightly modify the contraction mapping method used in BLP (1995). In particular, the “step” between each iteration of  $\xi_t$  is multiplied by  $\lambda$ , the nested logit parameter:<sup>13</sup>

$$(7) \quad \xi_{jt}^M = \xi_{jt}^{M-1} + \lambda[\ln s_{jt} - \ln s_{jt}(x_t, p_t, \xi_t, \theta_d)],$$

where  $M$  denotes the  $M$ th iteration,  $s_{jt}$  is the observed product share, and  $s_{jt}(x_t, p_t, \xi_t, \theta_d)$  is defined by equation (5).

The moment conditions used in estimation are based on restrictions of the form

$$(8) \quad E(\xi(x_t, p_t, s_t, \theta_d) | z_t) = 0,$$

where  $z_t$  is a vector of instruments. These moment conditions imply

$$(9) \quad E(h(z_t) \xi(x_t, p_t, s_t, \theta_d)) = 0,$$

for any vector of functions  $h(\cdot)$ . Intuitively, a method of moment estimation routine chooses  $\theta_d$  to make the sample analogue of the expectation in (9) as close to zero as possible.

The product-level unobservable  $\xi_{jt}$  accounts for a number of product characteristics, such as ticket restrictions and departure time, that are absent from our data source.<sup>14</sup> Prices are likely to be correlated with these product attributes. For example, refundable tickets are generally much more expensive than nonrefundable ones. We allow for an arbitrary correlation between  $\xi_{jt}$  and prices, and instrument prices. We also allow for the possible endogeneity of flight frequency (measured by the average number of daily departures). As we cannot allow for all product characteristics to be endogenous, we treat a number of them (such as distance and the number of connections) as exogenous.

<sup>13</sup> We iterate until the maximum difference between each iteration is smaller than  $10^{-12}$ :  $\|\xi^M - \xi^{M-1}\|_\infty = \max\{|\xi_1^M - \xi_1^{M-1}|, \dots, |\xi_K^M - \xi_K^{M-1}|\} < 10^{-12}$ . See Jean-Pierre Dubé, Jeremy T. Fox, and Che-Lin Su (2008) for an illuminating discussion of the importance of a stringent convergence rule. That paper also provides a computational algorithm for BLP (1995) that converges faster than the traditional “nested” algorithm that we use here.

<sup>14</sup> In practice, not all products are available at each point of time. For example, discount fares typically require advanced purchase and tend to disappear first. We use  $\xi_j$  to capture a ticket’s availability:  $\xi_j$  is high for products that are always available (or have fewer restrictions), and low for others that are less obtainable (or with more restrictions). Admittedly, this is a rough approximation. However, having an explicit model of the ticket availability when we do not have relevant data on availability across time does not seem appropriate. See Section IVD for further discussions. Monte Carlo results suggest that the bias in our application is likely to be insignificant.



Obviously, the instrument set should include exogenous variables that help to predict endogenous characteristics (prices and flight frequencies). The instruments also have to identify parameters that govern substitution patterns across products in a market, such as the type specific parameters  $\beta_r$  and  $\alpha_r$ ,  $\lambda$ , and the share of each passenger type  $\gamma_r$ . Intuitively, exogenous variation in choice sets across markets greatly helps to identify substitution patterns.<sup>15</sup> Our specific choice of demand instruments (as well as cost instruments) is considered in Section IVB, after we introduce the data in more detail.

Finally, we want to point out that a discrete model with  $r$  types of passengers is a parsimonious way to capture the correlation of tastes for different product attributes. Given the documented fact that some passengers (for example, business travelers) value the convenience of frequent departures and fewer layovers, while other passengers (for example, tourists) are more concerned about prices and less sensitive to differences in flight schedules, it is important to allow for correlations between taste parameters. A continuous random coefficient model requires the estimation of  $k$  means and  $k(k + 1)/2$  covariance elements. A discrete  $r$ -type model involves  $r \times k$  parameters, which are fewer than  $k(k + 3)/2$  if we have many product attributes but a few types. Another advantage of the discrete type model is the convenience of the analytic formula for the share equation, which is much simpler to evaluate than integrating the random coefficients with continuous distributions. Given the size of our dataset (with more than 200,000 products in different markets), the simplicity of an analytical formula dramatically reduces the computational burden of the estimation.

### B. Markups and Marginal Cost

We assume that prices are set according to a static Nash equilibrium with multi-product firms. Following BLP (1995), we compute equilibrium markups from knowledge of the demand data and parameters. Let  $b_{jt}(s_t, x_t, p_t, \theta_d)$  denote these markups. Marginal cost of product  $j$  in market  $t$  is<sup>16</sup>

$$(10) \quad mc_{jt} = p_{jt} - b_{jt}(s_t, x_t, p_t, \theta_d).$$

<sup>15</sup> Berry and Philip A. Haile (2009) consider this argument more formally, in a nonparametric context.

<sup>16</sup> The markup equation in matrix form is

$$MC = P + \left( \frac{\partial Q}{\partial P} \right)^{-1} Q,$$

$$\text{where } Q = (q_{1t}, \dots, q_{J_t t}) = (s_{1t}, \dots, s_{J_t t}) \times M_t, \left( \frac{\partial Q}{\partial P} \right) = \begin{pmatrix} \frac{\partial q_{1t}}{\partial p_{1t}} & \dots & \frac{\partial q_{J_t t}}{\partial p_{1t}} \\ \frac{\partial q_{1t}}{\partial p_{J_t t}} & \dots & \frac{\partial q_{J_t t}}{\partial p_{J_t t}} \end{pmatrix}. J_t \text{ is the number of products by}$$

firm  $f$  in market  $t$ ,  $M_t$  is the market size, and  $s_{jt}$  is defined by equation (5).



We posit a simpler version of marginal costs as compared to BCS (2006). The marginal cost function is given by

$$(11) \quad mc_{jt} = w_{jt}\psi + \omega_{jt},$$

where

- $w_{jt}$  is a vector of observed cost-shifters,
- $\psi$  is a vector of cost parameters to be estimated, and
- $\omega_{jt}$  is an unobserved cost shock.

Our specification differs from BCS (2007), who model marginal cost as depending on a “flexible” functional form in distance and a carrier’s total passenger flow on each segment of a route (“segment density”). This allows BCS (2006) to explain “hub economies” in an explicit model of the economies of scope. However, the presence of the endogenous segment density would render our counterfactual analysis (reported in Section VC) infeasible. It ties together all markets in which the product itineraries share common flight segments, and requires us to solve tens of thousands of prices simultaneously in the counterfactual analysis, which does not appear feasible.

As an alternative, we model the hub density effect through the direct inclusion of the “hub” variable in the marginal cost specification, and we have separate cost parameters for short-haul routes and long-haul routes. We are effectively assuming that the hub structure of a given airline does not change in our counterfactual analysis. The overall change in network structure (together with unmodeled changes in the product set) is left to the “unexplained” residual effect that contributes to carriers’ profit change during the sample period. See Section IVD for more detailed discussions.

Equations (10) and (11) imply that the cost-side unobservable is the difference between prices, markups, and the deterministic part of marginal cost:

$$(12) \quad \omega_{jt} = p_{jt} - b_{jt}(s_t, x_t, p_t, \theta_d) - w_{jt}\psi.$$

As with demand, we form moments that are expectations of the cost-side unobservable  $\omega$  interacted with cost-side instruments:

$$(13) \quad E(h(z_t)\omega(x_t, p_t, s_t, \theta_d, \psi)) = 0,$$

where  $z_t$  is a vector of instruments. These instruments include:

- exogenous elements of the marginal-cost shifters,  $w$ , and
- exogenous demand-side instruments that help to predict the markup term,  $b_{jt}(\cdot)$ .

In addition to estimating the marginal cost parameter  $\psi$ , the supply side restrictions in (13) also help to estimate the demand parameters  $\theta_d$  because these parameters enter the markup term. We allow for an arbitrary dependence between the cost shock

$\omega_{jt}$  and the unobserved product characteristic  $\xi_{jt}$ . We also allow for arbitrary correlations of  $(\xi_{jt}, \omega_{jt})$  among products within the same market. Note, however, that nothing in the estimation method allows us to estimate fixed costs.

### III. Data

There are three main data sources for this study. The Airline Origin and Destination Survey (DB1B), published by the US Department of Transportation (DOT), provides detailed information on flight fares, itinerary (origin, destination, and all connecting airports), the ticketing and operating carrier for each segment, and the number of passengers traveling on the itinerary at a given fare in each quarter.<sup>17</sup> The flight frequency is constructed using the scheduling data from Back Aviation Solutions, Inc. Flight delays are extracted from the Airline On-Time Performance Data, also published by the DOT. In the following section, we explain our market definition and sample selection. See the Appendix for further details.

#### A. Sample Selection

The DB1B data is a 10 percent random sample of airline tickets from US reporting carriers. Following Jan K. Brueckner and Spiller (1994) and BCS (2006), we keep round-trip itineraries within the continental US with at most four segments. We eliminate tickets cheaper than \$25, those with multiple ticketing carriers, or those containing ground traffic as part of the itinerary.

A market is defined as a directional pair of an origin and a destination airport. For example, Atlanta–Las Vegas is a different market from Las Vegas–Atlanta. This allows for the characteristics of the origin city to affect demand. As in BCS (2006), the market size is the geometric mean of the MSA population of the end-point cities.<sup>18</sup>

We focus on airports located in medium to large metropolitan areas with at least 850,000 people in 2006. There were 3,998 such markets in 1999 and 4,300 markets in 2006. These markets accounted for about 80 percent of total passengers, and roughly overlapped with the top 4,000 most traveled markets, which is the scope of focus in many empirical studies.<sup>19</sup>

There are two reasons for excluding small markets. The first one is computational; the estimation time increases substantially with the number of markets and products. The small airports accounted for only one-fifth of the passengers, but they constituted three-quarters of the markets and one-third of products. The main reason for excluding small markets, however, is the drastic difference between large and small markets. Even within our selected sample, the number of passengers and revenues in the largest markets are hundreds of times larger than the smallest ones. As

<sup>17</sup> The URL of the data source is (as of April 2008): <http://www.transtats.bts.gov/DataIndex.asp>.

<sup>18</sup> The data source (as of April 2008) for the MSA population is: <http://www.census.gov/population/www/estimates/CBSA-est2006-annual.html>.

<sup>19</sup> For example, the Government Accounting Office (GAO) focuses on the top 5,000 most traveled markets in their annual report of the airline industry.

demand patterns and operation costs are different among markets with diverse sizes, it is difficult for our stylized model to capture all of these differences.

Six groups of airports are geographically close.<sup>20</sup> Carriers in nearby airports might compete against each other as consumers can choose which airport to fly from. In one of our specifications, we group these nearby airports, and define markets based on the grouped airports.

For 2006, our sample contains 700,000 unique records, or 163 records per market. Given that the product shares need to be inverted at each iteration, both the memory requirement and the estimation time increase substantially with the number of products. In addition, conditioning on observed characteristics, many records have very similar fares (for example, a \$325 ticket and a \$328 ticket) and are not likely to be viewed by consumers as distinctive products. Therefore, we aggregate the records using a set of progressive fare bins conditioning on the itinerary and the ticketing carrier.<sup>21</sup> In summary, our product is a unique combination of the origin airport, the connecting airport, the destination airport, the ticketing carrier, and the binned fare. We have 214,809 products in 1999 and 226,532 products in 2006.

Back Aviation Solutions' schedule data report the departure time and arrival time for all domestic flights. To generate the number of departures for direct flights, we aggregate over all carriers that operate for a ticketing carrier in a given market. The number of departures for connecting flights is route specific. We restrict the connecting time to between 45 minutes and 4 hours. When there are multiple feasible connections, we only include the connection with the shortest layover time.<sup>22</sup> Using other departure measures, such as all feasible connections between 45 minutes to 4 hours, and the minimum number of departures between the two connecting segments, does not make much difference.

To evaluate changes in demand and supply between the late 1990s and the 2000s, we conduct the empirical analysis using two cross-sections of data: the second quarter in 1999 and the second quarter in 2006. We choose 2006 to avoid the few years right after 9/11 when carriers were adjusting to new security regulations.

## B. Data Summary

Table 1A reports the summary statistics of our sample. The top panel displays the mean and standard deviation for all regressors used in the estimation. There are several noticeable changes between 1999 and 2006. The average fare, in 2006 dollars, decreased from \$493 to \$451, a reduction of 8.5 percent. When weighted by the number of passengers travelling on each ticket, the average fare fell by 12 percent during the sample period, from \$371 to \$327. In 1999, 7.6 percent of the products

<sup>20</sup> The six groups of airports are: Dallas-Fort Worth International and Love Field in Dallas, TX; Baltimore/Washington International, Dulles, and National in Washington, DC; Midway and O'Hare in Chicago, IL; Kennedy, LaGuardia, and Newark in New York, NY; Los Angeles, Burbank, and Long Beach in Los Angeles, CA; San Francisco, Oakland, and San Jose in San Francisco, CA.

<sup>21</sup> In the base case specification, we use the following set of bins: \$20 for all tickets under \$700 (so tickets between \$300 and \$320 with the same itinerary and ticketing carrier are aggregated as one product), \$50 for tickets between \$700 and \$1,000, and \$100 for tickets above \$1,000.

<sup>22</sup> The Appendix explains in detail how this variable is constructed.

TABLE 1A—SUMMARY STATISTICS FOR THE DATASET

Variable	1999		2006	
	Mean	SD	Mean	SD
Fare (2006 \$100)	4.94	3.17	4.51	2.59
Product share	1.42E-04	6.37E-04	1.42E-04	5.26E-04
No. connections	1.25	0.97	1.14	0.99
No. daily departures	4.42	2.77	4.18	2.40
No. destinations (100 cities)	0.17	0.28	0.19	0.31
Hub	0.17	0.37	0.17	0.37
HubMC	0.87	0.34	0.78	0.41
Distance (1,000 miles)	2.73	1.40	2.78	1.42
Distance <sup>2</sup> (10 <sup>6</sup> miles)	9.42	8.44	9.72	8.66
Tourist place (FL/LAS)	0.13	0.33	0.13	0.34
Slot-control	0.36	0.76	0.35	0.75
SlotMC	0.21	0.41	0.20	0.40
Plane size (100)	1.35	0.33	1.23	0.34
Delay >= 30 minutes	0.14	0.07	0.13	0.07
American	0.16	0.37	0.18	0.39
Continental	0.10	0.29	0.08	0.28
Delta	0.19	0.39	0.15	0.36
American West	0.05	0.22		
Northwest	0.09	0.28	0.08	0.28
Trans World	0.09	0.28		
United	0.13	0.34	0.14	0.34
US Airways	0.10	0.30	0.15	0.36
JetBlue			0.01	0.12
Southwest	0.04	0.20	0.09	0.29
Other carrier	0.05	0.22	0.11	0.31
Observations	214,809		226,532	
<b>Market average</b>				
No. products	53.73	38.52	52.68	36.67
No. carriers	3.51	2.00	3.30	1.88
No. direct passengers (1,000)	20.13	40.45	22.75	43.66
No. connecting passengers (1,000)	3.52	4.10	2.71	3.13
No. markets with LCC entry			1,569	
No. markets	3,998		4,300	

Notes: Hub = 1 if the origin airport is a hub for the carrier. HubMC = 1 if either the origin, the connecting airport, or the destination is a hub for the carrier. Distance is the round-trip distance covered by the entire route of product  $j$ . Tourist Place = 1 if the origin airport is in Las Vegas or Florida. Slot-control is the number of slot-controlled airports the route of product  $j$  passes through. SlotMC = 1 if Slot-control > 0. Delay is the percentage of flights arriving more than 30 minutes later than the scheduled arrival time.

were priced above \$1,000; the fraction fell to 4 percent in 2006. Figure 3 and Figure 4 plot the fare density for direct and connecting flights, respectively. Compared to 1999, fares of connecting flights were lower at each quantile of the distribution in 2006. For direct flights, the fraction of both high-fare products ( $\geq$  \$1,000) and low-fare products ( $\leq$  \$200) shrank, while that of medium fare products increased. Not surprisingly, the average fare for connecting flights dropped by 12 percent, while the average fare for direct flights fell only by 4 percent.

The second pronounced development is the increasing number of direct passengers. Figure 5 displays the percentage of US domestic passengers on direct flights from 1995 to 2006. It varied between 63 percent and 64.5 percent from 1995 to 2001, and has steadily increased since then. By 2006, more than 67.3 percent of passengers traveled on direct flights. In our sample markets, the average number of direct passengers in a market increased by 13 percent from 1999 to 2006, while that of connecting passengers diminished by 23 percent.

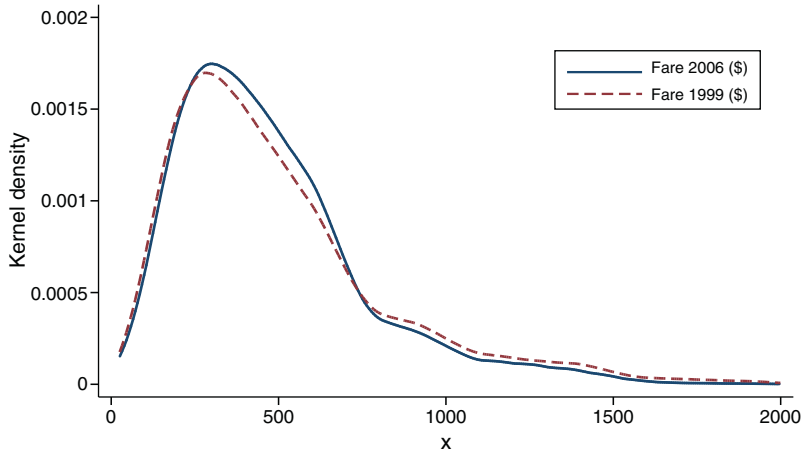


FIGURE 3. DIRECT FLIGHTS' FARE DISPERSION 1999–2006

*Note:* Calculation based on the sample markets.

*Source:* US DOT DB1B via BTS

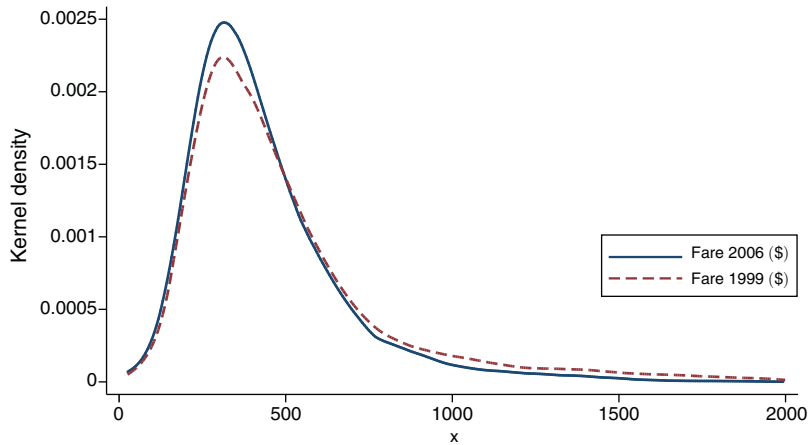


FIGURE 4. CONNECTING FLIGHTS' FARE DISPERSION 1999–2006

*Note:* Calculation based on the sample markets.

*Source:* US DOT DB1B via BTS

As shown in Figure 6, the trend away from connecting flights is almost universal—all legacy carriers flew fewer connecting passengers in 2006 (except for Continental).<sup>23</sup> American and Delta experienced the largest reduction, with the total number of connecting passengers decreasing by 29 percent and 40 percent from 1999 to 2006, respectively.

<sup>23</sup> After the domestic code-share agreement with Northwest in 1999, Continental started to issue connecting tickets with part of the route operated by Northwest, which led to a slightly higher fraction of connecting passengers in 2006.

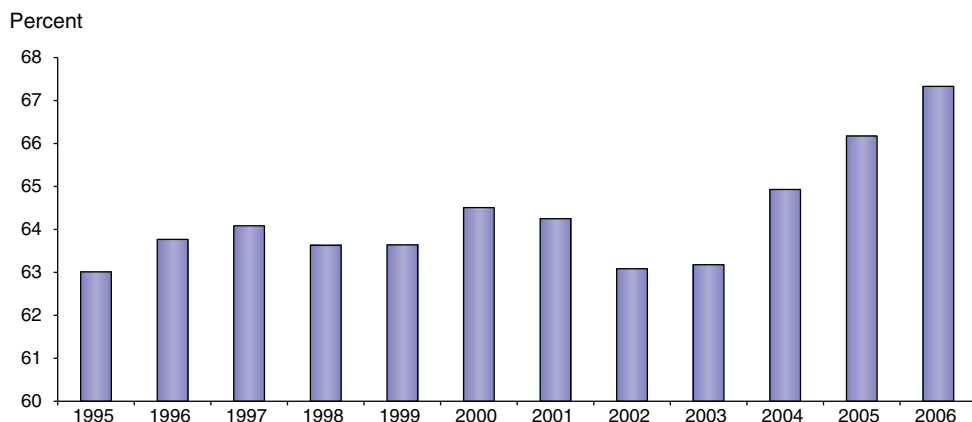


FIGURE 5. PERCENTAGE OF DIRECT PASSENGERS IN US

Source: US DOT DB1B via BTS and author’s calculation

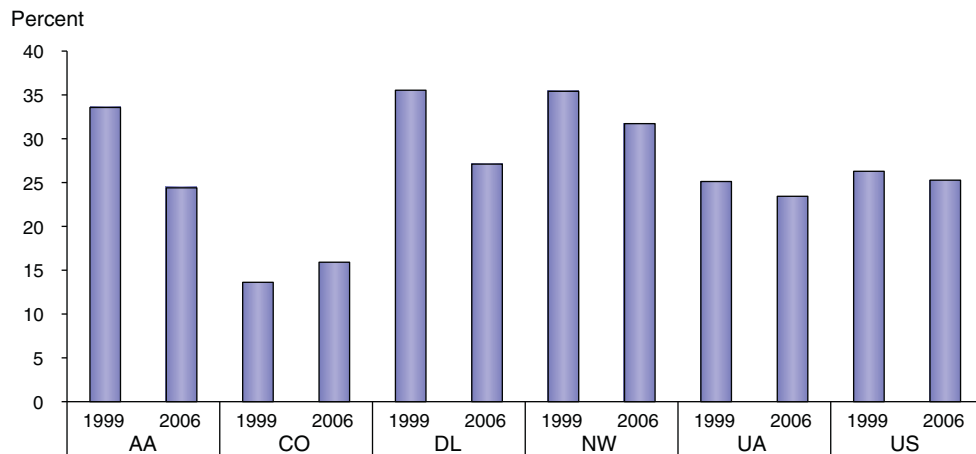


FIGURE 6. PERCENTAGE OF CONNECTING PASSENGERS BY CARRIER—1999 AND 2006

Note: Calculation based on the sample markets.

Source: US DOT DB1B via BTS

The declining number of connecting passengers during the sample period appears to be closely related to the recent “dehubbing” phenomenon in the airline industry. For example, Delta closed its hub in Dallas-Fort Worth International airport in January 2005, and cut 26 percent of flights at its Cincinnati hub in September 2005. US Airways downgraded Pittsburgh from a hub to a “focus city” in 2004.<sup>24</sup> By October 2007, it had reduced the daily departures out of Pittsburgh from over 500 in

<sup>24</sup> In the airline industry, a focus city is a location that is not a hub, but from which an airline has nonstop flights to multiple destinations other than its hubs.

TABLE 1B—SUMMARY STATISTICS BY LEGACY CARRIERS AND LCCs

Variable	1999		2006	
	Legacy	LCC	Legacy	LCC
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Fare (2006 \$100)	5.04 (3.23)	3.65 (1.84)	4.73 (2.73)	3.59 (1.53)
Product share	1.18E-04 (0.00)	4.43E-04 (0.00)	1.10E-04 (0.00)	2.83E-04 (0.00)
No. connections	1.28 (0.96)	0.89 (0.99)	1.18 (0.98)	1.00 (1.00)
No. daily departures	4.49 (2.75)	3.46 (2.90)	4.44 (2.37)	3.04 (2.18)
No. destinations (100 cities)	0.19 (0.28)	0.03 (0.06)	0.23 (0.34)	0.05 (0.10)
Hub	0.18 (0.38)	0.05 (0.22)	0.19 (0.40)	0.05 (0.21)
HubMC	0.91 (0.28)	0.29 (0.45)	0.88 (0.32)	0.34 (0.47)
Distance (1,000 miles)	2.78 (1.41)	2.21 (1.10)	2.79 (1.45)	2.71 (1.29)
Tourist place (FL/LAS)	0.13 (0.33)	0.15 (0.36)	0.12 (0.33)	0.16 (0.37)
Slot-control	0.39 (0.78)	0.06 (0.32)	0.40 (0.79)	0.17 (0.51)
SlotMC	0.22 (0.42)	0.04 (0.20)	0.22 (0.42)	0.11 (0.31)
Plane size (100)	1.36 (0.34)	1.30 (0.16)	1.22 (0.37)	1.31 (0.12)
Delay $\geq$ 30 minutes	0.15 (0.07)	0.06 (0.07)	0.13 (0.07)	0.12 (0.06)
Observations	198,706	16,103	184,267	42,265

Notes: Hub = 1 if the origin airport is a hub for the carrier. HubMC = 1 if either the origin, the connecting airport, or the destination is a hub for the carrier. Distance is the round-trip distance covered by the entire route of product  $j$ . Tourist Place = 1 if the origin airport is in Las Vegas or Florida. Slot-control is the number of slot-controlled airports the route of product  $j$  passes through. SlotMC = 1 if Slot-control  $>$  0. Delay is the percentage of flights arriving more than 30 minutes later than the scheduled arrival time. LCCs include AirTran, Frontier, JetBlue, Southwest, and a few smaller airlines like Spirit, ATA, and Sun Country. Columns 1 and 3 include legacy carriers as well as Alaska, Midwest, and a few other small carriers that are not classified as LCCs.

2000 to fewer than 70, and canceled service to more than 90 destination cities. With a few exceptions, most hubs serviced fewer connecting passengers in 2006 than in 1999.

As a result of the increasing number of direct flights, the average number of destination cities served by direct flights out of the origin airport increased from 17 to 19. The average number of daily departures dropped from 5.3 to 4.8, due to the carriers' recent capacity reduction. The average plane size fell from 135 seats to 123 seats, which reflected the increasing penetration of regional jets. All together, the six legacy carriers offered 77–78 percent of the products, and accounted for 66 percent of passengers in 1999 and 61 percent of the passengers in 2006.



TABLE 1C—SUMMARY STATISTICS BY MARKETS WITH OR WITHOUT LCC ENTRY

Variable	1999		2006	
	Markets with entry	Markets without entry	Markets with entry	Markets without entry
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Fare (2006 \$100)	5.08 (3.19)	4.82 (3.15)	4.42 (2.52)	4.60 (2.66)
Product share	1.11E-04 (0.00)	1.68E-04 (0.00)	1.29E-04 (0.00)	1.55E-04 (0.00)
No. connections	1.39 (0.92)	1.14 (0.99)	1.26 (0.97)	1.03 (1.00)
No. daily departures	4.07 (2.27)	4.70 (3.10)	3.76 (2.12)	4.58 (2.59)
No. destinations (100 cities)	0.16 (0.26)	0.19 (0.29)	0.16 (0.27)	0.23 (0.35)
Hub	0.14 (0.34)	0.19 (0.39)	0.12 (0.33)	0.21 (0.41)
HubMC	0.88 (0.33)	0.86 (0.35)	0.74 (0.44)	0.82 (0.38)
Distance (1,000 miles)	3.11 (1.32)	2.42 (1.38)	3.13 (1.34)	2.43 (1.41)
Tourist place (FL/LAS)	0.17 (0.38)	0.09 (0.29)	0.17 (0.37)	0.09 (0.29)
Slot-control	0.31 (0.73)	0.40 (0.78)	0.30 (0.71)	0.41 (0.79)
SlotMC	0.17 (0.38)	0.24 (0.43)	0.17 (0.38)	0.23 (0.42)
Plane size (100)	1.40 (0.30)	1.32 (0.35)	1.31 (0.29)	1.16 (0.37)
Delay > = 30 minutes	0.14 (0.06)	0.15 (0.08)	0.12 (0.06)	0.13 (0.07)
Observations	97,197	117,612	111,730	114,802

*Notes:* Hub = 1 if the origin airport is a hub for the carrier; HubMC = 1 if either the origin, the connecting airport, or the destination is a hub for the carrier. Distance is the round-trip distance covered by the entire route of product  $j$ . Tourist Place = 1 if the origin airport is in Las Vegas or Florida. Slot-control is the number of slot-controlled airports the route of product  $j$  passes through. SlotMC = 1 if Slot-control > 0. Delay is the percentage of flights arriving more than 30 minutes later than the scheduled arrival time. Columns 1 and 3 are for markets that experienced LCC entry during the sample period, while the other two columns are for markets that did not experience LCC entry.

The bottom panel of Table 1A documents the market average summary statistics. Both the number of products and the number of carriers per market declined slightly.<sup>25</sup> During the sample period, 39 percent of the markets experienced LCC entry.

We report the summary statistics by legacy carriers and LCCs separately in Table 1B.<sup>26</sup> Most of the differences are expected. Flights by LCCs are cheaper, with fewer connections, and are less frequent. Table 1C documents differences between markets

<sup>25</sup> This is probably a consequence of two major mergers in the 2000s: American merged with Trans World in 2001, and American West merged with US Airways in 2005.

<sup>26</sup> In Table 1B, the LCCs include AirTran, Frontier, JetBlue (in 2006), Southwest, and a few smaller ones like ATA, Spirit, and Sun Country. We grouped the legacy carriers with Alaska, Midwest, and a few other smaller carriers that are not classified as LCCs in columns headed "legacy."

that did not experience LCC entry and markets that had LCC entry. Compared to the rest of the markets, flights in markets with LCC entry had more connections and less frequent departures. Fares were higher in 1999 before entry, but were lower in 2006 after entry took place.

#### IV. Empirical Model

##### A. Model Specification

As mentioned in Section II, there is a well-documented correlation between price sensitivity and preference for convenience (few connections and frequent departures). Therefore, we allow three type-specific parameters: a constant, the fare coefficient, and the coefficient of the number of connections. We find that it is important to have a type-specific constant, which allows the model to fit the aggregate shares for both expensive and inexpensive tickets.<sup>27</sup>

We have spent a considerable amount of time experimenting with three or more types of passengers, without much success. The demand parameters common to all types are fairly robust, but the type-specific parameters and  $\lambda$  appear to be sensitive to small changes in the model's specification or the choice of instruments. Sometimes multiple parameter vectors deliver a similar fit for the data. Our conclusion is that the limited variation in the instruments prevents us from estimating an overly flexible model. Our two types of passengers can be described as tourists and business travelers.

We have also tried to model carriers' choices of flight frequencies together with pricing decisions, but faced three major challenges. First, some carriers mix different aircraft on the same route. For example, large jets are typically reserved for dense traffic during peak time, while smaller regional jets or turbo planes are often used for off-peak flights. Second, it is difficult to measure flight frequencies for connecting flights, which affects our ability to estimate marginal revenues generated by an additional departure. Lastly, to model how carriers balance between larger planes with fewer flights and smaller planes with more frequent flights, we need information on the type of the aircraft used, the flight schedule, and the number of passengers on each flight. Given our lack of such detailed data, we instrument flight frequencies without explicitly modeling how departures are determined. The exercise of modeling departures directly is left for future research.

In our base specification, demand is affected by the following product attributes: fares, the number of total connections round trip, the number of destinations,<sup>28</sup> the average number of daily departures, the round-trip distance (in thousand miles) covered by the entire route of the flight, distance squared, a tour dummy for airports in Florida and Las Vegas, the number of slot-controlled airports that the flight passes

<sup>27</sup> We also estimated the model with type-specific coefficients for flight frequencies and the tour dummy. The parameters were similar across types, and there was not much improvement in the model's fit.

<sup>28</sup> A product's number of destinations is the total number of cities to which its ticketing carrier serves direct flights from the origin airport.

through,<sup>29</sup> and carrier dummies.<sup>30</sup> Prices and the number of daily departures are endogenous and instrumented. We include the following regressors in the marginal cost equation: a constant, the round-trip distance in thousand miles, the number of connections, a hub dummy (equal to 1 if the flight departs from, transfers at, or lands at the carrier's hub airport), a slot dummy (equal to 1 if the flight passes through a slot-controlled airport), and carrier dummies. As different aircraft are used for short-medium haul routes and long haul routes, we allow two sets of cost parameters: one for markets shorter than 1,500 miles, and the other for markets longer than 1,500 miles.<sup>31</sup> We have also estimated eight other different specifications, which are discussed in detail in Section VA.

### B. Instruments

As is typical of demand studies with endogenous prices, we need instruments to identify the fare coefficients. One common strategy is to exploit the rival product attributes and the competitiveness of the market environment. All else being equal, products with closer substitutes have lower prices. A standard instrument is the number of products. In our data, the number of products in a market varies from 3 to 223, with an average of 53. However, we are concerned about the endogeneity of this variable because of the way it is constructed. A product is a group of tickets whose fares fall in a fixed bin. By construction, a market with a wider price dispersion has a larger number of products. Similar concerns extend to using rival product attributes as instruments.<sup>32</sup> We use the route level characteristics instead. For example, if there are 5 carriers offering 12 different routes, then the average route level characteristics is the average over these 12 routes. Our instruments along this line include the percentage of rival routes that offer direct flights, the average distance of rival routes, the number of rival routes, the number of all carriers, etc.<sup>33</sup>

A second identification strategy searches for variables that affect costs, but do not affect demand. One candidate is whether the destination is a hub for the ticketing carrier. It affects the marginal cost of a flight because larger and more fuel efficient planes can be used on routes with denser traffic, but is excluded from demand.<sup>34</sup>

<sup>29</sup> Four airports were under the slot control during the sample period: LaGuardia and Kennedy in New York, National in Washington, DC, and O'Hare in Chicago.

<sup>30</sup> In 1999, we included carrier dummies for American (the default group), American West, Continental, Delta, Northwest, Trans World, United, US Airways, Southwest, and a dummy for all other carriers. In 2006, we added a dummy for JetBlue (which started operation in 2000), and excluded dummies for American West (merged with US Airways in 2005) and Trans World (merged with American in 2001).

<sup>31</sup> The hub, slot, and carrier dummies are restricted to be same for both short-medium haul and long haul markets.

<sup>32</sup> For example, with a wide price dispersion and a large number of products, the sum of rival product attributes will be high as well.

<sup>33</sup> Some people might argue that if part of  $\xi_j$  includes ticket restrictions, and if ticket restrictions respond to rival attributes, then the validity of using rival attributes as instruments would be questionable. However, in the airline industry, carriers typically offer all levels of restrictions in all markets (for example, refundable and nonrefundable tickets are available in all markets). In other words, the "menu" of ticket restrictions seems more or less the same across markets. Besides the ticket restrictions, some components of  $\xi_{jt}$  capture the gate locations, the aircraft fleets, etc., that should be exogenous in the short run.

<sup>34</sup> Consumers value the hub status of the origin airport because of the frequent flier programs, convenient flight schedules, or easy gate/parking access. Most of these considerations do not apply to the destination airport. As a

The number of cities to which a carrier flies nonstop flights from the destination airport, which reflects the carrier's size of operation at the destination airport, serves a similar role. We also include a dummy for transferring at the hub, using similar arguments that costs are lower if the flight connects at a hub.<sup>35</sup>

The third group of instruments includes the fitted values of the twenty-fifth and the seventy-fifth quantile of fares in a given route.<sup>36</sup> As documented by Borenstein and Rose (1994, 2007), there is a wide fare dispersion across passengers travelling on the same route. The twenty-fifth and the seventy-fifth fitted fare quantiles are nonlinear functions of the exogenous route characteristics, and allow us to parsimoniously capture the price dispersion.

To construct instruments for flight frequencies, we first regress segment departures on characteristics of the end cities,<sup>37</sup> and then include the fitted segment departures as instruments.

The last group of instruments is the exogenous variables that directly enter the share equation (5) and the marginal cost equation (10). Finally, we include the interaction terms of the above variables provided they are not highly collinear.

### C. Identification

The identification of most parameters is straightforward. Here, we focus on  $\lambda$  and the type-specific parameters.  $\lambda$  is identified from changes in the aggregate market share when the number of products varies. In the extreme case of  $\lambda = 0$ , all products are perfect substitutes. The aggregate share remains fixed as the number of products changes, as long as the "best product" does not change. On the other hand, if  $\lambda = 1$ , the nested logit demand is reduced to a simple logit, and the aggregate market share is close to  $K/(K + 1)$  if there are  $K$  products with similar product attributes. Identification of the type-specific parameters follows from the random coefficient literature, as our model is a special case where the random coefficients take discrete values. Briefly, these type-specific parameters are identified from the substitution patterns among similar products when the mix of products varies across markets.

### D. Model Limitations

A distinctive feature of the airline industry is the prevalent practice of revenue management, which reserves certain seats for different fare categories. Consequently,

---

robustness check, we have also estimated the model without using the hub destination as an instrument. There is little change in the parameter estimates. Results are available upon request.

<sup>35</sup> There are two countervailing factors that affect the marginal cost of a flight that passes through a hub, or an airport with a large carrier presence. On the one hand, economies of scale reduce costs. On the other hand, congestion and delays associated with denser traffic tend to increase costs. In either case, these variables are valid instruments.

<sup>36</sup> The fitted fares are obtained from quantile regressions of fares on the following exogenous variables: carrier dummies, segment and route level characteristics (distance, difference in January temperatures between the end cities, whether or not in tourist locations, etc.), market size (measured by population), number of competitors, and the carrier's share of cities connected via nonstop flights at both the origin and the destination airport.

<sup>37</sup> The regressors that predict segment departures are similar to those in the fare quantile regression, except that we also include the hub status of both end cities.

cheaper tickets tend to disappear faster than expensive ones. As discussed in Section IIA, we do not have appropriate data to explicitly model missing products, and use  $\xi_{jt}$  to proxy for a product's availability. In Appendix Section II, we consider the impact of ignoring missing products on our parameter estimates, and discuss conditions under which the bias would be insignificant. Roughly speaking, the bias in the parameter estimates is small if the share of the outside good is large, the share of the missing products is small, or the covariance between the error in the computed  $\xi_{jt}$  and instruments is low. In most of our Monte-Carlo simulations designed to replicate the real data, the bias in the price coefficient varies from 0.1 percent to 0.5 percent. In our application, all of these factors are at play: the market share of the outside good exceeds 99 percent; each market has on average 53 products, with the share of any given product being fairly small; and our route-level instruments are unlikely to be highly correlated with product-specific approximation errors. These pieces of evidence point to a moderate bias in our parameter estimates.

An implicit assumption of our empirical model is that the network structure and the carriers that serve each market are taken as given. Ideally, we would like to model a three-stage game. First, carriers form their hubs. Second, given the hub structure, each carrier chooses a set of markets to serve. Third, given these entry decisions, carriers compete in prices and the frequency of flight departures. However, solving this game with a dozen carriers and thousands of markets is beyond our capability.

Alternatively, we focus on the last stage of the game and model consumers' choices of different products as well as carriers' price decisions. We instrument prices and departures using variables related to the network configuration, like the hub-spoke structure and the number of carriers. The hubs were formed mostly in the 1980s and 1990s. Market entry decisions involve acquiring gate access, optimizing flight schedules, obtaining aircraft and crew members, and advertising to consumers, all of which entail substantial fixed costs. The fact that capacity reduction is costly and that carriers are, in general, cautious about serving a new market suggests that the number of carriers is likely to be determined by long-term considerations and is uncorrelated with temporal demand shocks. This is admittedly a strong assumption, but is analogous to the standard assumption in the discrete-choice demand literature that variation in the set of available products and the number of firms across markets is exogenous (at least in the short run).

As we do not observe the day-to-day variation in fares and flight availability, we do not allow consumers to choose the date of purchase strategically. We also rule out the dynamic considerations in firms' pricing decisions.<sup>38</sup> Modeling this dynamic aspect is a difficult but promising topic. See Ciliberto (2008) for an interesting study on strategic deterrence in the airline industry.

Lastly, we do not observe the fixed cost of operating a flight, which limits our ability to estimate changes in net profit. Our profit estimates should be interpreted as variable profits.

<sup>38</sup> In light of these concerns, we also present demand estimates without using the supply side of the model (see Tables 3 and 4). We find similar patterns as in the base case specification.

## V. Results

The parameters from the base specification are presented first, followed by results from eight other specifications. The profit estimates are discussed next. Finally, we report results from the counterfactual exercises that isolate the effects of changes in demand, supply, and competition on legacy carriers' profits.

### A. Parameters

*Demand Parameters.*—As discussed in Section IVA, in our base specification, demand is affected by fares; the number of connections, destinations, and average daily departures; distance and distance squared; a tour dummy; the number of slot-controlled airports; and carrier dummies.

We expect consumers' utility to decrease with the number of connections. The number of destination cities captures the value of frequent flier programs. The larger the number of cities for which consumers can redeem frequent flier miles, the higher the value of these loyalty programs. In addition, a carrier that flies to many destination cities is likely to have more convenient gate access and offer better service.

Air-travel demand is usually U-shaped in distance. Short-haul flights compete with cars and trains, which become worse substitutes as distance increases, so demand initially grows with distance. As distance increases further, travel becomes less pleasant and demand starts to decrease. We include both distance and distance squared to capture the curvature of demand.

The tour dummy helps to fit the relatively high traffic volume in Florida and Las Vegas that cannot be explained by the observed product attributes. The slot variable captures the potential negative effect of congestion in slot controlled airports on air-travel demand.

The first two columns in Table 2 present the parameters for the base specification in 1999 and 2006, respectively. Most parameters are precisely estimated. Consistent with the story of the dot-com bubble bursting and the introduction of online ticketing sites, demand was more price sensitive in 2006. The price coefficient of tourists (labeled as type 1 in Tables 2, 3, and 4) increased in absolute value from 0.78 to 1.05, and the price coefficient of business travelers (labeled as type 2) rose from 0.07 to 0.10. In both cases, the differences between the two periods are statistically significant. The price elasticity was 31 percent larger for tourists and 43 percent larger for business travelers. In the meantime, the estimated percentage of business travelers rose from 41 percent to 49 percent, which moderated the increase in the overall price sensitivity of demand. The aggregate price elasticity, which is the percentage change in total demand when all products' prices increase by 1 percent, was 1.55 in 1999, and rose to 1.67 in 2006. David W. Gillen, William G. Morrison, and Christopher Stewart (2003) conducted a survey that collected 85 demand elasticity estimates from cross-sectional studies.<sup>39</sup>

<sup>39</sup> Out of these 85 estimates, 80 were taken from Tae H. Oum, Gillen, and S. F. Noble (1986) and represented US city-pair routes. All 85 studies were conducted between 1981 and 1986 and are slightly dated.

TABLE 2—BASE CASE PARAMETER ESTIMATES—1999 AND 2006

Demand variables	1999	2006	Cost variables	1999	2006
Fare 1	-0.78** (0.02)	-1.05** (0.03)	Constant_short	1.07** (0.06)	1.16** (0.06)
Connection 1	-0.53** (0.02)	-0.59** (0.03)	Distance_short	0.26** (0.01)	0.19** (0.01)
Constant 1	-5.79** (0.19)	-5.68** (0.19)	Connection_short	-0.06** (0.03)	0.07* (0.04)
Fare 2	-0.07** (0.00)	-0.10** (0.00)	Constant_long	1.61** (0.08)	1.59** (0.07)
Connection 2	-0.31** (0.02)	-0.51** (0.02)	Distance_long	0.09** (0.01)	0.04** (0.01)
Constant 2	-8.56** (0.40)	-8.60** (0.30)	Connection_long	-0.09** (0.03)	0.06 (0.04)
No. destination	0.38** (0.03)	0.27** (0.02)	HubMC	-0.02 (0.01)	-0.05** (0.01)
No. departures	0.04** (0.00)	0.11** (0.00)	SlotMC	0.08** (0.01)	0.03** (0.01)
Distance	0.30** (0.04)	0.53** (0.04)			
Distance <sup>2</sup>	-0.05** (0.01)	-0.08** (0.01)			
Tour	0.30** (0.03)	0.36** (0.03)			
Slot-control	-0.19** (0.01)	-0.18** (0.01)			
$\lambda$	0.77** (0.01)	0.72** (0.01)			
$\gamma$	0.69** (0.12)	0.63** (0.11)			
Demand carrier dummy			Cost carrier dummy		
Other carriers	-0.18** (0.03)	0.06** (0.02)	Other carriers	-0.03 (0.02)	-0.22** (0.02)
American West	-0.19** (0.03)		American West	-0.22** (0.02)	
Continental	-0.22** (0.02)	0.07** (0.02)	Continental	-0.03** (0.01)	-0.19** (0.01)
Delta	-0.13** (0.02)	-0.21** (0.02)	Delta	-0.10** (0.01)	-0.15** (0.01)
Northwest	-0.15** (0.02)	0.07** (0.02)	Northwest	-0.02 (0.01)	-0.04** (0.01)
Trans World	-0.17** (0.02)		Trans World	0.02 (0.01)	
United	0.16** (0.02)	0.08** (0.02)	United	-0.05** (0.01)	-0.06** (0.01)
US Airways	-0.19** (0.03)	0.06** (0.02)	US Airways	-0.08** (0.01)	-0.11** (0.01)
JetBlue		0.39** (0.06)	JetBlue		-0.32** (0.04)
Southwest	-0.05 (0.03)	0.08** (0.02)	Southwest	-0.12** (0.02)	-0.19** (0.02)
Function value	49.37	58.07			
Observations	214.8k	226.5k			

Notes: See Table 1A for the variable definitions. Standard errors are in parentheses.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.



The elasticities ranged from 0.181 to 2.01, with a median of 1.33. Our estimates seem quite reasonable.

Both tourists and business passengers exhibited a stronger preference for direct flights in 2006. The connection semi-elasticity, or the percentage reduction in demand when a direct flight becomes a connecting flight, jumped from 0.55 to 0.75 for business travelers, and from 0.75 to 0.80 for tourists. Combining both groups, the average connection semi-elasticity increased from 66 percent to 77 percent. In other words, the number of passengers on a direct flight would fall by almost four-fifths when a layover was added to its itinerary, holding the attributes of all other products (of rival carriers as well as the specified carrier) fixed. If all products became connecting flights, aggregate passenger traffic would decline by 27 percent in 1999 and 34 percent in 2006.<sup>40</sup>

These two results—a higher price sensitivity and a higher aversion toward connecting flights—are the most pronounced findings of changes in demand, and are present in all specifications that we have estimated. Both findings are supported by the data patterns (fare reductions and a smaller number of connecting passengers) documented in Section IIIB. While a fare reduction could also be rationalized by increasing competition or decreasing costs, the fact that fares dropped in markets with and without LCC entry, and that fares fell more for connecting flights that became more costly to operate,<sup>41</sup> provide ample evidence of a demand change during our sample period.

As we do not model carriers' choice of hub airports, we cannot examine how changes in demand affect the hub structure. However, it seems quite possible that the reduced demand for connecting flights is directly related to the recent hub downsizing phenomenon. For example, when Delta reduced the capacity that mostly served connecting passengers at its Cincinnati hub in 2005, it claimed that "connecting traffic is the least profitable for the airline."<sup>42</sup>

Many previous studies have pointed out the existence of a hub premium; carriers are able to charge higher fares for hub-originating flights, either because they offer more convenient gate access, or the frequent flier program is more valuable at hub airports. Borenstein (2005) and Borenstein and Rose (2007) pointed out that the hub premium declined over the past several years. Our parameter estimates are consistent with their findings. The coefficient of the number of destinations, which we use to capture a carrier's presence at the airport, dropped from 0.38 to 0.27. The result is very similar if we replace the number of destinations with the hub dummy. Either loyalty programs have become less valuable, or the difference in service between hub airports (or airports with a large carrier presence) and nonhub airports (or airports with a small carrier presence) has narrowed.

All other demand parameters have the expected signs. For example, demand increases in distance up to 1,600 miles (one-way) and then decreases. Tourist places

<sup>40</sup> The 34 percent change seems plausible, given that most markets with only connecting flights have very small passenger traffic. In any case, we want to be cautious about taking this number too literally, because it is an out-of-sample prediction that is very far from the observed data (two-thirds of the passengers travel on direct flights).

<sup>41</sup> See the next page for discussions on changes in marginal costs.

<sup>42</sup> Source: *Business Courier*, September 7, 2005.

attract more consumers, and flights through slot controlled airports have fewer passengers.

The business traveler accounted for 41 percent and 49 percent of total passengers in 1999 and 2006, respectively (see the third panel in Table 7). According to the 2001–2002 National Household Travel Survey, roughly 39 percent to 47 percent of air travel is taken for business purposes, depending on whether personal business trips are treated as business trips.<sup>43</sup> Our model's predictions match closely with the survey.

Interestingly,  $\lambda$  decreased from 0.77 in 1999 to 0.72 in 2006, which suggests that products became closer substitutes. It probably reflects the reduced differentiation among flights offered by different carriers as they cut down services and competed more intensively on prices.

Overall, the carrier dummies are broadly consistent with news reports. In 1999, American (the omitted group in Table 2) and United had the highest parameter values. They were also the most popular and successful carriers in the late 1990s. During the sharp downturn following 9/11, the legacy carriers, especially American and Delta, began to shift capacity to more lucrative international markets. These structural changes are reflected in their negative carrier dummies in 2006. The JetBlue dummy has a large positive coefficient, which is consistent with its popularity due to good on-time performance, new planes, free TV programs, etc. In fact, by 2006, it had been voted the number one US domestic airline by *Condé Nast Traveler* five years in a row.<sup>44</sup>

*Marginal Cost Parameters.*—Columns 3 and 4 in Table 2 report the marginal cost parameters, which include a constant, distance, the number of connections, a hub dummy, a slot dummy, and carrier dummies. We allow two sets of cost parameters: one for markets shorter than 1,500 miles round trip, and the other for markets longer than 1,500 miles.

Two offsetting factors affect the marginal cost of connecting flights. On the one hand, by channeling passengers from different origins and to different destinations through the connecting airport, carriers can generate denser traffic, increase the load factor, and defray costs with more passengers. On the other hand, a large fraction of the fuel is consumed at the landings and takeoffs. With an extra landing and takeoff and a longer travel distance, the fuel component of a connecting flight's marginal cost is much larger than that of a direct flight. The connection coefficient reflects the net effect of these two countervailing factors. The same economies-of-scale argument for connecting flights also applies to flights at the hub airports that tend to have denser traffic. Costs are probably higher at slot controlled airports due to higher landing fees, etc.

<sup>43</sup> The National Household Travel Survey was conducted on 26,000 households. According to the survey, 56 percent of trips longer than 50 miles were taken for pleasure, 16 percent for business, 13 percent each for commuting and for personal business (trips taken for family, personal, religious, or medical reasons), and 3 percent for other reasons. Air travel accounted for 7 percent of pleasure trips, 18 percent of business trips, 5 percent of personal business trips, and none of the commuting trips.

<sup>44</sup> Data source: [http://en.wikipedia.org/wiki/JetBlue\\_Airways#Awards](http://en.wikipedia.org/wiki/JetBlue_Airways#Awards).

The most noticeable difference between 1999 and 2006 is the connection coefficient, which changed signs during the sample period. In 1999, there was evidence of scale economies for connecting flights. Controlling for other variables, the marginal cost of serving a connecting passenger on a long route was \$18 less than that of a direct passenger, or roughly 12 percent of the average marginal cost. Unlike BCS (2006), who reported the existence of scale economies only on longer routes, our estimated marginal cost of connecting flights was lower on both long and short-medium routes in 1999.

The cost advantage of connecting flights disappeared in 2006. Controlling for other cost shifters, the marginal cost of a connecting flight was \$12 more expensive than that of a direct flight. The change was probably driven by increasing fuel costs in the sample period. Since the fraction of fuel consumed at takeoff and landing could be as high as 40 percent, rising fuel costs offset the benefit of denser traffic created by connecting flights.

All other parameters (except for the carrier dummies) are similar between the two periods, with the expected signs. Marginal cost increases with distance, and is higher for routes that pass through slot-controlled airports. Flights through hubs have a lower marginal cost.

The distance coefficient was smaller in 2006, which is puzzling given the higher fuel cost. The change probably reflects a combination of several factors, including reduced services and improved fuel efficiencies.

As expected, JetBlue and Southwest had lower marginal costs than the legacy carriers. Interestingly, American West also had a smaller marginal cost than the legacy carriers. According to US DOT Form 41, its total operating cost per available seat mile (CASM) was the lowest among all legacy carriers in the late 1990s and early 2000s. The coefficient of the Continental dummy in 2006 is comparable to that of Southwest, which seems an anomaly. These dummy variables presumably reflect various carrier specific factors that are not captured by the model.

As in most empirical studies, marginal cost is not directly observed. The parameters are identified from a “residual” regression where we “regress” the difference between the price and markup on cost instruments. To examine the sensitivity of the marginal cost parameters to the overidentifying restrictions, we regress the predicted marginal cost (the difference between prices and the estimated markup) on variables that affect marginal cost directly. The coefficients from this OLS regression are similar to the estimates from our full model, which suggests the robustness of the marginal cost instruments.<sup>45</sup>

Finally, we compare our cost estimates with the carriers’ reported operating costs per available seat mile in DOT Form 41. The average in 2006 dollars was 11.4 cents in 1999 and 12.5 cents in 2006. Our estimated marginal cost per mile is around 6 cents, about half of the average reported operating costs, which seems plausible.

<sup>45</sup> We tested the parameter differences using a standard Hausman test. All of them are insignificant, except for a couple of carrier dummies. Since there are no endogenous regressors in the marginal cost equation, the OLS estimates are consistent (although less efficient). The similarity between the OLS estimates and the estimates from our full model provides evidence that these results are not sensitive to our choice of instruments.

TABLE 3—PARAMETER ESTIMATES FROM DIFFERENT SPECIFICATIONS—1999

	Base case	No. MC	Delay	Combine airports	Small bin	Large bin	Airport dummy
Demand variables							
Fare 1	-0.78** (0.02)	-1.14** (0.49)	-0.78** (0.02)	-0.80** (0.02)	-0.80** (0.02)	-0.69** (0.02)	-0.74** (0.02)
Connection 1	-0.53** (0.02)	-0.47** (0.07)	-0.53** (0.02)	-0.53** (0.02)	-0.45** (0.02)	-0.63** (0.02)	-0.55** (0.02)
Constant 1	-5.79** (0.19)	-4.66 (4.05)	-5.84** (0.19)	-5.47** (0.18)	-6.05** (0.19)	-5.77** (0.17)	-6.33** (0.15)
Fare 2	-0.07** (0.00)	-0.09 (0.07)	-0.07** (0.00)	-0.06** (0.00)	-0.07** (0.00)	-0.07** (0.00)	-0.07** (0.00)
Connection 2	-0.31** (0.02)	-0.37** (0.04)	-0.31** (0.02)	-0.31** (0.02)	-0.28** (0.02)	-0.40** (0.01)	-0.36** (0.01)
Constant 2	-8.56** (0.40)	-8.35** (3.60)	-8.59** (0.40)	-8.48** (0.40)	-8.64** (0.40)	-8.07** (0.36)	-8.64** (0.27)
No. destination	0.38** (0.03)	0.32** (0.03)	0.34** (0.03)	0.34** (0.02)	0.36** (0.02)	0.40** (0.03)	0.48** (0.02)
No. departures	0.04** (0.00)	0.05** (0.01)	0.04** (0.00)	0.06** (0.00)	0.03** (0.00)	0.05** (0.00)	0.05** (0.00)
Distance	0.30** (0.04)	0.35** (0.06)	0.27** (0.04)	0.33** (0.04)	0.35** (0.04)	0.26** (0.04)	0.29** (0.04)
Distance <sup>2</sup>	-0.05** (0.01)	-0.05** (0.01)	-0.05** (0.01)	-0.05** (0.01)	-0.05** (0.01)	-0.05** (0.01)	-0.05** (0.01)
Tour	0.30** (0.03)	0.32** (0.04)	0.30** (0.03)	0.34** (0.04)	0.27** (0.03)	0.31** (0.03)	0.29** (0.03)
Slot-control	-0.19** (0.01)	-0.18** (0.01)	-0.21** (0.01)	-0.11** (0.01)	-0.19** (0.01)	-0.20** (0.01)	-0.13** (0.01)
Delay			0.76** (0.14)				
$\lambda$	0.77** (0.01)	0.72** (0.01)	0.77** (0.01)	0.69** (0.01)	0.76** (0.01)	0.79** (0.01)	0.83** (0.01)
$\gamma$	0.69** (0.12)	0.52 (1.77)	0.70** (0.12)	0.72** (0.12)	0.70** (0.12)	0.70** (0.11)	0.68** (0.09)
Cost variables							
Constant_short	1.07** (0.06)		1.07** (0.06)	1.29** (0.06)	0.85** (0.06)	0.88** (0.07)	0.81** (0.07)
Distance_short	0.26** (0.01)		0.26** (0.01)	0.28** (0.01)	0.26** (0.01)	0.26** (0.01)	0.23** (0.01)
Connection_short	-0.06** (0.03)		-0.06** (0.03)	-0.02 (0.03)	0.01 (0.02)	-0.08** (0.03)	-0.08** (0.03)
Constant_long	1.61** (0.08)		1.61** (0.08)	1.92** (0.07)	1.38** (0.08)	1.38** (0.09)	1.30** (0.09)
Distance_long	0.09** (0.01)		0.09** (0.01)	0.10** (0.01)	0.09** (0.01)	0.10** (0.01)	0.07** (0.01)
Connection_long	-0.09** (0.03)		-0.10** (0.03)	-0.07** (0.03)	-0.02 (0.03)	-0.10** (0.04)	-0.10** (0.03)
HubMC	-0.02 (0.01)		-0.02 (0.01)	-0.08** (0.02)	-0.03** (0.01)	0.00 (0.01)	0.00 (0.01)
SlotMC	0.08** (0.01)		0.08** (0.01)	0.11** (0.01)	0.08** (0.01)	0.09** (0.01)	0.06** (0.01)
Function value	49.37	46.54	49.44	40.22	51.15	42.90	44.89
Observations	214.8k	214.8k	214.8k	214.8k	238.5k	147.3k	214.8k

Notes: See Table 1A for the variable definitions. Column 1 is the base case. Column 2 does not use the markup condition. Column 3 adds delays to demand. Column 4 groups nearby airports. Columns 5 and 6 use a finer and a rougher set of fare bins, respectively. Column 7 includes 25 airport dummies. Standard errors are in parentheses. Carrier dummy estimates are available upon request.

\*\* Significant at the 5 percent level.

TABLE 4—PARAMETER ESTIMATES FROM DIFFERENT SPECIFICATIONS—2006

	Base case	No. MC	Delay	Combine airports	Small bin	Large bin	Airport dummy
Demand variables							
Fare 1	-1.05** (0.03)	-1.49** (0.37)	-1.06** (0.03)	-1.13** (0.03)	-1.09** (0.03)	-0.96** (0.03)	-1.04** (0.03)
Connection 1	-0.59** (0.03)	-0.33** (0.11)	-0.56** (0.03)	-0.46** (0.02)	-0.48** (0.02)	-0.72** (0.03)	-0.62** (0.03)
Constant 1	-5.68** (0.19)	-4.50 (3.89)	-5.61** (0.19)	-4.85** (0.24)	-5.87** (0.20)	-5.44** (0.22)	-6.06** (0.17)
Fare 2	-0.10** (0.00)	0.00 (0.06)	-0.10** (0.00)	-0.10** (0.00)	-0.10** (0.00)	-0.09** (0.00)	-0.10** (0.00)
Connection 2	-0.51** (0.02)	-0.60** (0.06)	-0.53** (0.02)	-0.67** (0.03)	-0.50** (0.02)	-0.52** (0.02)	-0.55** (0.02)
Constant 2	-8.60** (0.30)	-9.16** (3.23)	-8.55** (0.30)	-8.39** (0.34)	-8.64** (0.31)	-8.40** (0.37)	-8.85** (0.24)
No. destination	0.27** (0.02)	0.20** (0.03)	0.29** (0.02)	0.26** (0.02)	0.27** (0.02)	0.25** (0.02)	0.43** (0.02)
No. departures	0.11** (0.00)	0.10** (0.01)	0.11** (0.00)	0.13** (0.00)	0.09** (0.00)	0.12** (0.00)	0.12** (0.00)
Distance	0.53** (0.04)	0.55** (0.05)	0.53** (0.04)	0.41** (0.04)	0.52** (0.04)	0.55** (0.04)	0.58** (0.04)
Distance <sup>2</sup>	-0.08** (0.01)	-0.08** (0.01)	-0.08** (0.01)	-0.05** (0.01)	-0.08** (0.01)	-0.09** (0.01)	-0.09** (0.01)
Tour	0.36** (0.03)	0.37** (0.04)	0.35** (0.03)	0.41** (0.04)	0.34** (0.03)	0.37** (0.03)	0.34** (0.03)
Slot-control	-0.18** (0.01)	-0.18** (0.01)	-0.17** (0.01)	-0.10** (0.01)	-0.18** (0.01)	-0.19** (0.01)	-0.13** (0.01)
Delay			-0.82** (0.11)				
$\lambda$	0.72** (0.01)	0.67** (0.01)	0.72** (0.01)	0.63** (0.01)	0.72** (0.01)	0.72** (0.01)	0.77** (0.01)
$\gamma$	0.63** (0.11)	0.49 (1.69)	0.63** (0.11)	0.60** (0.14)	0.63** (0.12)	0.65** (0.13)	0.61** (0.09)
Cost variables							
Constant_short	1.16** (0.06)		1.16** (0.06)	1.30** (0.05)	1.02** (0.06)	1.22** (0.06)	1.07** (0.06)
Distance_short	0.19** (0.01)		0.19** (0.01)	0.22** (0.01)	0.19** (0.01)	0.21** (0.01)	0.17** (0.01)
Connection_short	0.07* (0.04)		0.10** (0.04)	0.25** (0.03)	0.14** (0.03)	-0.03 (0.04)	0.05 (0.04)
Constant_long	1.59** (0.07)		1.58** (0.07)	1.73** (0.07)	1.40** (0.07)	1.73** (0.07)	1.44** (0.07)
Distance_long	0.04** (0.01)		0.04** (0.01)	0.06** (0.01)	0.04** (0.01)	0.06** (0.01)	0.04** (0.01)
Connection_long	0.06 (0.04)		0.10** (0.04)	0.27** (0.04)	0.16** (0.04)	-0.07* (0.04)	0.05 (0.04)
HubMC	-0.05** (0.01)		-0.05** (0.01)	-0.07** (0.01)	-0.06** (0.01)	-0.06** (0.01)	-0.05** (0.01)
SlotMC	0.03** (0.01)		0.03** (0.01)	0.06** (0.01)	0.03** (0.01)	0.03** (0.01)	0.02** (0.01)
Function value	58.07	49.59	57.38	40.48	60.84	43.86	52.20
Observations	226.5k	226.5k	226.5k	226.5k	257k	146k	226.5k

Notes: See Table 1A for the variable definitions. Column 1 is the base case. Column 2 does not use the markup condition. Column 3 adds delays to demand. Column 4 groups nearby airports. Columns 5 and 6 use a finer and a rougher set of fare bins, respectively. Column 7 includes 25 airport dummies. Standard errors are in parentheses. Carrier dummy estimates are available upon request.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

*Other Specifications.*—In the base specification, we estimate demand parameters (especially price sensitivity) using both the share equation (5) and the pricing equation (10). As we are concerned about specification errors associated with our stylized pricing equation, we estimate the model again using only the share equation. The estimates are presented in the second column of Table 3 and Table 4. The  $\gamma$  parameter, or the percentage of tourists, is smaller than the base case, which leads to a noticeable change in the type-specific parameters (the parameters common to both types are very robust). For example, the tourists' price coefficient is much larger (in absolute value) in both years, partly to balance against a smaller fraction of price sensitive passengers. The average elasticities, however, are robust. The aggregate price elasticity in 1999 was 1.69, similar to the other specifications.<sup>46</sup> The pattern of a stronger preference for direct flights was very close to the base case. The connection semi-elasticity was 0.68 in 1999 and 0.76 in 2006. This is particularly reassuring, since these estimates do not suffer from potential specification errors associated with the pricing equation. These results provide evidence that our broad findings of changes in demand reflect genuine patterns in the data, and are not driven by our supply side assumptions, although the precision of the estimates for business travelers is improved by adding the supply side.

Flight delays can potentially explain the aversion toward connecting flights, since the possibility of missing a connection is directly affected by delays. However, there are some problems with the current delay statistics. First, in 1999, only the legacy carriers, Alaska and Southwest reported the on-time performance. The regional carriers that contracted with the legacy carriers and provided a significant amount of feeding traffic did not report to the DOT. Second, the delay statistics do not include passengers' waiting time for the extreme events like diverted flights or canceled flights. In column 3, we report the parameter estimates adding a delay variable, which is the percentage of flights arriving more than 30 minutes behind schedule. The connection's coefficients barely change. We have experimented with various other measures of delays, including the percentage of flights with delays longer than 15 minutes, with or without canceled or diverted flights. None of these experiments explains the increased disutility of connecting flights.

According to Stephane Bratu and Cynthia Barnhart (2005), when missed connections and flight cancelations are factored in, the average passenger delay is two-thirds longer than the official statistics. Given that the planes were fuller in 2006, the problem of delays was probably much worse, as it was harder to find seats on later flights in case of unexpected events. We expect that a better measure of the actual connecting time will help to explain the increased disutilities of connections.<sup>47</sup>

Six groups of airports are geographically close.<sup>48</sup> In column 4, we define markets based on the grouped airports. For example, all flights from either Midway or

<sup>46</sup> We cannot estimate the price elasticity in 2006 as the business travelers' price coefficient is pushed to the pre-imposed boundary of 0.

<sup>47</sup> Another possible explanation is the increasing marginal disutility of travel time. As travel time increases (due to long lines at the security check points, extra luggage handling, and longer waits for boarding and getting off the plane), consumers are increasingly less tolerant of connections that add additional travel time to the trip. We are currently pursuing data that allow us to look at this issue in more detail.

<sup>48</sup> See footnote 20 for a list of these airports.

O'Hare to Boston compete against each other. Products are nonetheless still defined by their origin-destination airport pair. Combining these airports affects 38 percent of the markets and doubles the number of products in some of the largest markets. Perhaps not surprisingly, the  $\lambda$  coefficient is somewhat smaller, since consumers face more similar choices in the grouped markets. The aggregate demand elasticity is  $-1.68$  in 1999, and  $-2.01$  in 2006, both of which are higher than the base case (which was  $-1.55$  and  $-1.67$ , respectively). With a more elastic demand, the marginal cost estimate is also higher than the base case.

Our products are generated by aggregating over a set of fare bins. To examine the robustness of the parameter estimates to changes in the bin size, column five and six in Table 3 and 4 report results with a finer set of bins and a rougher set of bins, respectively.<sup>49</sup> The aggregate price elasticity in 1999 in column 6 is  $-1.38$ , somewhat smaller than the base case. Most other elasticities are similar to the base case.

Some airports have higher traffic volume either because of historical reasons, or because of convenient geographic locations that are not captured by the model. In column 7, we add dummies to the 25 airports with the largest population. Similar to many studies with fixed effects, demand is less elastic, which leads to a lower estimate of the marginal cost in both years.

One might argue that the discovery of a stronger preference for direct flights is driven by changes in the supply side, rather than changes in demand. During our sample period, low-cost carriers expanded steadily, and offered a higher fraction of point-to-point service. The more negative connection coefficient in 2006 could be driven by the decreasing shares of the legacy carriers which happened to offer more connecting flights. To address this concern, we re-estimate the model using only markets that did not experience LCC entry between 1999 and 2006. If LCC chose to enter markets where people derive a higher value from direct services, then our connection disutility parameter would be biased toward zero in 2006. The results are presented in the first two columns of Table 5. The layover semi-elasticity is  $0.67$  in 1999, and  $0.74$  in 2006. Once again, we find evidence that consumers prefer direct flights even in markets that are not affected by LCC entry.<sup>50, 51</sup>

As mentioned at the beginning of this paper, the advent of new regional jets allows carriers to tailor the aircraft size to the market size and provide point-to-point service to markets traditionally dependent on connecting service. Another competing explanation for our finding is that consumers' preference has not changed, but there are more direct flights available. To tease out the regional-jet effect, we restrict the sample to markets longer than 1,500 miles one way, which exceeds the range of most regional jets. We lose about 70 percent of the markets, and our instruments have much less variation compared to the full sample. Distance squared is nearly

<sup>49</sup> In column 5, the set of bins was \$10 for fares under \$300, \$20 for fares between \$300 and \$700, \$50 for fares between \$700 and \$1,000, and \$100 for fares above \$1,000. In column 6, the bins were \$50 for fares under \$1,000, and \$100 for fares above \$1,000.

<sup>50</sup> The carrier dummies were not reported here, as there are too many parameters. Results are available upon request.

<sup>51</sup> The price elasticity for the specification that uses only markets without LCC entry is similar to all other specifications:  $-1.66$  in 1999 and  $-1.84$  in 2006.



TABLE 5—ROBUSTNESS CHECK

	Markets without LCC entry		Markets longer than 1,500 miles	
	1999	2006	1999	2006
<b>Demand variables</b>				
Fare 1	-0.85** (0.03)	-1.20** (0.04)	-0.88** (0.05)	-1.60** (0.10)
Connection 1	-0.48** (0.02)	-0.53** (0.04)	-0.53** (0.03)	-0.54** (0.05)
Constant 1	-5.68** (0.31)	-5.38** (0.39)	-3.44** (0.95)	-1.74** (0.75)
Fare 2	-0.07** (0.00)	-0.11** (0.00)	-0.06** (0.00)	-0.12** (0.00)
Connection 2	-0.36** (0.02)	-0.48** (0.03)	0.00 (0.11)	-0.55** (0.04)
Constant 2	-8.55** (0.71)	-8.75** (0.60)	-8.88* (5.00)	-7.36** (0.80)
No. destination	0.38** (0.03)	0.26** (0.03)	0.39** (0.06)	0.27** (0.05)
No. departures	0.05** (0.00)	0.12** (0.01)	0.02** (0.01)	0.07** (0.01)
Distance	0.31** (0.05)	0.51** (0.05)	-0.16** (0.02)	-0.14** (0.02)
Distance <sup>2</sup>	-0.05** (0.01)	-0.08** (0.01)		
Tour	0.18** (0.05)	0.18** (0.04)	0.51** (0.05)	0.63** (0.05)
Slot-control	-0.15** (0.01)	-0.18** (0.01)	-0.16** (0.01)	-0.17** (0.01)
$\lambda$	0.75** (0.01)	0.72** (0.01)	0.75** (0.02)	0.67** (0.01)
$\gamma$	0.71** (0.21)	0.63** (0.23)	0.85 (0.76)	0.57 (0.35)
<b>Cost variables</b>				
Constant_short	1.09** (0.06)	1.24** (0.06)		
Distance_short	0.29** (0.01)	0.23** (0.01)		
Connection_short	0.04 (0.03)	0.09* (0.05)		
Constant_long	1.66** (0.09)	1.68** (0.09)	3.13** (0.13)	2.22** (0.10)
Distance_long	0.10** (0.01)	0.05** (0.01)	0.04** (0.01)	0.06** (0.01)
Connection_long	0.04 (0.04)	0.09 (0.06)	-0.12** (0.06)	0.10* (0.06)
HubMC	-0.08** (0.02)	-0.07** (0.01)	-0.01 (0.04)	-0.13** (0.02)
SlotMC	0.10** (0.01)	0.05** (0.01)	0.10** (0.02)	0.04** (0.01)

Notes: Columns 1 and 2 only use markets that did not experience LCC entry between 1999 and 2006. Columns 3 and 4 use markets longer than 1,500 miles that are less likely to be affected by the regional jets.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

collinear with the distance variable and is omitted from the regressors.<sup>52</sup> Demand is much more elastic than the base case, but the pattern of a stronger preference for direct flights remains: the connection semi-elasticity was 0.63 in 1999 and 0.80 in 2006.

We have estimated many other specifications that are not reported here. For example, we have estimated a model restricting the cost parameters to be the same across all markets for all specifications.<sup>53</sup> We have also experimented with type-specific tour and flight frequency parameters. Finally, we estimated a pure logit demand model using only the within market variation.<sup>54</sup> Our major findings (more price sensitive demand, a much stronger preference for direct flights, and changes in marginal cost favoring direct flights) are robust and appear in almost every set of parameter estimates. We are convinced that these findings reveal inherent data patterns and are not due to our modeling assumptions. The intuition for these results is straightforward. A negative supply shock should induce a smaller quantity *and* a higher price. In our data, fewer passengers flew connecting flights even though fares were uniformly lower in 2006—at each quantile of the fare distribution and in markets with or without entry of LCCs.

*Marginal Effects.*—To better understand the magnitude of the parameters, Table 6 tabulates changes in demand with varying product attributes. The effect of carrier airport presence on demand appears to be mild. Doubling the number of destinations for all products raises the aggregate demand by 11 percent in 1999 and 9 percent in 2006. On the other hand, adding one daily departure to all flights drives up aggregate demand by 6 percent in 1999 and 16 percent in 2006. Changes in distance barely affect demand. In contrast, both the tour dummy and the slot variable have a significant impact. Adding the tour dummy to all products boosts the number of passengers by 32 percent in 1999 and 39 percent in 2006, while congestion in slot controlled airports reduces demand by 22 percent. These marginal effects do not vary much across specifications.

*Elasticities, Marginal Cost, and Markups.*—In Tables 7 and 8, we summarize the elasticities, the percentage of each type of passenger, marginal cost, and markups for different specifications. The aggregate price elasticity ranged from  $-1.37$  to  $-1.69$  in 1999 and  $-1.58$  to  $-2.01$  in 2006. The increase in price elasticity over the sample period varied from 6.5 percent to 20 percent, with an average of 13 percent. The connection semi-elasticity is relatively stable across different specifications, with an average increase of 16 percent. The scale economies of connecting flights appear

<sup>52</sup> There is only one set of cost parameters, since all selected markets are longer than 1,500 miles.

<sup>53</sup> Results are available upon request. The cost parameters were more robust when we restricted them to be the same across long-haul and short-medium haul markets, but we prefer our reported specifications, as there were significant cost differences between these markets (for example, the type of aircraft used were different).

<sup>54</sup> We estimated a pure logit model instead of a two-type nested logit model because within market variation alone does not allow us to identify the type-specific parameters or the nested logit parameter  $\lambda$ . We find that the connection semi-elasticities (both the levels and the changes) are similar to the full model (the connection semi-elasticity increases from 0.62 in 1999 to 0.75 in 2006). The 1999 demand elasticity is similar to the full model ( $-1.77$ ), but the demand elasticity in 2006 cannot be reliably estimated, mainly because our instruments and regressors do not have enough within market variation. Results are available upon request.

TABLE 6—PERCENTAGE CHANGES IN DEMAND WHEN PRODUCT ATTRIBUTES CHANGE

	1999	2006
No. destination doubles	11%	9%
Add one daily departure	6%	16%
Distance up 10 percent	−1%	−1%
Tour dummy changes from 0 to 1	32%	39%
Slot changes from 0 to 1	−22%	−22%

*Note:* The table reports the percentage change in market demand when the relevant product attribute is changed as specified. For example, in 2006, adding one departure to all products increases the market demand by 16 percent on average.

TABLE 7A—ELASTICITY ESTIMATES FROM DIFFERENT SPECIFICATIONS—1999

	Base case	No. MC	Delay	Combine airports	Small bin	Large bin	Airport dummy
Price elasticity							
Type 1	−5.01	−7.81	−5.01	−5.64	−4.90	−4.77	−4.40
Type 2	−0.44	−0.65	−0.44	−0.46	−0.42	−0.48	−0.43
Both types	−1.96	−2.16	−1.96	−2.35	−1.95	−1.63	−1.62
Aggregate price elasticity	−1.55	−1.69	−1.55	−1.68	−1.53	−1.38	−1.37
Connection semi-elasticity							
Type 1	0.75	0.73	0.75	0.78	0.69	0.79	0.74
Type 2	0.55	0.64	0.55	0.59	0.51	0.63	0.58
All	0.66	0.68	0.66	0.71	0.61	0.71	0.66
Percentage of passengers							
Type 1	0.59	0.47	0.59	0.64	0.57	0.58	0.54
Type 2	0.41	0.53	0.41	0.36	0.43	0.42	0.46

TABLE 7B—ELASTICITY ESTIMATES FROM DIFFERENT SPECIFICATIONS—2006

	Base case	No. MC	Delay	Combine airports	Small bin	Large bin	Airport dummy
Price elasticity							
Type 1	−6.55		−6.57	−8.09	−6.41	−6.66	−6.10
Type 2	−0.63		−0.63	−0.70	−0.61	−0.63	−0.60
Both types	−2.10		−2.15	−2.94	−2.15	−1.97	−1.89
Aggregate price elasticity	−1.67		−1.70	−2.01	−1.63	−1.66	−1.58
Connection semi-elasticity							
Type 1	0.80	0.63	0.79	0.77	0.74	0.86	0.80
Type 2	0.75	0.83	0.76	0.88	0.75	0.76	0.76
All	0.77	0.76	0.77	0.83	0.74	0.80	0.77
Percentage of passengers							
Type 1	0.51	0.47	0.52	0.59	0.48	0.55	0.48
Type 2	0.49	0.53	0.48	0.41	0.52	0.45	0.52

*Notes:* The aggregate price elasticity measures the percentage change in total demand when all products' prices increase by 1 percent. Connection semi-elasticity measures the percentage change in product  $j$ 's demand when it switches from a direct flight to a connecting flight, fixing other products' attributes.

to have disappeared over the sample period, and the increase in connecting flights' marginal cost is much bigger than that of direct flights. Rising costs, combined with lower fares, leads to a sizable reduction of the markups of connecting flights.

The Lerner Index, or the ratio of markups to fares, for the top 10 percent most expensive products dropped substantially over the sample period, from 90 percent in 1999 to less than 70 percent in 2006. The reduction in the profitability of these high-end products, together with the shrinking profit of connecting flights, is an important explanation of the legacy carriers' financial stress in recent years.

TABLE 8A—MARGINAL COST AND LERNER INDEX FROM DIFFERENT SPECIFICATIONS—1999

Marginal cost (\$)	Base case	Delay	Combine airports	Small bin	Large bin	Airport dummy
Connecting flights	160	160	190	153	141	125
Direct flights	149	149	170	126	132	120
All products	156	156	183	142	138	123
Lerner Index						
Connecting flights	0.60	0.60	0.53	0.60	0.69	0.69
Direct flights	0.66	0.66	0.61	0.68	0.78	0.74
All products	0.63	0.63	0.56	0.63	0.72	0.71

TABLE 8B—MARGINAL COST AND LERNER INDEX FROM DIFFERENT SPECIFICATIONS—2006

Marginal cost (\$)	Base case	Delay	Combine airports	Small bin	Large bin	Airport dummy
Connecting flights	167	173	229	165	157	149
Direct flights	138	137	158	120	147	124
All products	155	158	199	145	153	139
Lerner Index						
Connecting flights	0.56	0.54	0.41	0.54	0.60	0.60
Direct flights	0.66	0.66	0.60	0.66	0.69	0.69
All products	0.60	0.59	0.49	0.60	0.64	0.64

Note: See Table 3 for explanations of the specification in each column.

TABLE 9—CARRIER PROFIT AND REVENUE PER MARKET

Year		Profit (\$100k)			Revenue (\$100k)		
		All fares	Bottom 90% fares	Top 10% fares	All fares	Bottom 90% fares	Top 10% fares
1999	All flights	17.80	11.77	6.03	26.38	19.79	6.60
	Direct	14.95	10.17	4.77	21.90	16.62	5.29
	Connecting	2.86	2.14	0.72	4.48	3.64	0.84
2006	All flights	14.46	12.19	2.27	23.92	20.72	3.19
	Direct	12.53	11.03	1.50	20.53	18.31	2.23
	Connecting	1.94	1.62	0.32	3.38	2.93	0.45

### B. Profit and Revenue Estimates

The average number of products offered by a carrier in a given market differed slightly between 1999 and 2006.<sup>55</sup> To avoid the complication of the changing number of products (which might reflect the changing dispersion of prices rather than the changing number of distinct products), we analyze a carrier's average profit and revenue per market, instead of the average profit per product. We also focus on the legacy carriers throughout this analysis. We first report the profit estimates and the counterfactual results using the base case parameters, and then describe the general patterns across all specifications.

<sup>55</sup> There are, on average, 54 products and 3.5 carriers per market in 1999 and 53 products and 3.3 carriers per market in 2006.

Table 9 displays the legacy carriers' profit and revenue separately for connecting flights and direct flights. For connecting flights, 2006 witnessed fewer passengers, lower revenues, higher costs, and lower profits. Compared with 1999, the average demand shrank by 14 percent, and the average fare was 12 percent lower. As a result, the average revenue fell by 25 percent, and profit fell even further, by 32 percent. Profit for the top 10 percent most expensive products decreased by 56 percent, which was driven by a bigger reduction in fares and demand among these high-end products.

The picture for direct flights is more complicated. The average number of direct passengers per carrier per market increased by 8 percent, but average revenue was down by 6 percent, and average profit was 16 percent lower. The changes across different quantiles of the fare distribution reveal that all of the profit reduction occurred among the 10 percent most expensive products. In 1999, these products generated an average profit of \$477k per carrier per market, and accounted for 32 percent of total profits from all direct flights. By 2006, profits from the top 10 percent products declined to merely \$150k, and constituted only 12 percent of total profits. As our parameter estimates suggest, demand in 2006 was more price sensitive. Even though consumers displayed a stronger preference for direct flights, they generally stayed away from high-end products and switched to flights with low or medium fares. Profits and revenues from the bottom 90 percent of flights were about 8–10 percent higher in 2006 than in 1999. However, the higher profitability from the low- and medium-fare flights was overwhelmed by the profit declines among the most expensive flights. Profits from all direct flights fell by 16 percent.

When we combine both direct and connecting flights, the legacy carriers transported 4 percent more passengers, but generated 9 percent fewer revenues and 19 percent fewer profits in 2006 than in 1999.

### C. Counterfactual Analysis

To examine how legacy carriers' profits were affected by changes in demand, changes in marginal cost, and LCC's expansion, we calculate counterfactual profits and revenues for the following five scenarios:

- using 2006 observed product attributes and marginal cost parameters, but 1999 demand parameters;
- using 2006 observed product attributes and marginal cost parameters, but 1999 demand parameters and  $\xi_j$  that "replicates" its distribution in 1999;
- using 2006 observed product attributes and demand parameters, but 1999 marginal cost parameters;
- using 2006 observed product attributes, demand and marginal cost parameters, but excluding LCCs from the markets they entered between 1999 and 2006;
- using 2006 observed product attributes, but 1999 demand and marginal cost parameters,  $\xi_j$  that replicates its distribution in 1999, and excluding LCCs from the markets they entered during the sample period.

TABLE 10—CARRIER PROFIT AND REVENUE PER MARKET  
FOR DIFFERENT COUNTERFACTUAL SCENARIOS: CONNECTING FLIGHTS

Different scenarios	Profit (\$100k)			Revenue (\$100k)		
	All fares	Bottom 90% fares	Top 10% fares	All fares	Bottom 90% fares	Top 10% fares
1999 base case	2.86	2.14	0.72	4.48	3.64	0.84
2006 base case	1.94	1.62	0.32	3.38	2.93	0.45
1999 demand $\theta_d$	2.47	1.97	0.50	4.05	3.43	0.63
1999 demand $\theta_d$ and $\xi$	2.45	1.91	0.54	3.95	3.27	0.68
1999 MC parameters	2.02	1.64	0.38	3.51	2.99	0.52
No LCC expansion	2.01	1.62	0.39	3.51	2.97	0.54
All factors	2.59	2.04	0.56	4.15	3.45	0.69

Notes: We use 2006 product attributes for all counterfactual exercises. In each row, we solve for a new vector of the optimal prices that satisfy the first order conditions incorporating the parameter changes as specified.

In each exercise, we solve for a new vector of optimal prices that satisfy the first order conditions incorporating changes specified above.<sup>56</sup> The first exercise quantifies the effect of changes in demand, including the increased price sensitivity and the higher aversion to connecting flights.

As discussed in Section IIA,  $\xi_j$ , the utility from the unobserved product attributes (like refundability, advance purchase requirements, etc.), plays an important role in determining demand. The difference in  $\xi_j$  between 1999 and 2006 is a combination of changes in taste and changes in unobserved product characteristics. If these product attributes are similar across the two years, then the difference in  $\xi_j$  reflects changes in consumers' utility from these attributes, and constitutes an important component of the demand change. In the second exercise, we incorporate changes in  $\xi_j$  by replicating its 1999 distribution conditional on fares separately for direct and connecting flights. For example, given all direct flights priced at \$350, we replace the first quantile of  $\xi_j^{06}$  with the first quantile of  $\xi_j^{99}$ , etc. Then we solve for the counterfactual prices using 1999 demand parameters and the constructed vector of  $\xi_j$ .<sup>57</sup>

The third exercise analyzes the effect of changes in marginal cost on legacy carriers' profits, the fourth one examines competition from LCCs, while the last exercise combines all the factors discussed above.

Table 10 summarizes the counterfactual results for connecting flights. Overall, the model does a decent job in explaining the profit change for connecting flights. Replacing the 2006 demand parameters with the 1999 values explains 58 percent and 61 percent of the profit and revenue reduction, respectively. Results are similar when we incorporate  $\xi_j$ 's 1999 distribution.

Using the 2006 demand parameters and the 1999 cost parameters accounts for about 9 percent of the profit and revenue decrease between 1999 and 2006. The

<sup>56</sup> In solving for optimal prices, we restricted the first order condition to be smaller than  $10^{-9}$ . The convergence was slow for the third and fourth counterfactual exercise, so we set the tolerance level to  $10^{-8}$ . There was not much difference in profit estimates whether using tolerance level  $10^{-8}$  or  $10^{-9}$ .

<sup>57</sup> We replicate  $\xi_j$ 's distribution conditioning on fares because the unobserved attributes are likely to differ between cheap tickets and expensive ones.

TABLE 11—CARRIER PROFIT AND REVENUE PER MARKET FOR DIFFERENT COUNTERFACTUAL SCENARIOS: DIRECT FLIGHTS

Different scenarios	Profit (\$100k)			Revenue (\$100k)		
	All fares	Bottom 90% fares	Top 10% fares	All fares	Bottom 90% fares	Top 10% fares
1999 base case	14.95	10.17	4.77	21.90	16.62	5.29
2006 base case	12.53	11.03	1.50	20.53	18.31	2.23
1999 demand parameters	10.97	9.62	1.35	18.08	16.28	1.80
1999 demand parameters and $\xi$	15.06	11.72	3.34	22.11	17.67	4.44
1999 MC parameters	11.99	10.41	1.58	19.85	17.48	2.36
No LCC expansion	12.81	11.20	1.61	20.85	18.49	2.36
All factors	14.80	11.46	3.34	22.03	17.55	4.48

Notes: We use 2006 product attributes for all counterfactual exercises. In each row, we solve for a new vector of the optimal prices that satisfy the first order conditions incorporating the parameter changes as specified.

marginal cost is higher in 2006, which leads to higher fares, lower demand, and lower profits.

Around 40 percent of the markets experienced LCC entry during the sample period.<sup>58</sup> Compared with changes in demand, competition from LCCs has a modest impact on connecting flights' profit. Removing LCCs explains 15 percent of a legacy carrier's profit drop in markets affected by entry, or 8 percent when averaged over all markets. There are a couple of explanations for this modest impact. First, many new products introduced by the low-cost carriers are direct flights. As discussed below, LCC entry accounts for a much larger fraction of the direct flights' profit reduction. Second, the legacy carriers have gradually developed strategies (for example, lowering fares, adding departures) to compete with low-cost carriers.

When we incorporate all factors, the model is able to replicate 72 percent of the profit reduction during the sample period. The model performs well even when we look at high-fare and low-fare products separately. It explains 81 percent of the profit change for the bottom 90 percent of products, and 60 percent of the profit change for the top 10 percent most expensive products.

Results for direct flights are presented in Table 11. Using the 1999 demand parameters and  $\xi_j$ 's 1999 distribution, the predicted profit from all direct flights is close to the observed level in 1999.<sup>59</sup> As the marginal cost of direct flights was higher in 1999, using 1999 cost parameters reduces profits by 4 percent. Removing LCCs explains 25 percent of the legacy carriers' profit reduction in markets that experienced LCC entry, and 12 percent when averaged over all markets. Combining all factors, we are able to replicate 94 percent of the observed changes in direct flights' profits.

Although the model replicates the average profit of direct flights, it does not fit as well the profit increase for low-and-medium-fare products and the profit decrease for high-fare products. For example, using the 1999 demand parameters, the predicted profit is comparable to 1999's observed profit for the bottom 90 percent of direct

<sup>58</sup> Some of these markets already had low-cost carriers in 1999 (like AirTran, Frontier, or Southwest). A market experienced LCC entry if a new low-cost carrier established a service in that market between 1999 and 2006.

<sup>59</sup>  $\xi_j$  was an important factor in determining demand for high-end products. Its dispersion among high-end direct flights was much wider in 1999 than it was in 2006. Replicating  $\xi_j$ 's distribution in 1999 helped us to duplicate demand for high-fare flights.



TABLE 12A—PERCENTAGE OF PROFIT CHANGES EXPLAINED BY DIFFERENT COUNTERFACTUAL SCENARIOS—CONNECTING FLIGHTS

Different scenario	Base case	Delay	Combine airports	Small bin	Large bin	Airport dummy
1999 demand parameters and $\xi$	0.56	0.46	0.49	0.56	0.53	0.54
1999 MC parameters	0.09	0.14	0.33	0.14	0.11	0.20
No LCC expansion	0.08	0.08	0.08	0.08	0.08	0.06
All factors	0.72	0.66	0.87	0.76	0.70	0.77

TABLE 12B—PERCENTAGE OF PROFIT CHANGES EXPLAINED BY DIFFERENT COUNTERFACTUAL SCENARIOS—DIRECT FLIGHTS

Different scenario	Base case	Delay	Combine airports	Small bin	Large bin	Airport dummy
1999 demand parameters and $\xi$	1.05	1.02	0.85	1.29	0.70	0.87
1999 MC parameters	-0.22	-0.23	-0.20	-0.16	0.18	0.07
No LCC expansion	0.12	0.11	0.18	0.14	0.08	0.08
All factors	0.94	0.90	0.77	1.26	0.90	0.98

*Notes:* We use 2006 product attributes for all counterfactual exercises. In each row, we solve for a new vector of the optimal prices that satisfy the first order conditions incorporating the parameter changes as specified.

flights, but is only 28 percent of the observed profit for the top 10 percent of direct flights. The model fits changes in the mean reasonably well, but does not do a good job of fitting different quantiles of the data distribution.

We have repeated the above counterfactual exercises for all other specifications and summarized the results in Table 12A and B. For connecting flights, changes in demand account for around 46–56 percent of the profit reduction, changes in cost account for 9–33 percent of the profit reduction, and entry of LCCs account for 6–8 percent of the profit reduction. For direct flights, demand is by far the most important factor. LCC expansion contributes to 8–18 percent of the profit reduction. The changes in marginal cost have mixed effects. In four specifications (including our base case), the marginal costs of direct flights were lower in 2006, while in the other two specifications, the marginal costs were higher in 2006. When we combine all factors, the model replicates 66–87 percent of connecting flights' profit decline, and 77–126 percent for direct flights.<sup>60</sup>

## VI. Conclusions

We find that compared to the late 1990s, air-travel demand was more price sensitive in 2006. Passengers displayed a stronger preference for direct flights. In addition,

<sup>60</sup> We have also conducted a counterfactual analysis using the 1999 products, but 2006 preference and/or cost parameters. The results are similar to those reported here. For example, using 1999 products, but 2006 preference and cost parameters, the average profit for direct and connecting flights is \$1.29 million and \$222,000, respectively. Results are available upon request.

We base our counterfactual analysis on the 2006 products mainly because it allows us to explicitly analyze the impact of LCC entry on legacy carriers' profits. All we need to do is to remove LCCs from the markets that they entered during the sample period. In addition, since we are interested in understanding what could explain the observed carriers' financial stress in 2006, using the 2006 product mix seems more natural than using the 1999 product mix.

the change in marginal cost favored direct flights. These three factors, together with the expansion of LCCs, explain more than 80 percent of the observed reduction in legacy carriers' profits. Despite the media's emphasis on increasing fuel costs and competition from LCCs, the change in demand is also a very important factor in understanding the legacy carriers' profit losses. The changes in demand and costs are also consistent with observed changes in network structure: less emphasis on hub airports and a greater number of direct flights.

We conclude with some caveats. First, as costs are not directly observed, we obtain estimates of costs from our admittedly stylized supply side. Therefore, the magnitude of the impact of changes in costs on profitability should be interpreted accordingly. In addition, we estimate changes in variable profits, not in net profits, because we do not observe fixed costs. Second, we find that the impact of LCC entry in the 2000s is modest compared to the change in demand. If the expansion of LCCs has contributed to the change in demand by affecting consumers' search behavior and their awareness of fare dispersion, then LCCs' general equilibrium effect could be much larger. Last, our model is static and does not include dynamic considerations, like the choice of capacity, network formation, or improvements in technological efficiency. Modeling these dynamic elements is an interesting question for future research.

## APPENDIX

### I. Constructing Departures and Flight Delays

In this section, we explain how we construct flight frequencies and the delay variable. The scheduling data from Back Aviation Solutions report the scheduled departure time and arrival time for all flights operated by US carriers that file with Official Airline Guides. Obtaining the number of departures for direct flights is straightforward. We count the total number of flights by all carriers that operate for a ticketing carrier in a given market. Constructing the number of departures for connecting flights is slightly more involved. We combine the schedules of all carriers that operate for a ticketing carrier in a given market, and restrict the connecting time to between 45 minutes and 4 hours. When there are multiple feasible connections, only the connection with the shortest layover time is included.<sup>61</sup>

DOT publishes the flight-level on-time arrival data for nonstop domestic flights.<sup>62</sup> We first obtain the on-time performance for each operating carrier for each airport pair, and then aggregate over operating carriers using the number of departures as weights to generate the delay measure for each ticketing carrier in each airport pair. For connecting flights, the delay variable is the average over the two flight segments.

<sup>61</sup> Suppose we have the following flight schedule among airports A, B, C: flight 1001 departs from A at 8 a.m., arrives at B at 2 p.m.; flight 1002 departs from A at 10 a.m. and arrives at B at 4 p.m.; flight 1003 departs from B at 5:30 p.m. and arrives at C at 7:30 p.m. Even though both flight 1001 and 1002 can be connected with flight 1003, we only count the connection with the shortest layover time. In this example, the carrier operates one connecting departure in market A–C.

<sup>62</sup> In 1999, only the major carriers—American, Continental, Delta, Northwest, Trans World, United, and US Air, plus Alaska, American West, and Southwest—reported the delay statistics to the DOT. In 2006, some of the largest regional carriers reported the delay statistics as well.

## II. A Model of Product Availability

As discussed in Section IIA and IVD, we use  $\xi_{jt}$  to proxy for a product's availability (as well as other unobserved product attributes). In this section, we consider the impact of ignoring product availability on our parameter estimates. We first analyze a logit model with missing products, and discuss conditions under which the bias would be insignificant. Then we extend our results to nested logit models. Finally, we present some Monte Carlo evidence, where we demonstrate that the magnitude of bias in our parameter estimates is likely to be small in our application.

*A Logit Model.*—We begin with a pure logit model. Missing products lead to varying choice sets. Let  $C$  denote a choice set in a market, and  $\mathcal{C}$  denote the space of all possible choice sets. If product  $j$  is offered in choice set  $C$ , then we write  $j \in C$ . The probability that a consumer faces  $C$  is  $\phi_C$ . The probability that product  $j$  is available is

$$\bar{\phi}^j = \sum_{C:j \in C} \phi_C.$$

Conditional on product  $j$  being available, the probability of observing choice set  $C$  is

$$\phi_{C/j} \equiv \frac{\phi_C}{\bar{\phi}^j}.$$

The expected market share of good  $j$ , averaged across the possible choice sets, is

$$\begin{aligned} s_j &= \sum_{C:j \in C} \phi_C \frac{e^{\delta_j}}{1 + \sum_{k \in C} e^{\delta_k}} \\ &= \bar{\phi}^j \sum_{C:j \in C} \phi_{C/j} \frac{e^{\delta_j}}{1 + \sum_{k \in C} e^{\delta_k}} \\ &= \sum_{C:j \in C} \phi_{C/j} \frac{e^{\delta_j + \ln(\bar{\phi}^j)}}{1 + \sum_{k \in C} e^{\delta_k}}, \end{aligned}$$

where in the last expression, the probability that product  $j$  is available is incorporated into the numerator. If we define the “mean utility adjusted for product availability” as

$$\tilde{\delta}_j = \delta_j + \ln(\bar{\phi}^j),$$

then the market share of product  $j$  is

$$s_j = \sum_{C:j \in C} \phi_{C/j} \frac{e^{\tilde{\delta}_j}}{1 + \sum_{k \in C} e^{\delta_k}} = e^{\tilde{\delta}_j} \sum_{C:j \in C} \phi_{C/j} \frac{1}{1 + \sum_{k \in C} e^{\delta_k}} = e^{\tilde{\delta}_j} \tilde{s}_0,$$

where  $\tilde{s}_0 = \sum_{C:j \in C} \phi_{C/j} \frac{1}{1 + \sum_{k \in C} e^{\delta_k}}$  is the average share of the outside good conditional on product  $j$  being available. Solving for the adjusted mean product utility, we have:

$$\ln(s_j) - \ln(\tilde{s}_0) = \tilde{\delta}_j.$$

Since product availability is unknown, we do not observe the conditional log share  $\ln(\tilde{s}_0)$ . Instead, we replace it with the observed log share  $\ln(s_0)$ , and regress  $\ln(s_j) - \ln(s_0)$  on product attributes and prices to estimate taste parameters:

$$\begin{aligned} \ln(s_j) - \ln(s_0) &= \tilde{\delta}_j + (\ln(\tilde{s}_0) - \ln(s_0)) \\ (A1) \qquad \qquad \qquad &= X_j \beta - \alpha P_j + (\xi_j + \ln(\bar{\phi}^j)) + (\ln(\tilde{s}_0) - \ln(s_0)), \end{aligned}$$

where  $(\ln(\tilde{s}_0) - \ln(s_0))$  can be interpreted as our model residual from ignoring product availability.

There are several cases where the model residual is likely to be small. If both the conditional and unconditional shares of the outside good are close to one, then the difference  $\ln(\tilde{s}_0) - \ln(s_0)$  will be small. In addition, if  $s_j$  is small, then  $\ln(s_j)$  is a large negative number, and the proportional error in calculating  $\tilde{\delta}_j$  will be small.<sup>63</sup>

Whether or not parameter estimates have a large bias depends on the covariance between  $\ln(\tilde{s}_0) - \ln(s_0)$  and instruments. If  $\ln(\tilde{s}_0) - \ln(s_0)$  is small, then the covariance is also necessarily small. Regardless of the magnitude of  $\ln(\tilde{s}_0) - \ln(s_0)$ , a small covariance between the model residual and instruments will result in little bias in the estimated parameters.<sup>64</sup>

*A Nested Logit Model.*—The arguments above extend easily to a nested logit model, where product  $j$ 's share is

$$\begin{aligned} s_j &= \sum_{C:j \in C} \phi_{C/j} \frac{e^{(\delta_j + \lambda \ln(\bar{\phi}^j))/\lambda}}{\sum_{k \in C} e^{\delta_k/\lambda}} \frac{\left[ \sum_{k \in C} e^{\delta_k/\lambda} \right]^\lambda}{1 + \left[ \sum_{k \in C} e^{\delta_k/\lambda} \right]^\lambda} \\ &= e^{\tilde{\delta}_j} \sum_{C:j \in C} \phi_{C/j} \left[ \frac{e^{\tilde{\delta}_j/\lambda}}{\sum_{k \in C} e^{\delta_k/\lambda}} \right]^{1-\lambda} \frac{1}{1 + \left[ \sum_{k \in C} e^{\delta_k/\lambda} \right]^\lambda}. \end{aligned}$$

<sup>63</sup> When the conditional and unconditional shares of the outside good are far less than 1, the products might still vary in a way that keeps these two shares close, but obviously different outcomes are possible depending on the data generating process.

<sup>64</sup> When some products are missing, the constant in the inverted share equation (A1) is  $\beta_0 - \ln(\bar{\phi}_j)$ , and our constant parameter estimate  $\hat{\beta}_0$  will reflect this.

Inverting the shares and re-arranging terms, we have:

$$\begin{aligned} \ln s_j - \ln s_0 &= \tilde{\delta}_j + (1 - \lambda) \ln s_{j|g} \\ &+ \ln \left\{ \sum_{C:j \in C} \phi_{C/j} \left[ \frac{e^{\tilde{\delta}_j/\lambda}}{\sum_{k \in C} e^{\delta_k/\lambda}} \right]^{1-\lambda} \frac{1}{1 + \left[ \sum_{k \in C} e^{\delta_k/\lambda} \right]^\lambda} \right\} \\ &- (1 - \lambda) \ln s_{j|g} - \ln \tilde{s}_0 + (\ln \tilde{s}_0 - \ln s_0), \end{aligned}$$

where  $s_{j|g}$  is product  $j$ 's within group share  $s_{j|g} = \sum_{C:j \in C} \phi_{C/j} \left[ \frac{e^{\tilde{\delta}_j/\lambda}}{\sum_{k \in C} e^{\delta_k/\lambda}} \right]^{1-\lambda}$ , and  $\tilde{s}_0$  is the share of the outside good conditional on product  $j$  being available  $\tilde{s}_0 = \sum_{C:j \in C} \phi_{C/j} \left( \frac{1}{1 + \left[ \sum_{k \in C} e^{\delta_k/\lambda} \right]^\lambda} \right)$ . The last two rows are once again our model residuals. As in the logit example, both a large share of the outside good (or a small share of the missing product) and a moderate covariance between instruments and the model residual help to alleviate the potential parameter bias.

*Monte-Carlo Evidence.*—We have simulated a large variety of nested logit models. The parameters are chosen such that the share of the outside good is above 98 percent,<sup>65</sup> and the aggregate price elasticity is between  $-1$  and  $-2$ . We experimented with different products (from 5 to 50), various missing products (from 1 to 20), different attributes that make the missing products more or less attractive than products that are always available, as well as different values of  $\lambda$ .<sup>66</sup> Our Monte-Carlo experiments produce moderate bias in the price coefficient, ranging from 0.1 percent to 5 percent in most cases.

Our full model presented in the main body of this paper is a discrete mixture of nested logit models and is somewhat more complicated. However, the same argument applies. The degree to which the misspecified share equation affects our parameter estimates depends on the magnitude of the model residual, as well as its covariance with instruments. Given that in our application, the outside good's market share exceeds 99 percent, each market has on average 53 products, with the share of any given product being fairly small, and most of our instruments are at the route level and unlikely to be highly correlated with product-specific approximation errors, we believe that the bias in our parameter estimates should be small.

## REFERENCES

- Armantier, Olivier, and Oliver Richard. 2008. "Domestic Airline Alliances and Consumer Welfare." *RAND Journal of Economics*, 39(3): 875–904.
- Berry, Steven T. 1990. "Airport Presence as Product Differentiation." *American Economic Review*, 80(2): 394–99.

<sup>65</sup> The average share of the outside good is 99 percent in our data.

<sup>66</sup>  $\lambda$  affects the magnitude of the model residuals. In our experiments, with small product shares (and hence large negative  $\delta_j$ ), the model residual is smaller when  $\lambda$  is closer to 0.

- Berry, Steven, Michael Carnall, and Pablo T. Spiller.** 2006. "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity." In *Advances in Airline Economics: Competition Policy and Antitrust*, Vol. 1, ed. Darin Lee, 183–214. Amsterdam: Elsevier.
- Berry, Steven, and Philip A. Haile.** 2009. "Identification of Discrete Choice Demand from Market Level Data." <http://www.econ.upenn.edu/system/files/Empirical030520090.pdf>.
- Berry, Steven, James Levinsohn, and Ariel Pakes.** 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–90.
- Borenstein, Severin.** 2005. "U.S. Domestic Airline Pricing, 1995–2004." University of California, Berkeley Competition Policy Center Working Paper CPC05-48.
- Borenstein, Severin, and Nancy L. Rose.** 1994. "Competition and Price Dispersion in the U.S. Airline Industry." *Journal of Political Economy*, 102(4): 653–83.
- Borenstein, Severin, and Nancy L. Rose.** 2007. "How Airline Markets Work...Or Do They? Regulatory Reform in the Airline Industry." National Bureau of Economic Research Working Paper 13452.
- Bratu, Stephane, and Cynthia Barnhart.** 2005. "An Analysis of Passenger Delays Using Flight Operations and Passenger Booking Data." *Air Traffic Control Quarterly*, 13(1): 1–27.
- Brueckner, Jan K., and Pablo T. Spiller.** 1994. "Economies of Traffic Density in the Deregulated Airline Industry." *Journal of Law and Economics*, 37(2): 379–415.
- Brunger, William.** 2007. "The Impact of Internet on Airline Pricing." Unpublished.
- Ciliberto, Federico, and Elie Tamer.** 2009. "Market Structure and Multiple Equilibria in Airline Markets." *Econometrica*, 77(6): 1791–1828.
- Ciliberto, Federico.** 2008. "Strategic Deterrence in a Dynamic Game: Evidence from the Airline Industry." Unpublished.
- Conlon, Christopher T., and Julie Holland Mortimer.** 2008. "Demand Estimation under Incomplete Product Availability." National Bureau of Economic Research Working Paper 14315.
- Dana, James D., and Eugene Orlov.** 2009. "Internet Penetration and Capacity in the US Airline Industry." Northeastern University College of Business Administration Research Paper 08-001.
- Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su.** 2008. "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation." <http://www.ftc.gov/be/workshops/>.
- Forbes, Silke J.** 2008. "The Effect of Air Traffic Delays on Airline Prices." *International Journal of Industrial Organization*, 26(5): 1218–32.
- Gayle, Philip G.** 2004. "Does Price Matter? Price and Non-Price Competition in the Airline Industry." <http://www-personal.ksu.edu/~gaylep/comp.pdf>.
- Gillen, David W., William G. Morrison, and Christopher Stewart.** 2003. "Air Travel Demand Elasticities: Concepts, Issues and Measurement." Department of Finance Canada Report. [http://www.fin.gc.ca/consultresp/airtravel/airtravstdy\\_-eng.asp](http://www.fin.gc.ca/consultresp/airtravel/airtravstdy_-eng.asp).
- Goolsbee, Austan, and Chad Syverson.** 2008. "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines." *Quarterly Journal of Economics*, 123(4): 1611–33.
- Ito, Harumi, and Darin Lee.** 2003. "Incumbent Responses to Lower Cost Entry: Evidence from the U.S. Airline Industry." Brown University Department of Economics Working Paper 2003-22.
- McFadden, Daniel.** 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data with Econometric Applications*, ed. Charles F. Manski and Daniel McFadden, 198–272. Cambridge, MA: MIT Press.
- Mead, Kenneth M.** 2000. "Internet Sales of Airline Tickets." Statement of the Inspector General of the U.S. Department of Transportation before the Committee on Commerce, Science, and Transportation. Report CR-2000-111. <http://dot.cantongroup.com/sites/dot/files/pdfdocs/cr2000111.pdf>.
- MIT Airline Data Project.** <http://web.mit.edu/airlinedata/www/Traffic&Capacity.html>. (accessed April 1, 2009.)
- Morrison, Steven A., and Clifford Winston.** 2005. "What's Wrong with the Airline Industry? Diagnosis and Possible Cures." Statement, Hearing before the Subcommittee on Aviation Committee on Transportation and Infrastructure, United States House of Representatives, Washington, DC, September 28, 2005. <http://www.brookings.edu/views/testimony/20050928winston.pdf>.
- Morrison, Steven A., Clifford Winston, Elizabeth E. Bailey, and Alfred E. Kahn.** 1989. "Enhancing the Performance of the Deregulated Air Transportation System." *Brookings Papers on Economic Activity: Microeconomics*: 61–123.
- Mozdzanowska, Aleksandra L.** 2004. "Evaluation of Regional Jet Operating Patterns in the Continental United States." Master's thesis. MIT.
- Oum, Tae H., David W. Gillen, and S. E. Noble.** 1986. "December for Fare Classes and Pricing in Airline Markets." *Logistics and Transportation Review*, 22(3): 195–222

- Peters, Craig.** 2006. "Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry." *Journal of Law and Economics*, 49(2): 627–49.
- Puller, Steven L., Anirban Sengupta, and Steven N. Wiggins.** 2007. "Testing Theories of Price Dispersion and Scarcity Pricing in the Airline Industry." Unpublished.
- Reiss, Peter C., and Pablo T. Spiller.** 1989. "Competition and Entry in Small Airline Markets." *Journal of Law and Economics*, 32(S2): S179–202.
- SITA.** 2008. SITA Airline IT Trends Survey. [http://www.sita.aero/News\\_Centre/Publications/Air\\_Transport\\_IT\\_Review\\_Issue\\_2/Features/Airline\\_IT\\_Trends\\_Survey.htm](http://www.sita.aero/News_Centre/Publications/Air_Transport_IT_Review_Issue_2/Features/Airline_IT_Trends_Survey.htm). (accessed April 1, 2009.)



**This article has been cited by:**

1. Stephen Martin. 2012. Market Structure and Market Performance. *Review of Industrial Organization* . [[CrossRef](#)]
2. Dan Mahoney, Wesley W. Wilson Chapter 11 The Size and Growth of Airports **3**, 233-273. [[CrossRef](#)]
3. Anirban Mukherjee, Vrinda Kadiyali. 2011. Modeling Multichannel Home Video Demand in the U.S. Motion Picture Industry. *Journal of Marketing Research* 985-995. [[CrossRef](#)]