An Empirical Analysis of Marketing Alliances between Major US Airlines

Oliver M. Richard
University of Rochester
richard@ssb.rochester.edu

December 1999
Preliminary version; Do not quote.

Abstract

The paper proposes an empirical framework with endogenous flight frequency and quantity decisions on an airline market. The framework is built around the hypothesis that passengers value not only the ticket price but also the cost of delay associated with an airline. At sample mean values, the cost of delay is estimated to account for 34% of the $280 full price of a seat on a flight.

Marketing alliances increase the demand from connecting passengers for seats on a flight. Simulation results show that airlines respond to this increase by flying fewer local (nonstop) passengers rather than by increasing their flight frequency on a market.

1. Introduction

In January 98, Northwest and Continental Airlines announced that they were forming a marketing alliance and, in April 98, American and US Airways, Delta and United Airlines followed with similar statements. While marketing alliances had long been a part of the airline industry, this was the first time that major US airlines announced their intention to link up their domestic flight operations[1].

The definition of a marketing alliance between airlines is that the "carriers involved will be able to market and sell tickets on their partner's flights, a practice known as code sharing. Frequent flyer programs will also be linked, giving members more destinations to choose from" (New York Times, 04/24/98). This paper analyzes how code sharing changes, all else equal, the frequency, quantity, and pricing choices of airlines within alliances[2]. It is written amid growing concern in policy circles that ticket prices are rising and service quality is decreasing. The fear is that marketing alliances among the few remaining major airlines will only worsen these problems.

To better understand the concept of code sharing, some terminology is necessary. Airline customers are identified by an origin and a destination airport. For example, B-D customers fly from airport B to airport D. A local B-D passenger is a B-D customer who takes a nonstop flight on market B-D. A connecting B-D passenger travel on an

[1] Notes: [1] As of the summer of 1999, only the Northwest-Continental alliance has evolved to code sharing. The American-US Airways and Delta-United alliances have only, to this point, linked frequent flyer programs. [2] In the 1980's, while expanding their hub-and-spoke operations, major airlines entered into alliances with (and some eventually acquired) the commuter airlines feeding passengers to their hub airports from small and mid-size communities. In the recession of the early 1990's, US airlines turned to foreign carriers for alliances and, thereby, extended the reach of their domestic network abroad (i.e., cabotage and foreign ownership laws prevent an airline from offering flights within a foreign country).

[2] Issues such as whether alliances will encourage collusion, lead partners to serve new markets or exit from jointly served markets, are important topics, but they are corollaries to the definition of a marketing alliance. They are not considered in this paper. Note that there have always been secondary aspects to alliances. The 1980 alliances were a way for hub airlines to secure the pipeline of passengers to their hubs. Commuter airlines had also the know-how and aircraft equipment necessary to serve at lower costs small and mid-size communities. The 1990 alliances were acompanied by major cash and equity investments by foreign carriers into their US counterparts. There are also equity considerations to the 1998 alliance between Northwest and Continental. It is also argued that the US carriers are trying to extend the antitrust immunity which the Department of Justice granted international alliances, to their domestic alliances.

indirect route from airport B to D. An indirect route is a path made up of flights which links two airports and requires at least one stop at an intermediate airport. Hence, on a flight on a amrket B-D, an airline can, and does in practice, fly both local B-D and connecting passengers, where these connecting passsengers have for origin and/or destination an airport other than B or D.

It is a fact that connecting passengers dislike switching airlines on an indirect route and there are revenue (lack of) sharing incentives for airlines to keep a connecting passenger on its flights. Code sharing allows an airline to brand one of its partner's flights as its own and establishes a formal framework for revenue sharing. This expands the range of indirect routes an airline's flights can be a part of. This, in turn, increases the number of connecting passengers an airline can sell seats to on a flight.

The policy concern with code sharing is that, in response to the increase in the demand from connecting passengers, airlines will raise ticket prices for local passengers and, thereby, decrease the number of passengers flying nonstop. For example, Bamberger and Carlton (1993) provide empirical evidence to that effect.

I analyze this issue from an empirical framework which endogenizes airlines' flight frequency and quantity choices. This framework is built upon the hypothesis that a customer values not only the ticket price charged by an airline but the cost of delay associated with traveling with that airline; that is, customers value the *full price* of a seat. The full price is defined as the sum of the ticket price and the cost of delay.

The cost of delay specification incorporates both the positive and negative externalities that come with increased passenger volume on an airline market. A higher passenger volume on market increases the cost of delay (a negative externality) because a passenger is more likely to encounter crowded aircraft and longer than usual check-in lines, plane boarding/exiting times, and baggage retrieval times. However, higher passenger volumes have, at some point, to be accompanied by an increase in the number of flights. A higher frequency of flights decreases the cost of delay (a positive externality) since a passenger is more likely to find an flight departure closer to his desired departure time. This concept dates back to Douglas and Miller (1972), and full price specifications have long been a part of the theoretical airline literature (Dorman (1976), Panzar (1979), Berechman and Shy (1993), Lederer (1993)).

The model specifies aggregated demand functions for each of local and connecting passengers, and cost functions for each of flight frequency, connecting and local passengers. The estimated systems of equations are the first-order conditions and demand equations which characterize optimal/equilibrium flight frequency and quantity choices on market. From these and the estimation results (based on third quarter 1993 data), the reader is able to explicitly account for the interaction between flight frequency, quantity, and ticket prices on an airline market. There lies the benefits of this model. It is also the first empirical model, to my knowledge, to endogenize the strategic effects of flight frequency choices.

There are two main findings to the estimation results. First, the data provide support for the empirical framework in that estimated parameter values have signs consistent with the economic and institutional details of the model. For example, slope parameters have the correct signs and are significantly different from zero; customers assign a positive valuation to the cost of delay. Second, at sample mean values, the cost of delay is estimated to amount to 34% of the cost of delay of the $280 full price of a seat to a local passenger. The cost of frequency delay accounts for 90% ($85) of this cost of delay. These results also provide the marginal worth to a local passenger (flying coach) of an extra flight per day, per airline on a market. For example, if an airline increases its number of daily flights from 2 to 3 on a market, that additional flight is worth $18.52 to a local B-D passenger. If the increase is from 3 to 4 daily flights, that additional flight is worth $11.04.

Upon estimation of the model, I modify the data for the sample markets to reflect a setting with marketing alliances between each of Continental and Northwest, American and US Airways, Delta and United. This defines an *alliance sample data*. Given the estimated parameter values for the model, I simulate the flight frequency, quantities and prices for the markets in the *alliance sample data*.

Alliances increase the demand from connecting passengers for seats on a flight on a market. In response, the simulation results show that airlines increase the number of connecting passengers they serve. However, they respond to this increase by serving fewer local passengers rather than by increasing their flight frequency on the market. Fewer passengers fly nonstop in an industry with alliances.

A final comment is warranted. To estimate a complete model of airline interaction, I must parametrize the functional forms for the demand and costs functions in the model. Results are dependent upon these specification choices. It is therefore important to view the paper as providing an empirical framework with endogenous decision-making for the analysis of marketing alliances. There lies the primary interest of the paper.

(greater discussion of results, with numbers, forthcoming in next draft, due very shortly) for marketing alliances stress a very simple point. Ticket prices may rise for local passengers due to an increase in stochastic delay, but full prices may fall due to a decrease in frequency delay. Given local passengers value the full price of a seat, they stand to gain from alliances. More passengers fly nonstop on the sample markets with alliances.

The economic intuition is sraightforward. Since the marginal cost of a passenger is lower than that of flight frequency, an exogenous increase in the demand from connecting passengers increases more than proportionately the passenger volume relative to flight frequency. This raises ticket prices. However, frequency delay is assigned a higher dollar value than stochastic delay. This means that a marginal increase in flight frequency may lead to a greater decrease in the cost of delay than the increase in delay generated from the marginal increase in the passenger volume. The interest of the paper is that I endogenously estimate all relevant parameters to this argument. In fact, the analysis goes beyond this simple argument since the model provides the expected frequency, quantity, and prices for market with alliances.

Section 2 presents the theoretical components to the model. Section 3 introduces the reader to the data. Section 4 discusses the empirical model and its estimation is detailed in section 5. Section 6 contains the results and section 7 concludes. Most tables are listed after the conclusion. The appendices provide additional information on the data, the model, the estimation and the results.

2. The Theoretical Model

2.1. Flights and Indirect Routes

A flight denotes a single nonstop (direct) trip identified by an origin and a destination airport and a departure time. An indirect route is a path made up of flights which links two airports and requires at least one stop at an intermediate airport. Flights and indirect routes are, to the extent both are present on an airport-pair market, substitutes.

Customers are identified by an origin and destination airport. For example, B-D customers have airport B for origin and airport D for destination. A local B-D passenger is a B-D customer who takes a flight on market B-D. A connecting B-D passenger is a B-D customer who travels on an indirect route from airport B to airport D. An airline with flight(s) on a market from airport B to airport D (i.e., market B-D) is competing for B-D passengers with airlines offering indirect routes on market B-D (see figure 1).

2.2. Amendments to the Definition of an Indirect Route

Connecting passengers who take all flights on an indirect route with the same airline are said to connect online. If a connecting passenger switches airlines on an indirect route, that passenger is said to interline. It is a fact that connecting passengers prefer online to interline connections and that there are revenue (lack of) sharing incentives for airlines to keep a connecting passenger on its flights[3]. Morrison and Winston (1995,

[3]Carlton, Landes & Posner (1980) estimate that travelers would be willing to pay nearly $37 (in 1993 dollars) more for a flight with an on-line connection than one with an interline connection. The benefits from on-line connections include a shorter walk in the terminal to ctach the connecting flight, greater

p.22) document that, by 1994, fewer than 1% of connecting passengers switch airlines on their indirect routes. This leads me to restrict the definition of an indirect route to paths where all flights are served by the same airline.

An airline offers flights across a network of markets. It can therefore include a flight on a market B-D into the path of indirect routes between some airports A_i and E_j, with $A_i \neq B$ and/or $E_j \neq D$ (see figure 1 for an illustration). For example, airport A_i (E_i) may be any origin (destination) airport on a market to airport B (from airport D) where the airline has flights. Given I do not model the choice of departure times for flights, I do not differentiate among departure times for indirect routes. Namely, an indirect route on market B-D is redefined as a path made up of markets with flights from the same airline which links airports B and D and requires at least one stop at an intermediate airport. To illustrate, all indirect routes from Seattle (SEA) to Miami (MIA) made up of a SEA-Dallas (DFW) and a DFW-MIA flight count as one indirect route from SEA to MIA, irrespective of differences in departure times.

The number of indirect routes a flight on market B-D can be a part of is, therefore, proportional to the number of markets which an airline services with flights, into airport B and from airport D. The greater the number of these markets, the greater the number of A_i-E_j markets the airline may draw connecting passengers from. In this paper, an airline with flights on market B-D is said to face the demands from two aggregated groups of customers: local B-D and connecting A-E customers.

2.3. Marketing Alliances

The concept of a marketing alliance is that the "carriers involved will be able to market and sell tickets on their partner's flights, a practice known as code sharing. Frequent flyer programs will also be linked, giving members more destinations to choose from" (New York Times, 04/24/98). If airlines 1 and 2 enter in a marketing alliance, markets with flights from airline 2 can now be used by airline 1 to create new paths for indirect routes

coordination with the connecting flight (especially in case of delays),and less chance of lost luggage (source: Morrison and Winston (1995), p. 23).

(and vice-versa). Since airline networks overlap, this changes the range of connecting A-E customers each airline may sell seats to on a B-D flight. All else equal, the demand from connecting A-E passengers for seats on a B-D flight should increase relative to the demand from local B-D passengers.

2.4. Concern for Local Passengers

The policy concern is that the increase in the demand from connecting passengers may lead to an increase in the prices for local passengers and, thus, a decrease in the number of passengers flying nonstop.

This argument presumes the existence of negative externalities between local and connecting passengers. While a model with increasing aircraft marginal costs per passenger would allow for such externalities, the airline environment may not be so easily characterized. Indeed, the empirical literature provides evidence of economies of density in aircraft costs on a market: holding flight frequency constant, higher passenger volumes on a market accommodated through large aircraft or denser seat configurations, lower aircraft costs per passenger (Christensen, Caves, & Threteway (1984), Brueckner & Spiller (1991), Berry, Carnall & Spiller (1996)).

2.5. Literature review: A Capacity Story

In an empirical analysis of hub airports, Bamberger and Carlton (1993) provide a basis for negative externalities among local and connecting customers. They formulate a setting with customer heterogeneity, capacity constraints and demand uncertainty. In short, local and connecting customers differ in their willingness to pay for a seat on a flight and in the time at which they are expected to join the demand for seats on flights. There is uncertainty in the number of each type of customers. For a given seat capacity, an exogenous increase in the expected demand from connecting passengers, some of whom have high willingness to pay for a seat, would, under appropriate conditions, decrease the likelihood that airlines sell a given seat to a local passenger. This would decrease the expected number of seats allocated to local passengers and increase their expected ticket fare[4].

Empirically, Bamberger and Carlton regress the average ticket prices for local B-D passengers on the flight frequency on market B-D, the percentage of all B-D passengers which are connecting passengers, and a number of descriptive market variables. They find that local fares are higher on markets where a higher percentage of all passengers are connecting.

There are several problems with such a capacity-based/seat allocation story. Theoretically, these models are dynamic ones given new information on bookings for a particular flight becomes available over time. Their formulation and resolution require stringent assumptions on how information is revealed over time and on how the endogenous variables may be updated in light of the new information. As Bamberger and Carlton note, airlines (and their yield management groups) struggle mightily to solve this seat allocation problem despite making considerable simplifying assumptions. Empirically, the necessary data to develop this story into a structural empirical framework are not available. Namely, data on seat allocations per customer type, selling dates for tickets, and individual load factors for particular flights are not publicly available.

2.6. Flight Frequency Valuation

The airline environment is further complicated by the presence of positive externalities between local and connecting passengers. Airline customers' desired departure times are distributed on the 24-hr clock and there are only a small, discrete number of flights (or indirect routes) scheduled on a market. This means that customers suffer a time difference between their desired departure time and their actual departure time given the schedule. Customers are said to incur a cost of frequency delay. Given aircraft have limited seat capacity, an increase in the passenger volume on a market must, at some point, be met with additional flights. As an airline adds flights to a market, Richard

[4] Dana (?) provide a similar analysis (check reference).

(1999) shows that it spreads its departures over daytime hours. This decreases (non-linearly, in standard spatial games à la Hotelling[5]) the distance between a customer's actual and desired departure times and reduces the cost of frequency delay to customers. In the context of the paper, an increase in the number of connecting passengers on a market may increase the number of flights on that market and this, in turn, reduces the cost of delay to local passengers. The number of local passengers on the market may increase.

There is ample casual and empirical evidence that such positive externalities exist. Morrison and Winston (1995)'s comprehensive analysis of airline behavior and traveler welfare shows that passengers value increases in flight frequency. Bamberger and Carlton also extend their analysis to document empirically (with regression work similar to that outlined above) that, albeit fares for local passengers have risen on markets with hub airports, the number of flights on these markets has increased, in reponse to a rise in the number of connecting passengers at hub airports, and this has lead to an increase in the number of local passengers despite the fare hike.

2.7. Full Price Specification

This discussion leads me to characterize the airline environment with a full price model. In this model, customers value not the ticket price but the full price of a seat on a flight. Formally, the full price of a seat on a flight is defined as the sum of the ticket price charged by the airline and the cost of delay of flying on the airline on that market.

The cost of delay function is defined as the sum of the cost of frequency delay (see earlier discussion) and the cost of stochastic delay. Frequency delay assigns a dollar value to the difference, in minutes, between a customer's desired departure time and his/her actual departure time given the schedule. It is, in this paper, a monotonic, decreasing function of the number of flights the airline offers on the market.

[5] For example, consider a spatial model à la Hotelling where customers are uniformly distributed on the [0,1] line segment. Assume that there is only one firm and that it covers the market with 1 flight. If the firm locates optimaly n outlets, a customer must at most travel a distance of 1/(n+1) to purchase from an outlet.

Customers' desired departure times are distributed on the 24hr-clock and, as an airline adds flights to a market, Richard (1999) shows that it spreads its departures over daytime hours. In standard spatial games à la Hotelling, we may expect this to yield a non-linear decrease in the distance between a customer's actual and desired departure times[6].

Stochastic delay assigns a dollar value to airline-specific delays which result whenever the passenger encounters longer than usual check-in lines, plane boarding/exiting times, and longer baggage retrieval times. It is a function of the identity of an airline, its passenger volume and its number of flights on the market. In that sense, stochastic delay introduces a measure of differentiation among airlines with flights. Since passengers are known to dislike high passenger volumes on aircraft and crowded gate facilities[7], this function is monotonically increasing in the passenger volume, thereby introducing negative externalities into the model.

The basic premise of this full price specification is that all customers agree that a lower passenger volume and a higher flight frequency is better. The only loss of generality comes from all customers transforming, on a per-airline basis, these attributes into dollar values in precisely the same way. The stochastic and frequency delay functions may themselves be redefined as proxies for service quality where service quality increases as the cost of delay decreases. These concepts of frequency and stochastic delays (and their definition) date back to Douglas and Miller (1972), and full price specifications have long been a part of the theoretical airline literature (Dorman (1976), Panzar (1979), Berechman and Shy (1993), Lederer (1993)). This paper is the first, to the best of my

[6]For example, consider a spatial model à la Hotelling where customers are uniformly distributed on the [0,1] line segment. Assume that there is only one firm and that it covers the market with 1 flight. If the firm locates optimaly n outlets, a customer must at most travel a distance of 1/(n+1) to purchase from an outlet.

[7]A review of recent Congressional inquiries into passenger welfare easily attest to this. On an anecdotal level, Pieter Bouw, managing director and 22-year veteran of KLM Airlines, writes: "most significant is that overcrowded terminals, runway congestion, substantial delays, and missed connections result in a loss of service to customers. Just think about missed business meetings and frustrated holidays; less comfort and more uncertainty; and longer transfer times and lost baggage as a result of missed connections. The most important implication is that passengers are avoiding the airline system." [Airline Business, June 1989].

knowledge, to empirically estimate such a specification.

2.8. The Model

The model is defined for a market B-D and built to analyze the flight frequency and quantity decisions of airlines with flights (i.e., airlines which offer a strictly positive number of flights) on market B-D. The entry decisions of these airlines on market B-D are, at this point, taken as a given. Entry is discussed, later on, when talking about the estimation methodology. The setting is one of perfect information and the number and identity of airlines with flights on market B-D are, thus, known to all.

In the model, K entrants (airlines with flights) on market B-D simultaneously select their flight frequency and number of local B-D and connecting A-E passengers. Frequency and quantity decisions are chosen to be simultaneous given aircraft schedules remain rather flexible once published. Flights can be added or cancelled and the cost of doing so, while not necessarily low, is not prohibitive.

There are three levels to the economic environment on market B-D: [1] competition for local B-D passengers among airlines with flights on market B-D; [2] competition for B-D passengers among airlines with flights and airlines with indirect routes on market B-D; [3] competition for connecting A_i-E_j passengers among airline with flights on market B-D.

Let me start with describing the competitive setting in [1]. The K entrants face one aggregated market demand from local B-D passengers. These passengers value the full price of a seat on a B-D flight and are assumed to have identical and rational expectations; that is, their expectations are fulfilled in equilibrium (re: Katz and Shapiro (1985)). They purchase from the airline with lowest full price and, in equilibrium, ticket prices adjust across the K airlines to yield full price equality:

$$\overbrace{p_k^l + \underbrace{delay\left(f_k, q_k^l + q_k^c, k\right)}_{\text{cost of delay for airline k}}}^{\text{full price for airline k}} \quad = \quad fct\left(exog, \sum_{m=1}^{K} q_m^l, q^i\right)$$

$$with$$

$$p_k^l + delay\left(f_k, q_k^l + q_k^c, k\right) \quad = \quad p_h^l + delay\left(f_h, q_h^l + q_h^c, h\right)$$

$$for\ k \quad \neq \quad h \qquad\qquad k, h = 1, ..., K$$

where the subscripts k, h identify the airlines and the superscripts l, c, i identify, respectively, a variable for local B-D, connecting A-E and connecting B-D passengers. Namely, p_k^l denotes airline k's ticket price for local B-D passengers, f_k is the number of flights for airline k on market B-D, q_k^l is the number of local passengers carried by airline k, q_k^c is the number of connecting A_i-E_j passengers carried by airline k, and q^i is the number of connecting B-D passengers carried by airlines with indirect routes on market B-D. The term *exog* stands for exogenous market characteristics.

The specification allows for flights and indirect routes on market B-D to be imperfect substitutes. On a travel time basis, Richard (1999) reports that the median length of a flight is, in July 1993, 70 minutes (average is 87 minutes) while the median length of an indirect route is 314 minutes (with layovers; this statistics covers indirect routes with one and two stops).

Turning now to the competitive settings in [2] and [3], these present a formidable obstacle. When it comes to connecting passengers and indirect routes, the competitive environments for all markets which are part of an indirect route come into play. To carry a connecting A-E passenger on a flight on market B-D requires carrying this passenger on all other flights of her indirect A-E route. Hence, the decision to carry this passenger has to be made globally across all markets this passenger will travel through. However, the operations research literature on the airline industry clearly documents that the sheer size of airline networks and heterogeneity across airline markets prohibit any analysis of global decision-making across markets.

Reiss and Spiller (1989) is, to my knowledge, the only paper to attempt to model the strategic interactions between airlines with flights and airlines with indirect routes on a market (i.e., setting [2]). They consider the decision of a single airline to offer flights on a market when it is faced with competition from airlines with indirect routes on that market. From a model with differentiated products, they find that Cournot conjectures best characterize the competitive setting between the airline with flight(s) and its indirect competitors.

I propose turn to the data to examine whether any simplifying assumptions may be given some basis for.

3. The Data

The data are from the third quarter of 1993 and are compiled from five databases[2]. Databank 1A, a database from the Department of Transportation (DOT), is a 10% random sample of all airline tickets sold each yearly quarter. It provides, for this analysis, the local and connecting ticket prices (with fare class) per airline and the ratio of local to connecting passengers per airline, per market. The data do not include information on the time-of-day or day of travel, on possible ticket restrictions, or on frequent flyer status. Databank DS T-100, another DOT database, provides the total number of passengers and flights per route, per month. Only the major U.S. airlines and their (directly-owned) subsidiaries (e.g., shuttle, commuter airlines) are required to report to Databank T-100. The Official Airline Guide North American editions provide a complete listing of all scheduled flight operations for the summer of 1993. Routes with nonstop flight operations not reported to Databank T-100 are deleted since I have no passenger totals for these routes. The data in Databank T-100 identify the type of aircraft flown on a route. The Aircraft Quarterly Operating Costs and Statistics periodical, an AVMARK Inc. publication which summarizes Form 41 data, details operating costs per aircraft type, per airline. I reconcile both databases to create airline cost variables. The Census data provide the data on market demographics and characteristics. All necessary details on the construction of the sample data are provided in appendix A.

The available data limit the scope of my analysis in several ways. First, since Databank 1A is on a quarterly basis, all decision variables (e.g., flight frequency, quantities) are defined on a quarterly basis. This paper can therefore be characterized as an analysis of expected quarterly flight frequency and passenger volumes. Second, while airlines report to Databank 1A the fare classes for tickets, the data on first-class ticket prices are notoriously unreliable. I limit the analysis to coach/economy class passenger volumes[3]. Third, only nine percent of airline markets have three or more airlines with flights. To anticipate, incorporating these markets in the estimation phase of the paper considerably burdens the empirical structure while it adds few observations. I proceed to limit the scope of the empirical analysis to markets with at most two airlines with flights. Fourth, sample markets are directional and only one direction per market is included in the sample data. This yields a sample of 605 markets with a single airline with flights and 185 markets with two airlines with flights.

4. The Empirical Model

This section describes the functional forms for the model and provides an empirical framework for cross-market strategic interactions. Three important remarks are warranted. First, the nature of the model requires parametric specifications for the functional forms. Second, I select simple functional forms (i.e., linear, log-linear specifications) which are commonly found throughout the empirical literature at-large. Third, the empirical framework in this paper is flexible on the choice of functional forms. Results have shown much robustness, across trial estimation runs, to changes in the specifications.

4.1. The Local B-D Demand

Let there be K airlines with flights on market B-D. The inverse demand function for local B-D demand is linear:

$$p_k^l + \underbrace{\frac{\gamma^l + \alpha_k \left(q_k^l + q_k^c\right)}{\sqrt{f_k}}}_{\text{cost of delay for local B-D passenger}} = exog^d - \beta^l \sum_{k=1}^{K} q_k^l - \beta^i q^i \qquad with$$

$$p_k^l + \frac{\gamma^l + \alpha_k \left(q_k^l + q_k^c\right)}{\sqrt{f_k}} = p_h^l + \frac{\gamma^l + \alpha_h \left(q_h^l + q_h^c\right)}{\sqrt{f_h}} \qquad \forall\, k, h = 1, .., K$$

$$exog^d = \beta_0 + \beta_1 \ln(POP) + \beta_2 POPSQK + \beta_3 INC + \beta_4 MILES + u_1$$

where, to recall, p_k^l is airline k's ticket price for local passengers, f_k is the number of flights for airline k on market B-D, q_k^l is the number of local passengers carried by airline k, q_k^c is the number of connecting A-E passengers carried by airline k, and q^i is the number of connecting B-D passengers carried by airlines with indirect routes on market B-D. Symbols $\alpha_k, \gamma^l, \beta_0, ..., \beta_4$ are parameters to be estimated. The term u_1 accounts for unobserved demand factors (i.e., I do not observe any of the demand and cost functions).

The cost of delay specification is:

$$\text{Cost of delay on airline k} = \frac{\gamma^l + \alpha_k \left(q_k^l + q_k^c\right)}{\sqrt{f_k}} \qquad with$$

$$\text{frequency delay } = \frac{\gamma^l}{\sqrt{f_k}} \quad \text{and} \quad \text{stochastic delay } = \frac{\alpha_k \left(q_k^l + q_k^c\right)}{\sqrt{f_k}}$$

Frequency delay is non-linear in the flight frequency and stochastic delay is related to the passenger volume per aircraft. To reflect that higher flight frequency is nevertheless accompanied by higher overall passenger volumes, stochastic delay is reduced less than proportionately by increases in the number of flights.

The variables in the $exog^d$ term measure the size of the demand on market B-D.

These variables include the mileage of the market ($MILES$), and the (natural log of) average population (POP), population per square miles ($POPSQK$), and income level (INC) across the cities for each of airports B and D.

4.2. Competition between Nonstop and Indirect Airlines

I now examine the data on sample markets to see whether any simplifying assumptions on cross-market strategic interactions may be given some basis for. All of the numbers I discuss below are quarterly averages, unless otherwise mentioned, for the sample markets based on observed flight schedules and ticket sales. Since the distributions for the sample statistics are rather symmetric around their means/medians, these figures provide representative information on the data.

Connecting A-E passengers account for 58% of an airline's passenger volume on a sample market B-D. The airline draws connecting A-E passengers from 57 indirect routes with one intermediate stop and 54 indirect routes with two stops. Indirect routes with one stop contribute twelve times more connecting A-E passengers than indirect routes with two stops.

Connecting passengers are rather evenly drawn from across markets A-E since each such market contributes, on average, 1.68% (median is 0.89) of the airline's total connecting A-E passenger volume. Thirty percent of the markets A-E an airline draws connecting passengers from have nonstop flights. The airline draws 28% of all of its connecting A-E passengers from markets A-E with flights.

On a sample market B-D, 18% of all B-D passengers travel on an indirect B-D route rather than a B-D flight. There are 6 airlines (i.e., indirect B-D airlines) carrying passengers across a total of 18 different indirect routes. The median herfindal index of concentration in connecting B-D passengers across these airlines is 0.34 (mean is 0.39). In other words, out of an average of 6 airlines carrying connecting B-D passengers, there are about 2.94 (= 1/herfindal = 1/0.34) effective competitors. Given 91% of markets with flights have two or fewer airlines with flights, this represents, by airline standards, a fair amount of competition.

In addition, an indirect B-D airline with multiple indirect B-D route carries the bulk of its connecting B-D passengers across two indirect routes. To the extent a market has about K effective competitors per market, this means that connecting B-D passengers spread, on average, across 6 indirect B-D routes. In fact, the number of indirect B-D routes is more highly correlated with the number of connecting B-D passengers than the number of airlines with indirect B-D routes (correlation of 0.63 vs. 0.55). There is also some issue as to how much coordination there is for a given airline across its indirect routes on a market. A substantial amount of coordination demands a global approach decision-making and we are back to the basic problem, which motivates this dicussion, of how to model such a setting.

For markets with no flights (i.e., in the data at large), there are 2 airlines offering a total of 3.5 indirect routes. The number of effective competitors (in terms of passenger volume) is 1.3 on these markets. The total passenger volume on markets with no flights is eleven time smaller on average than that on market with flights.

It is clear from this statistical summary that the data provide no definite answers as to how to model cross-market strategic interactions. Nevertheless, a few traits can be highlighted. First, airlines carry connecting passengers from a large number of markets and passengers from each of these markets represent a very small percentage of the airline's total passenger volume on market B-D (i.e., 1.68% of 58% = 0.97%). Second, there are a fair, by airline standards, number of airlines with indirect routes on a sample market and connecting B-D passengers are carried across a large number of different indirect B-D routes.

In light of these statistics, I specify a revenue schedule specification for connecting A-E passengers. This schedule provides the number of connecting A-E passengers an airline may carry for a given full price. Namely, an airline with flights on market B-D is assumed to act as a price-taker with regards to its connecting A-E demand. In the model, the airline selects the average ticket price for its connecting A-E passengers, where this average price is computed across all observed prices for the entire itinerary (i.e., across all flights on the indirect routes) of all of its connecting A-E passengers on market B-D in the third quarter of 1993.

Connecting A-E passengers value the full-price of a seat on a B-D flight with airline k. The full price for a connecting passenger is defined as the sum of the ticket price and the cost of delay associated with airline k. The specification is:

$$\text{Full price for airline k } = p_k^c + \frac{\gamma^c + \alpha_k\left(q_k^l + q_k^c\right)}{\sqrt{f_k}} + \delta_{5k} DUM_k \qquad with$$

$$\text{Cost of delay on airline k } = \frac{\gamma^c + \alpha_k\left(q_k^l + q_k^c\right)}{\sqrt{f_k}} + \delta_{5k} DUM_k \qquad and$$

$$\text{cost of frequency delay} = \frac{\gamma^c}{\sqrt{f_k}} \qquad \text{cost of stochastic delay} = \frac{\alpha_k\left(q_k^l + q_k^c\right)}{\sqrt{f_k}}$$

DUM_k proxies for the cost of delay on the non B-D flight(s) of the indirect route

where p_k^c is the average ticket price paid by a connecting passenger for her entire itinerary (i.e., across all flights on her indirect route), and $\gamma^c, \alpha_k, \delta_{5k}$ are parameters to be estimated.

The frequency delay valuation parameter γ^c is allowed to differ across the specifications for local and connecting passengers. A flight on market B-D is only one flight on a connecting passenger's indirect route and this passenger should, thus, be expected to value it differently than a local passenger. The stochastic delay specification for flight B-D is assumed identical across both local and connecting passengers; this assumption is made to simplify the structure of the estimated models. Finally, given the aggregated nature of the analysis, the cost of delay on the non B-D flight(s) of a passenger's indirect route may not be accounted for and I proxy for it with an airline's specific dummy variable.

The revenue schedule for connecting A-E demand for airline k on market B-D is specified as a log-linear function:

$$\ln\left(q_k^c\right) = exog_k^c - \delta\left(p_k^c + \frac{\gamma^c + \alpha_k\left(q_k^l + q_k^c\right)}{\sqrt{f_k}} + \delta_{5k} DUM_k\right) \qquad with$$

$$exog_k^c = \delta_0 + \delta_1 \ln(ROUT_k) + \delta_2 RMILES_k + \delta 3 HUB_k + \delta_4 \ln(MILES) + u_{2k}$$

where $\delta, \delta_0, ..., \delta_4$ are parameters to be estimated, and u_{2k} accounts for unobserved factors.

Characteristics for airline k on market B-D are the number of *potential* indirect routes including flight B-D ($ROUT_k$) and the average of the ratio of the mileage of each potential indirect route to the shortest mileage itinerary available for that market ($RMILES_k$). These variables are based upon indirect routes constructed, by appending markets with flights, from the same airline, to one another (irrespective of flight times) given entry decisions across non B-D markets (see appendix B for ampler details). The variable HUB_k is a dummy variable denoting the presence of a hub for airline k at airports B and/or D[8]. I expect the presence of a hub for airline k on the B-D market to increase the demand from connecting passengers for seats on a B-D flight.

4.2.1. The Reaction Function for Connecting B-D Passengers on Market B-D

Each indirect airline on market B-D decides upon the number of connecting B-D passengers it carries within the context of its quantity and frequency decisions on the markets which make up its indirect route(s). As documented in the previous sub-section, connecting B-D passengers make up only a very small fraction of the airline's total passenger volume on these markets. Given the assumed empirical framework for connecting A-E passengers, I treat the airlines with indirect routes as a price-taking competitive fringe and specify a reaction function for the total number of connecting B-D passengers. This

[8] These airports are designated as hub airports for the third quarter of 1993: [1] Minneapolis-St Paul, Detroit DTW, Memphis for Northwest; [2] Atlanta, Salt Lake City, Cincinnati, Dallas-Fort Worth for Delta; [3] Chicago O'Hare, Denver, LAX, San Francisco SFO, Washington IAD for United; [4] Pittsburgh, Charlotte, Baltimore BWI, Philadelphia for USAir; [5] Dallas-Fort Worth, Chicago O'Hare, Miami, Nashville, Raleigh-Durham for American; [6] St Louis for TWA; [7] Houston IAH, Newark, Cleveland, Denver for Continental; [8] Phoenix for America West.

Only 10% of sample markets do not have a hub airline with flights.

function is increasing in the full price of a seat on a flight on market B-D:

$$\begin{aligned} q^i &= exog^i - \lambda \left(p_k^l + \frac{\gamma^l + \alpha_k \left(q_k^l + q_k^c \right)}{\sqrt{f_k}} \right) \qquad with \\ exog^i &= \lambda_0 + \lambda_1 AREA + \lambda_2 ICON + \lambda_3 \ln(MILES) + \lambda_4 IRATE + u_3 \end{aligned}$$

where $\lambda, \lambda_0, ..., \lambda_4$ are parameters, u_3 is an error term, and q^i is the total number of connecting B-D passengers for market B-D. These passengers travel on indirect routes from airport B to airport D.

Variables for market B-D include the square miles area for the cities including airports B and D ($AREA$; i.e., a demand variable), the number of *potential* indirect routes on market B-D ($ICON$), the ratio of their average mileage to the mileage of market B-D ($IRATE$).

4.2.2. Cost Functions

The Aircraft Quarterly Operating Costs and Statistics periodicals contain airline cost data, for the first quarter of 1993, on a per hour of utilization, per aircraft type, per airline basis. The data in Databank T-100 provides the type of aircraft utilized on a market at that time. I combine the information from both databases to construct airline cost variables for the sample markets (which are third quarter data). These variables help specify the cost functions of the model and are instrumental in its identification. The variables were assigned to the various cost functions on the basis of regression work (not documented).

Conceptually, airline costs on market B-D can be defined as a function of the number of flights and number of passengers on market B-D:

$$\begin{aligned} &\text{airline costs}\,(\#\text{ of flights, }\#\text{ of passengers}) = \text{cost}(\#\text{ of passengers}) \\ &+\ [\#\text{ of flights}] \times [\text{cost per flight}\,(\#\text{ of passengers per flight, cost variables})] \end{aligned}$$

For the purpose of this paper, I write

$$\begin{aligned}\text{costs for airline k} &= \cos t_k^l q_k^l + \cos t_k^c q_k^c + f_k \left(\cos t_k \left(\frac{q_k^l + q_k^c}{f_k} \right) + \cos t_k^f \right) \\ &= \left(\underbrace{\cos t_k^l + \cos t_k}_{=_{notation} \cos t_k^l} \right) q_k^l + \left(\underbrace{\cos t_k^c + \cos t_k}_{=_{notation} \cos t_k^c} \right) q_k^c + f_k \left(\cos t_k^f \right)\end{aligned}$$

where the terms $\cos t_k^f$, $\cos t_k$, $\cos t_k^l$ and $\cos t_k^c$ represent linear combinations of cost characteristics for airline k.

The hub-and-spoke structure of airline networks is a defining characteristic of this industry and there is empirical evidence in the literature that hub airports raise the hub airlines' yields and ticket prices for local passengers (Borenstein (1989), General Accounting Office (1988-1998), Morrison & Winston (1995)). In particular, the system of *bank* scheduling at hub airports (i.e., flight arrivals and departures are concentrated at given time periods) has been argued to promote inefficient usage of airport facilities and aircraft resources[9]. This would raise flight frequency and passenger costs at the hub airports.

Each of the cost specifications in the paper includes two hub variables: HUB_k, which is airline-specific, and $MAJHUB$. This latter variable denotes whether each of airports B and/ D are one of the major US hubs (e.g., Atlanta, Dallas Forth-Worth, Houston IAH, Los Angeles LAX, Miami, New York JFK, and Chicago O'Hare).

The contribution to the literature is that, within a structural framework with endogenous decision-making, I identify, and provide measures for, the impact of hub airports on the flight frequency, local and connecting ticket prices. The analysis is from a cost-based perspective albeit, it must be said, the use of dummy variables to account for

[9] In 1994, the C.E.O. of Continental Lite (a former point-to-point subsidiary of Continental Airlines) explains that "CO Lite wants 9.5 gate turns per day [only 4 gate turns per day at CO's hubs], 11.3 hours of aircraft utilization, and 6 hours 'hard flying' per crew-member. These are the things that drive costs. Although Denver [a CO hub] and Greensboro [not a CO hub] have roughly the same number of flights per day, because of the way the banks run at Denver, we've got to have 20 gates manned between 1pm and 2pm. Greensboro only has six gates and they are always manned." [Air Transport World, June 1994].

hub airports does not prevent some type of fixed (i.e., independent of quantity levels) markup interpretation[10]. For example, Borenstein (1989) argues that factors such as frequent flyer programs, close ties to travel agencies, greater advertising exposure, provide comparative advantages to hub airlines at their hub airports.

For the marginal cost per flight, I have:

$$\begin{aligned}\cos t_k^f &= \exp(\kappa_0 + \kappa_1 FUEL_k + \kappa_2 CASM_k + \kappa_3 HUB_k + \kappa_4 MAJHUB \\ &\quad + \kappa_5 SLOT + \kappa_6 \ln(MILES) + \kappa_7 DL_k + u_{4k})\end{aligned}$$

where u_{4k} accounts for unobserved factors and κ_0,...,κ_6 are parameters to be estimated. The airline-specific cost variables are aircraft fuel costs ($FUEL_k$) and operating costs per available seat-mile ($CASM_k$). This latter variable is a standard measure of costs in the literature. The variable $SLOT$ is a dummy denoting the presence of landing and take-off lots at Chicago O'Hare, New York JFK and Laguardia, and Washington National airports. The dummy variable DL_k denotes whether or not airline k is Delta Airlines. This dummy variable was added during the estimation process as it proved to significantly improve the fit of the model.

If the specifications for the $\cos t_k^l$ and $\cos t_k^c$ terms should conceptually share some variables and parameters, I do not impose parameter restrictions across both specifications and they include some different variables. The main reason is that the ticket prices for connecting and local passengers are measured over differing itineraries since the ticket prices for connecting passengers apply to their entire itinerary, not just for the flight on market B-D. Hence, I use mileage characteristics for the indirect routes rather than market B-D in the marginal cost function for connecting passengers.

For the marginal cost per local passenger, I specify:

$$\cos t_k^l = \omega_0 + \omega_1 \ln(CSBH_k) + \omega_2 MAINT_k + \omega_3 MAJHUB + \omega_4 MILES +$$

[10]The objective of the paper is not a detailed analysis of pricing policies at hub airports. This explains the approach I have opted for. To introduce expliclty a mark-up story would require a more complicated demand system and a different model.

$$+\omega_5 HUB_k \times MILES + u_{5k}$$

where u_{5k} accounts for unobserved factors and ω_0,...,ω_5 are parameters to be estimated. The airline cost variables are the (natural log of) operating costs per seat, per hour of utilization ($CSBH_k$) and aircraft maintenance costs ($MAINT_k$). Costs may possibly differ at hub airports and I include two hub variable, $MAJHUB$ and an airline-specific interaction variable, $HUB_k \times MILES$[11]. Mileage is one of the primary determinants of airline costs and an interaction variable offers, here, more interest at the interpretation level than a simple dummy (note that it also yields a better fit).

For the marginal cost per connecting passenger, I have:

$$\cos t_1^c = \psi_0 + \psi_1 TOTMIL_k + \psi_2 \ln(CSBH_k) + \psi_3 MAJHUB + \psi_4 HUB_k + u_{6k}$$

where u_{6k} accounts for unobserved factors and ψ_0,...,ψ_4 are parameters to be estimated. The $TOTMIL_k$ variable denotes the average total mileage for all *potential* indirect routes from airline k which include flight B-D. In this specification, a dummy variable for hub effects, HUB_k, proves a better fit than an interaction variable.

4.2.3. Profit-Maximization Problem: Market with one Nonstop Airline

Given these choices of functional forms, a single airline (e.g., airline 1) on a market B-D selects its flight frequency and quantities of passengers to maximize profits:

$$\max_{q_1^l, q_1^c, f_1} \quad p_1^l q_1^l + p_1^c q_1^c - \cos t_1^l q_1^l - \cos t_1^c q_1^c - \cos t_1^f f_1$$

$$such \quad that$$

$$p_1^l + \frac{\gamma^l + \alpha_1 \left(q_1^l + q_1^c\right)}{\sqrt{f_1}} = exog^d - \beta^l q_1^l - \beta^i q^i \qquad \text{(M.1)}$$

$$\ln\left(q_1^c\right) = exog_1^c - \delta \left(p_1^c + \frac{\gamma^c + \alpha_1 \left(q_1^l + q_1^c\right)}{\sqrt{f_1}} + \delta_{51} DUM_1\right) \qquad \text{(M.2)}$$

[11] I recognize, however, that some of the value of the estimated parameter a_5 could be attributed to market power at the hub, rather than to cost levels.

$$q^i = exog^i - \lambda \left(p_1^l + \frac{\gamma^l + \alpha_1 \left(q_1^l + q_1^c \right)}{\sqrt{f_1}} \right) \tag{M.3}$$

There are three equations to the demand system: equations M.1 to M.3.

4.2.4. Profit-Maximization Problem: Market with two Nonstop Airlines

Airlines 1 and 2 on a market B-D select their flight frequency and quantities of passengers to maximize profits:

$$\max_{q_1^l, q_1^c, f_1} \quad p_1^l q_1^l + p_1^c q_1^c - \cos \mathrm{t}_1^l q_1^l - \cos \mathrm{t}_1^c q_1^c - \cos \mathrm{t}_1^f f_1$$

$$\max_{q_2^l, q_2^c, f_2} \quad p_2^l q_2^l + p_2^c q_2^c - \cos \mathrm{t}_2^l q_2^l - \cos \mathrm{t}_2^c q_2^c - \cos \mathrm{t}_2^f f_2$$

$$such \quad that$$

$$p_1^l + \frac{\gamma^l + \alpha_1 \left(q_1^l + q_1^c \right)}{\sqrt{f_1}} = exog^d - \beta^l \left(q_1^l + q_2^l \right) - \beta^i q^i \tag{D.1}$$

$$p_1^l + \frac{\gamma^l + \alpha_1 \left(q_1^l + q_1^c \right)}{\sqrt{f_1}} = p_2^l + \frac{\gamma^l + \alpha_2 \left(q_2^l + q_2^c \right)}{\sqrt{f_2}} \tag{D.2}$$

$$\ln \left(q_k^c \right) = exog_k^c - \delta^c \left(p_k^c + \frac{\gamma^c + \alpha_k \left(q_k^l + q_k^c \right)}{\sqrt{f_k}} + \delta_{5k} DUM_k \right) \quad k = 1, 2 \tag{D.3,D.4}$$

$$q^i = exog^i - \lambda \left(p_1^l + \frac{\gamma^l + \alpha_1 \left(q_1^l + q_1^c \right)}{\sqrt{f_1}} \right) \tag{D.5}$$

There are five equations to the demand system: equations D.1 to D.5 (i.e., equation D.3 for k=1 and equation D.4 for k=2).

4.2.5. First-order Conditions

The set of first-order conditions is:

$$p_k^l - \frac{\beta^l}{1 + \lambda \beta^i} q_k^l = \frac{\alpha_k \left(q_k^l + q_k^c \right)}{\sqrt{f_k}} + \cos t_k^l \tag{M.4 D.6,D.7}$$

$$p_k^c = \frac{\alpha_k \left(q_k^l + q_k^c \right)}{\sqrt{f_k}} + \cos t_k^c \tag{M.5 D.8,D.9}$$

$$\ln\left(\begin{array}{c}\left[\frac{\alpha_k\left(q_k^l+q_k^c\right)+\gamma^l}{2\sqrt{f_k^3}}\right]\left[1+\lambda\beta^i\right]q_k^l \\ +\left[\frac{\alpha_k\left(q_k^l+q_k^c\right)+\gamma^l}{2\sqrt{f_k^3}}\right]q_k^c\end{array}\right)=\ln\left(\cos t_k^f\right) \qquad \text{(M.6 D.10,D.11)}$$

There are three first-order conditions for a market with one airline: equations M.4 to M.6 with k =1. There are six first-order conditions for a market with two airlines: equations D.6, D.8 and D.10 for $k = 1$, equations D.7, D.9 and D.11 for $k = 2$.

4.3. Estimated Systems of Equations

Equations M.1 to M.6 determine optimal flight frequency and quantity choices for a single airline with flights on market B-D. These equations represent the estimated system of equations for sample markets with one airline. For notation purposes, the vector $y_j^m = \left(f_1, q_1^l, q_1^c, p_1^l, p_1^c, q^i\right)$ denotes the variables endogenous to the system for a sample market j with one airline (i.e., superscript m for monopoly in (nonstop) flights), x_j^m stands for the vector of exogenous variables, and the vector of residuals for the system of equations M.1 to M.6 is denoted by

$$u\left(y_j^m, x_j^m, N_j = 0, \theta\right) = (u_1, u_2, u_3, u_4, u_5, u_6)$$

where N_j is a dummy denoting the presence of a second airline with flight on market j and θ is the vector of parameters.

Equations D.1 to D.11 determine (Nash) equilibrium frequency and quantity choices for two airlines with flights on market B-D. Ticket prices are constructed from Databank 1A which is a 10% random sample of airline tickets and equation D.2, which denotes full price equality for local passengers on market B-D, does not hold, as such, for sample observations. This leads me to add an error term u_7 to equation D.2 which now reads

$$p_1^l + \frac{\gamma^l + \alpha\left(q_1^l + q_1^c\right)}{\sqrt{f_1}} = p_2^l + \frac{\gamma^l + \alpha\left(q_2^l + q_2^c\right)}{\sqrt{f_2}} + u_7 \qquad \text{(D.2*)}$$

Equations D.1, D.2*, D.3 to D.11 form the estimated system of equations for sample markets with two airlines. For notation purposes, the vector $y_j^d = \left(f_1, f_2, q_1^l, q_1^c, p_1^l, p_1^c, q_2^l, q_2^c, p_2^l, p_2^c, q^i\right)$ denotes the variables endogenous to the system for a sample market j with two airlines (i.e., superscript d for duopoly), x_j^d stands for the vector of exogenous variables, and the vector of residuals for the system is denoted by

$$u\left(y_j^d, x_j^d, N_j = 1, \theta\right) = (u_1, u_{21}, u_{22}, u_3, u_{41}, u_{42}, u_{51}, u_{52}, u_{61}, u_{62}, u_7)$$

Some additional comments... First, all residual terms are assumed to have mean zero. This is the only distributional assumption on the residual terms. Second, with the exception of stochastic delay parameters α_k and δ_{5k}, parameter values for the model do not vary with the number and identity of the airlines with flights on a market. Namely, I estimate only one of each demand and cost functions for the whole model. For the parameter α_k, I experimented with creating groups of airlines and assigning a parameter to each group. I found little variation in estimated parameter values except with regards to Delta and American Airlines (see discussion later in the results). The estimated model thus includes 2 different α_k parameters: one for Delta and American Airlines, one for all other airlines. Finally, the variable DUM_k in the full price specification for connecting passengers represents a single dummy variable for Delta Airlines[12].

4.4. Marketing Alliances

4.4.1. Data

The data for the analysis of alliances is built to represent an industry with alliances between each of Continental and Northwest, American and US Airways, Delta and United. These alliances change the network characteristics of these airlines. The specifications contain five variables denoting network characteristics: $ROUT_k$ and $RMILES_k$ in the connecting A-E schedule, $ICON$ and $IRATE$ in the function for connecting B-D passengers, and $TOTMIL_k$ in the marginal cost for a connecting A-E passenger.

[12] A dummy for American Airlines was initially added, then dropped (for parsimony) since its estimated parameter value was found not to differ significantly (at 10%) from zero.

For estimation purposes, these variables are constructed, for airline k, from airline k's entry decisions across markets other than B-D. Markets with flights from airline k are appended to one another (irrespective of flight times) given airline k's entry decisions across non B-D markets[13]. This yields a data of *potential* indirect routes per airline per market. When simulating an alliance between airlines k and h, I add to these data for airline k, those potential indirect routes constructed from <u>both</u> airline k and airline h's entry decisions across markets other than B-D.

In the third quarter of 1993, 95% of markets have four or fewer potential indirect routes per airline. The Databank 1A data, which contain tickets sold to passengers, indicate that an airline carried its connecting passengers for a market through four or fewer indirect routes 98.8% of the time. In an industry with alliances, 95% of markets now have twelve or fewer potential indirect routes per airline. Namely, alliances provide greater flexibility to an airline in its routing of connecting passengers for a market. They also provide a greater array of possible departure times to connecting passengers and increase the relative attractiveness of indirect B-D routes to a flight on market B-D.

To capture this increase in flexibility in an industry with alliance while maintining some bound on the number of potential indirect routes, I limit the number of indirect routes per airline per market to the eight indirect routes with shortest mileage[14]. This applies to the construction of the data for potential indirect routes for both the estimation and the alliance samples. Note that the identity of the eight shortest indirect routes for an airline on a market may differ across both samples. In an industry with alliances, shorter potential paths (for indirect routes) become available on some markets; this can decrease the number of connecting A-E markets for an airline on some markets.

In the sample data for alliances, the number of connecting A-E markets, $ROUT_k$, changes (with regards to the sample values used for the estimation) for 890 of the 975

[13] I do not consider any indirect routes for a market with mileage 3.5 times greater than the mileage of a flight, if there is one, or of the shortest indirect route, if there are no flights on the market.

[14] In an industry with alliances, 91% of sample markets have 8 or fewer indirect routes. Note that, without this limit, the number of potential indirect routes is very large number of potential indirect routes for some markets (up to 100). It is a fact that airlines do not route connecting passengers for a market through tens of itineraries.

airline observations in the sample (i.e., 605 markets with 1 airline, 185 markets with 2 airlines). The average increase in $ROUT_k$ is 64, while the average increase in $\ln(ROUT_k)$ is 0.87. In 13 of these 890 instances, the value of $ROUT_k$ decreases. Meanwhile, the number of indirect routes per sample market, $ICON$, increases, on average, by 5.97[15].

In summary, upon estimation of the model, I modify the data for the sample markets to reflect an airline environment with marketing alliances. This changes the sample values for the variables $ROUT_k$, $RMILES_k$, $ICON$, $IRATE$, and $TOTMIL_k$. This updated sample defines the *alliance sample data*.

4.4.2. Comparative Statics

When airline k enters into an alliance, the values of the $exog_k^c$, $\cos t_k^c$ and $exog^i$ terms in the system of equations determining its optimal/equilibrium decisions (either equations M.1-M.6 or D.1-D.11) change. These changes impact on airline k's flight frequency and quantity decisions. Some intuition may be provided.

Let airline k's demand from connecting A-E passengers increases as a result of the alliances. Since the marginal cost of a passenger is lower than that of a flight, an increase in the number of connecting passengers increases more than proportionately the airline's passenger volume relative to its flight frequency. This raises the cost of stochastic delay and, in turn, raises local ticket prices (equations M.4, D.6-D.7). Nevertheless, the marginal increase in the flight frequency may sufficiently decrease the overall cost of delay that full prices for local passengers fall. Hence, the number of local B-D passenges may rise in an industry with alliances.

Having said this, the price for local B-D passengers need not even rise. Alliances increase the number of indirect routes on a market B-D. This increases the number of B-D passengers which travel on indirect routes and puts downward pressure on the price for local B-D passengers.

The contribution of this paper is that it accounts for all these effects and endogenously

[15] Two identical indirect routes for market B-D which are code-shared across two airlines within an alliance count only as one indirect route.

estimates all relevant parameters to the model. The framework further provides the expected changes in frequency, quantity, and ticket prices for markets with alliances.

5. Estimation

5.1. GMM Estimation

The systems of equations M.1-M.6 and D.1-D.11 are estimated with a Generalized Method of Moments (GMM) method (re: Gallant (1986)). Instrumental variables for non-linear systems commonly include the variables exogenous to the model and lower order monomials thereof. I make no attempt to find the most efficient set of instruments and use for instruments (almost all of[16]) the variables exogenous to the model. For a listing, see table 1.

Table 1. Listing of Instrumental variables for the GMM estimation of the models

$\ln(POP)$	$AREA$	INC	$POPSQK$	$MILES$	
$\ln(ROUT_k)$	$RMILES_k$	$SLOT$	HUB_k	$MAJHUB$	
$CASM_k$	$FUEL_k$	$\ln(CSBH_k)$	$MAINT_k$	$ICON$	$IRATE$

Note: There are 16 instumental variables for the system for 1 airline and 23 for the system with 2 airlines

5.2. Sample Selection Bias

I have specified the (variable) profit-maximization model for a given number of airlines with flights on a market. Sample selection bias is a possibility; that is, the unobserved factors in the profit-maximization model may be correlated with unobserved factors determining the number of airlines on the market. If this correlation is not dealt with when estimating the model, estimated parameter values will be biased.

While a structural model of entry is desirable, it is not applicable to this analysis. Two econometric issues would have to be resolved. First, variable profits are a non-

[16]Variables $\ln(MILES)$ and $TOTMIL_k$ are not included in the lists of instrumental variables due to high correlation with the variable $MILES$. To provide a measure of the collinearity among instrumental variables, the ratio of the highest to the lowest eigenvalue to the instrumental vaiable matrix is 101 for markets with 1 airline and 263 for markets with 2 airlines.

trivial function of airline characteristics. Depending upon the identities of the entrants, it is conceivable that the same market could contain different number of firms. This non-uniqueness of entry equilibria is a major obstacle to any empirical estimation of entry[17]. Second, variable profits are a non-linear function of the residual terms (in each of systems M.1-M.6 and D.1-D.11). This makes it essentially impossible to relate the distributions of unobserved factors across the entry conditions and profit-maximization models. On a related data problem, the data for first-class passengers are known to be unreliable in the early 1990's. First-class passengers, as previously mentioned, are not dealt with in this paper and their revenues may account for a fair portion of market revenues depending upon the market and the time of the year.

5.3. Correcting for Sample Selection Bias

To deal with the possibility of a sample selection bias, I proceed to characterize the profit-maximization model as a *marginal* model. This means that, unconditionally on actual entry choices, the model describes the optimal choices for the decision variables (i.e., frequency, quantities) on a market with one or two airlines of given characteristics. For a market j, the model is then said to produce the notional output vector y_j^m (defined earlier) when there is one airline and y_j^d when there are two airlines. Note that, implicitly, any structural model of entry would, as I believe airlines do in practice, compare among such notional values.

Given this interpretation of the models, any possible sample selection bias may be corrected for by appropriately weighting the moment conditions used to estimate the profit-maximization systems. Let $y_j = \left(y_j^m, y_j^d\right)$, $x_j = \left(x_j^m, x_j^d\right)$, and let $z_j = \left(z_j^m, z_j^d\right)$ denote the instrumental variables, with z_j^m as the vector of instrumental variables for a market j with one airline and z_j^d that for a market j with two airlines. The theoretical moment conditions for market j can be written as:

[17]For example, Berry (1992) assumes that variable profits depend only on the number of airlines in the market, not their identity or characteristics. Bresnahan and Reiss (1990) allow for firm-specific unobservables in variable profits but limit their analysis to two potential entrants. As Berry (1992) points out (and deals with), there are a large number of potential entrants in the airline industry.

$$
\begin{aligned}
m\left(y_j, x_j, N_j, \theta^o\right) &= (1-N_j) \times \left(\frac{u\left(y_j^m, x_j^m, N_j=0, \theta^o\right)}{\Pr(N_j=0 \mid y_j, x_j)} \otimes z_j^m\right) \\
&\quad + N_j \times \left(\frac{u\left(y_j^d, x_j^d, N_j=1, \theta^o\right)}{\Pr(N_j=1 \mid y_j, x_j)} \otimes z_j^d\right)
\end{aligned}
$$

where θ^o is the vector of true parameter values for the model. These weighted moment conditions yield consistent estimated parameter values since

$$
\begin{aligned}
& E_{y,N}\left[m\left(y_j, x_j, N_j, \theta^o\right)\right] \\
&= E_{y,N}\left[(1-N_j) \times \left(\frac{u\left(y_j^m, x_j^m, N_j=0, \theta^o\right)}{\Pr(N_j=0 \mid y_j, x_j)} \otimes z_j^m\right) + N_j \times \left(\frac{u\left(y_j^d, x_j^d, N_j=1, \theta^o\right)}{\Pr(N_j=1 \mid y_j, x_j)} \otimes z_j^d\right)\right] \\
&= E_y\Big[\quad E_N\left[(1-N_j) \mid y_j, x_j\right] \times \left(\frac{u\left(y_j^m, x_j^m, N_j=0, \theta^o\right)}{\Pr(N_j=0 \mid y_j, x_j)} \otimes z_j^m\right) \\
&\quad + E_N\left[N_j \mid y_j, x_j\right] \times \left(\frac{u\left(y_j^d, x_j^d, N_j=1, \theta^o\right)}{\Pr(N_j=1 \mid y_j, x_j)} \otimes z_j^d\right) \quad \Big] \\
&= E_y\Big[\quad \Pr(N_j=0 \mid y_j, x_j) \times \left(\frac{u\left(y_j^m, x_j^m, N_j=0, \theta^o\right)}{\Pr(N_j=0 \mid y_j, x_j)} \otimes z_j^m\right) \\
&\quad + \Pr(N_j=1 \mid y_j, x_j) \times \left(\frac{u\left(y_j^d, x_j^d, N_j=1, \theta^o\right)}{\Pr(N_j=1 \mid y_j, x_j)} \otimes z_j^d\right)\Big] \\
&= E_y\left[u\left(y_j^m, x_j^m, N_j=0, \theta^o\right) \otimes z_j^m + u\left(y_j^d, x_j^d, N_j=1, \theta^o\right) \otimes z_j^d\right] \\
&= 0 \qquad \text{(by definition of the parameters of interest)}
\end{aligned}
$$

In practice, I only observe, for a market j, one of y_j^m or y_j^d. Hence, the GMM method, as I have outlined it, is not applicable without a *working* assumption. The model produces the frequency, quantities, and prices for an airline both when it is the only airline with flights on a market and when it faces a competitor with flights on a market.

I define an aggregated vector y_j^* of output variables for a market j such that

$$y_j^m = (f_1, q_1^l, q_1^c, p_1^l, p_1^c, q^i) \qquad \text{when there is one airline with flights}$$

$$y_j^d = \left(\frac{f_1+f_2}{2}, \frac{q_1^l+q_2^l}{2}, \frac{q_1^c+q_2^c}{2}, \frac{p_1^l\, q_1^l + p_2^l\, q_2^l}{q_1^l+q_2^l}, \frac{p_1^c\, q_1^c + p_2^c\, q_2^c}{q_1^c+q_2^c}, q^i\right)$$
$$\text{when there are two airlines with flights.}$$

With regards to the exogenous variables, I only observe the characteristics of the entrants on market j. For airlines not on market j, I can construct the network variables but not the cost variables[18]. This leads me to define an aggregated vector x_j^* of exogenous variables for market t. This vector includes:

[1] all variables for market B-D which do not depend upon the identity of the airline(s). These variables are $MILES$, $\ln(MILES)$, $\ln(POP)$, $AREA$, INC, $POPSQK$, $MAJHUB$, $SLOT$, $ICON$, $IRATE$);

[2] network characteristics for airlines with flights on market B-D and, for markets when there is only one airline, the network characteristics for the airline with highest value for $\ln(ROUT)$ which has no flights on market B-D (i.e., a most likely entrant). These variables are $\ln(ROUT_k), RMILES_k, HUB_k$;

[3] the cost characteristics of the airline on market B-D when it is the only one with flights, and the average cost characteristics across both airlines on market B-D in duopoly settings. These variables are $CASM$, $FUEL$, $MAINT$, $\ln(CSBH)$, and $TOTMIL$;[19]

[4] variables which may impact on entry on market B-D in the third quarter of 1993 but are not part of the model of profit-maximization. These variables include the

[18] I constructed cost variables for airlines not on market j from the aircraft assignment information on markets of similar mileage. These variables proved to have different means from the ones constructed for airline with flights on market j. This meant that, in the probit estimation I'll shortly characterize, these were highly significant variables. Given the arbitrary nature of the method used to construct these variables, I felt it was more appropriate to discard them.

[19] I could have included a $TOTMIL$ variable for a second airline (as with the other network characteristics in [2]); I did not do so for the specific estimation run I discuss in this paper.

average herfindal index of concentration in flight departures at airports B and D for the first quarter of 1993 (HFL), the number of airlines with hubs at airports B and D ($HUBAIR$), and the number of airports within each of the cities which airports B and D belong to ($AIRP$) (see appendix D for a complete listing of the variables in x^*).

For the weighted GMM method to yield consistent parameter values, I impose the *working* assumption that:

$$\Pr(N_j = 0 \mid y_j, x_j) = \Pr(N_j = 0 \mid y_j^*, x_j^*)$$

This is an implicit conditional independence assumption[20]: conditionally on the contents of y_j^* and x_j^*, nothing more is learned on the presence of a second airline with flights on market B-D from the remaining information contained in y_j and x_j.

The entry probabilities, $\Pr(N_j = n \mid y_j^*, x_j^*)$ for $n = 0, 1$, are estimated with a standard probit regression. Predicted probability values, $\Pr(N_j \widehat{= n \mid} y_j^*, x_j^*)$ for $n = 0, 1$, are used to weight the sample moment conditions:

$$\frac{1}{n}\sum_{j=1}^{n}\Bigg[\left(\frac{1-N_j}{\Pr(N_j \widehat{= 0 \mid} y_j^*, x_j^*)}\right) \times \left(u\left(y_j^m, x_j^m, N_j = 0, \theta^o\right) \otimes z_j^m\right) + \left(\frac{N_j}{\Pr(N_j \widehat{= 1 \mid} y_j^*, x_j^*)}\right) \times \left(u\left(y_j^d, x_j^d, N_j = 1, \theta^o\right) \otimes z_j^d\right) \Bigg]$$

The GMM estimation proceeds in two steps. Let $u_j^m(\theta) = u\left(y_j^m, x_j^m, N_j = 0, \theta\right)$, $u_j^d(\theta) = u\left(y_j^d, x_j^d, N_j = 1, \theta\right)$, $p_j^m = \Pr(N_j \widehat{= 0 \mid} y_j^*, x_j^*)$, and $p_j^d = \Pr(N_j \widehat{= 1 \mid} y_j^*, x_j^*)$.

[20] I feel reasonably comfortable with this assumption. I feel particularly comfortable with it when I consider the assumptions an alternate method of dealing with entry would have demanded.

The first-stage finds a preliminary estimate $\widehat{\theta}$ of θ^0:

$$\begin{aligned}
\widehat{\theta} &= \arg\min_{\theta} \quad S(\theta, V) \\
&= \arg\min_{\theta} \left(\sum_j \begin{matrix} u_j^m(\theta) \otimes z_j^m \\ u_j^d(\theta) \otimes z_j^d \end{matrix} \right)' V^{-1} \left(\sum_j \begin{matrix} u_j^m(\theta) \otimes z_j^m \\ u_j^d(\theta) \otimes z_j^d \end{matrix} \right) \\
& \textit{with} \\
& V = \begin{pmatrix} \frac{I_6}{\widetilde{p_j^m}} \otimes \sum_j z_j^m z_j^{m\prime} & 0 \\ 0 & \frac{I_{11}}{\widetilde{p_0^d}} \otimes \sum_j z_j^d z_j^{d\prime} \end{pmatrix} \\
& \textit{and} \qquad \widetilde{p_j^h} = \frac{1}{n^h} \sum_{j=1}^{n^h} \frac{1}{p_j^h p_j^h} \qquad\qquad h = m, d \quad \text{(markets with 1,2 airline(s))}
\end{aligned}$$

where I_n is a $n \times n$ identity matrix, n^m and n^d denote, respectively, the number of sample markets with one and two airlines. The variance-covariance matrix V is block-diagonal since the sample includes no markets where the number of airlines with flights changes during the sample period.

The preliminary estimate $\widehat{\theta}$ of θ^0 is consistent but not efficient. Greater efficiency, and some protection against heteroskedasticity, is achieved by updating the variance-covariance matrix on the basis of the first-stage residuals:

$$\hat{V} = \begin{pmatrix} \frac{1}{n^m} \sum_j \left(\frac{u_j^m(\widehat{\theta})}{p_j^m} \right) \left(\frac{u_j^m(\widehat{\theta})}{p_j^m} \right)' \otimes \sum_j z_j^m z_j^{m\prime} & 0 \\ 0 & \frac{1}{n^d} \sum_j \left(\frac{u_j^d(\widehat{\theta})}{p_j^d} \right) \left(\frac{u_j^d(\widehat{\theta})}{p_j^d} \right)' \otimes \sum_j z_j^d z_j^{d\prime} \end{pmatrix}$$

and obtaining a revised estimate $\theta^{\#}$ of θ^0 with $\theta^{\#} = \arg\min_{\theta} S\left(\theta, \hat{V}\right)$. Upon estimation of these sample moment conditions, I test for heteroskedasticity in the residual terms. The natural log of squared residual values are regressed on selected squared exogenous variables. Given the low R^2 values (all are inferior to 0.2), no additional correction is found necessary.

Prediction values are calculated, from estimation results, for the output vectors $y_j^m = \left(f_1, q_1^l, q_1^c, p_1^l, p_1^c, q^i\right)$ on markets with a single airline and $y_j^d = \left(f_1, f_2, q_1^l, q_1^c, p_1^l, p_1^c, q_2^l, q_2^c, p_2^l, p_2^c, q^i\right)$

on markets with two airlines. These predicted values are computed, by bootstrap, on the basis of a random draw of 100 residual terms for each of the two systems of equations. Correlation among observed and predicted values attest to the goodness of fit of the model.

For the analysis of alliances, I modify the estimation sample to create the aliance sample (as explained earlier). Given the estimated parameter values for the model, I simulate, by bootstrap and on the basis of the same residual draws as for earlier predictions, the flight frequency, quantities and prices for the airline markets in the *alliance sample data*. This produces updated values for output vectors y_j^m on markets with one airline and y_j^d on markets with two airlines in a setting with marketing alliances. A comparison of the prediction results across the estimation sample and the alliance sample is the basis for the discussion of the alliance results.

6. Results

6.1. Probit results for Entry

The probit regression correctly predicts 92% of sample markets (e.g., 573 of 605 markets with 1 airline and 151 of 185 markets with two airlines). Estimates, listed in table 3, show that the likelihood of a second airline with flights on a market increases with the size of the market (as measured by the variables $AREA$ and INC) and decreases with the level of airport concentration (HFL). A second airline is also more likely to enter on a market with one of its hub airports (variables HUB_2 and $SHORTHUB_2$). However, the larger is the connecting A-E demand for an airline on market (as measured by $\ln(ROUT_1)$), the lower is the likelihood that another airline enters that market.

The inclusion (in the probit) of the six aggregated output variables in vector y^* (e.g., quantities, etc.) allows me to test for the relevance of sample bias corrections. I test the null hypothesis that $\Pr(N = 1|x^*) = \Pr(N = 1|x^*, y^*)$; that is, entry is independent of y^*. Comparing the log-likelihood values for a probit regression including vector y^* to one ran without it, I find that $-2 \times$ (difference in log-likelihood values) $= 26.5 = \chi^2_6$.

This means that the null hypothesis is rejected at a 1% significance level. The output variables significantly contribute to the predictive power of the probit regression.

Having said this, no signs (e.g., positive or negative) or significance results for the estimated parameter values are reversed if I weight the sample moment conditions by either $\Pr(N = 1|x^*)$ or $\Pr(N = 1|x^*, y^*)$. Qualitatively, correcting for sample selection bias changes little to the interpretation of the results.

6.2. Estimation Results for the Profit-Maximization Model

6.2.1. Cost of Delay Valuation and Full Prices

The results deliver estimates of the cost of delay to airline passengers (see table 5). The cost of delay is the sum of the costs of stochastic and frequency delay. At sample mean values, the cost of stochastic delay to a passenger amounts to $9.30, while the cost of frequency delay is $86.87. The cost of delay to a local B-D passenger averages, thus, to $96.17 across sample markets in the third quarter of 1993.

The average ticket price for a local B-D passenger is $187.09. This means that the full price of a seat to a local passenger averages to $283.26. The cost of delay accounts for 33.95% of the amount of this full price.

The higher cost of frequency delay relative to stochastic delay is expected. Sample markets average 3.5 flights per day while customers' desired departure times are distributed on the 24-hr clock (primarily, from 6am to 10pm). Hence, the average time difference (frequency delay) between desired and actual departure times across passengers should be greater than the average stochastic delay encountered on aircraft/at airports on the day of travel.

The results (in table 5) also provide the marginal worth to a local passenger (flying coach) of an extra flight per day, per airline on a market. For example, if an airline increases its number of daily flights from 2 to 3 on a market, that additional flight is worth $18.52 to a local B-D passenger. If the increase is from 3 to 4 daily flights, that additional flight is worth $11.04. This type of results is novel to the literature (including the study of airline service in Morrison and Winston (1995)).

6.2.2. Hub airports

Hub airlines have significantly higher marginal costs per flight on markets with their hub airport(s) (re: parameter κ_3 in table 5). This is consistent with claims that hub operations lower aircraft utlization rates and inefficiently use airport and gate facilities[21]. The marginal cost of a flight is, however, relatively lower (re: parameter κ_4) at the major international hub airports (e.g., Atlanta, Dallas Forth-Worth, Houston IAH, Los Angeles LAX, Miami, New York JFK, and Chicago O'Hare). The sheer size of airline operations at these airports may provide some economies of scale and greater competition on fuel and maintenance costs.

In line with a finding of higher frequency costs, hub airlines are found to charge local passengers 1.48 cents more per mile on markets to/from their hub airports. On a market of 850 miles (sample average), this amounts to 6.8% of the $185 sample average ticket price for local passengers. Morrison and Winston (1995) cite a hub premium of 4 to 7% in 1993 (p.48). Ticket prices for local passengers are, further, estimated to be $10 higher on a market from/to one of the major hubs[22].

As for connecting ticket prices, these are estimated to be $23 lower on hub airlines on markets with their hub airport(s). This decrease may result from greater efficiency with regards to handling of the connecting passenger's flight transfer (e.g., greater efficiency in baggage handling, greater ability to deal with delays and rebookings). However, connecting ticket prices are $10 higher on markets with one of the major hub airports. These airports are major *nerve* centers for the industry where congestion probems through-

[21]It is worth noting that sample markets with a hub airport average 820 miles while markets with no hub airports average 1007 miles. Given the bulk of aircraft costs is incurred at take-off, a given aircraft is more expensive to fly on a shorter route. This may be argued to account for some of the higher costs per flight on markets with hub airports (beyond what the ln(MILES) variable already accounts for). For example, Morrison and Winston (1995, p. 46) have such an argument with regards to some hub studies. I did experiment with an interaction $HUB_k \times \ln(MILES)$ variable, instead of a hub dummy, but found it to yield a (marginally) poorer fit.

[22]This finding is a bit surprising in light of the lower relative frequency costs at these airports.

out the system tend to accumlate. Since congestion primarily impacts on connecting passengers, this could explain the relative increase in costs.

6.2.3. Goodness of fit: Estimated Parameter values

The full price specification requires that customers assign a positive (dollar) value to the cost of delay. The parameter values for the costs of frequency and stochastic delay, α_k, γ^l, γ^c are estimated to be positive and significantly different from zero (see table 4). Similarly, other estimated parameter values have signs consistent with the economic and institutional details of the model. The inverse local B-D demand and the connecting A-E schedule are downward sloping, while the reaction function for connecting B-D customers is upward sloping in the full price for a local B-D passenger. The local B-D demand is found to significantly increase with the population of the market (variable $\ln(POP)$) and the mileage of the route ($MILES$). The connecting A-E demand significantly increases with the number of potential connecting A-E markets ($\ln(ROUT)$) while it decreases as the ratio of the mileage of these potential indirect routes to the shortest mileage itinerary available for that market increases ($RMILES_k$). The number of connecting B-D passengers increases with the number of indirect B-D routes ($ICON$). The mileage of a market is also found to significantly increase airlines' marginal costs.

This is not to say that all parameter estimates necessarily meet expectations[23]. In particular, the point estimate for the slope of the indirect quantity, β^i, in the inverse local B-D demand is higher than the point estimate for the slope for the local B-D quantity, β^l. However, I fail to reject the null hypothesis that $\beta^i = \beta^l$ at a 5% significance level[24].

[23]For example, the variable IRATE is found to have a positive and significant impact on the number of connecting B-D passengers. The main issue here is that the value of IRATE rises with the number of indirect B-D routes (with the value of ICON) so that a higher value of IRATE measures, in some sense, an increase in the number of indirect routes on market B-D.

[24]The point estimate for β^i has consistently been higher than the point estimate for the slope for the local B-D quantity, β^l, across all trial runs with various functional forms and specifications. This is also the case if I run a standard OLS or two stage least-squares regression on the inverse demand function. In addition, I experimented with writing the connecting B-D quantity as a function of only the flight frequency on market B-D or of the quantity of local B-D passengers. In neither cases, did I get a point estimate for β^i smaller than that for β^l.

Connecting A-E passengers are also estimated to incur a cost of delay about three times as great as local passengers. While additional flights may significantly reduce the time connecting passengers may spend waiting at an airport, the magnitude of the difference in valuation is a bit surprising. It remains that while point estimates for frequency delay for local passengers have ranged from 1200 to 1400 across estimation runs, those for connecting passengers have varied from 3300 to 4900 depending upon the choice of instruments and specifications.

6.2.4. Goodness of fit : Test and Predictions

The criterion value for the GMM's objective function is, at optimum, equal to 125 in the first stage and 969 in the second stage[25]. Gallant (1986) indicates that this value is a Chi-square with $(6 \times 16) + (11 \times 23) - 44 = 305$ degrees of freedom. Clearly, the model's specification is statistically rejected. Two comments are warranted. First, as Gasmi, Laffont & Vuong (1992) emphasize, rejection is typical for large structural models. Second, method of moments estimation techniques are geared towards the estimation of sample mean values. This paper's endogenous variables (e.g., quantities, etc.) happen to exhibit, across markets, substantial heterogeneity around their sample mean values (descriptive statistics are in appendix A).

A better measure of fit for the model is the correlations between observed and predicted values for the output variables in each of the y^m and y^d vectors. The model predicts well the decision variables on markets with a single airline. Predicted mean values for the flight frequency, quantities and prices are close to their observed sample mean values. Correlations among predicted and observed values across sample markets range from 0.62 to 0.8 (see table 6)[26].

The model predicts poorly, however, the endogenous variables on markets with two

[25] Residuals were checked for outliers. For the estimation results presented in the paper, two percent of sample markets (17 out of 790) are not included in the estimation sample. Their deletion from this sample decreased the optimal value for the first-stage objective function by 30%.

[26] For comparison's sake, I ran least-square regressions for each demand and cost functions (with and without endogenous variables on the right-hand side). Adjusted R-square values range from 0.4 to 0.7 and their magnitude mirrors closely that of the correlation values for the structural model.

airlines. While correlation values range as well from 0.5 to 0.8, mean predicted values differ substantially from mean sample values (see table 6). There are some explanations. First, the only distributional assumptions on the residuals is that they have zero mean. While this is appealing from an estimation point of view, it does not facilitate the computation of predictions. Second, I estimate one set of demand and cost parameters across sample markets. Since there are about 3.5 times fewer sample markets with two airlines than sample markets with one airline, one expects the model to fit better that data on markets with one airline. There is also as much heterogeneity in the sample values of the output variables across markets with two airlines as there is across markets with one airline. Third, full price equality does not hold (as such) at sample values across markets with two airlines. The resolution algorithm for predictions has to adjust quantities and prices as to yield full price equality. This can yield extreme predicted values for the quantities of local passengers.

6.3. Results for Marketing Alliances

6.3.1. Simulation Results

This discussion compares amongst predicted output values for the estimation sample and the alliance sample. Given the model's relatively low predictive power for markets with two airlines, I base my analysis on simulated output values for the 556 sample markets with a single airline where this airline belongs to an alliance[27].

Alliances decrease the number of local B-D passengers on 82% of sample markets. The median change is -32%. While the number of connecting B-D passengers rises on 98% of sample markets (median change, Δ, is +47%), this increase is not sufficient to palliate for the decrease in the number of local B-D passengers. The total passenger volume on a market B-D decreases on 80% of sample markets with a median change of -15%.

[27] 49 of the 605 markets have flights from an airline not in one of the three alliances. These arlines are America West, TWA, Midwest Express, Kiwi, and Braniff.

The number of connecting A-E passengers for an airline with flights on a market B-D increases across 77% of markets (median Δ is 14.6%), while the ticket price for connecting A-E passengers is lower on 69% of markets and falls, on average, by 0.07%.

The total passenger volume for an airine with flights on market B-D (i.e.,both local B-D and connecting A-E passengers) decreases on 64% of sample markets (median Δ is -9.7%) and the airline's flight frequency on a market B-D falls, on average, by 1.5%. Consequently, stochastic delay decreases (median Δ is -5%), while frequency delay increases (median Δ is 4.5%), across 63% of sample markets . The cost of delay is higher on 74% of sample routes.

The end result is that, on the average, the ticket price for local B-D passengers falls by 2% while the full prices for these passengers only decreases by 0.07%. Ticket prices and full prices are lower across 80% of markets.

Two separate can be identified. First, the increase in the number of indirect routes on a market B-D (i.e., increase in $exog^i$) increases the level of competition for B-D passengers. This puts downward pressure on the ticket price for local B-D passengers. In response, airlines shift away from local B-D passengers towards connecting A-E passengers. Second, the exogenous increase in the demand from connecting A-E passengers (i.e., increase in $exog^c$) increases the number of connecting passengers on a market B-D. This increase does not, however, lead to an increase in the number of flights an airline offers on market B-D. Rather, airlines compensate for the increase in connecting passengers by decreasing the number of local passengersthey serve.

These two effects (i.e., increase in $exog^i$ and in $exog^c$) both work to shift airlines away from flying local passengers. The question remains as to which effect might dominate. Hence, I ask: holding the number of indirect routes constant, how does the increase in the demand from connecting A-E passengers change flight frequency and quantity decisions?

To provide an answer, I modify slightly the composition of the alliance sample. If I update, as earlier, the values for variables $\ln(ROUT_k)$, $RMILES_k$, and $TOTMIL_k$, I do not update anymore the values for $ICON$ and $IRATE$. Namely, the value f the

$exog^i$ in this modified sample is identical to its value in the estimation sample.

In this experiment, the local B-D passenger volume decreases on 73% of markets and the median change is -27%. The number of connecting B-D passengers increases slightly (median Δ is 3.3%) across 77% of markets. The total passenger volume for an airline with flights on market B-D on 62% of markets (median Δ is -8%) while the flight frequency changes little across markets (median Δ is -1%).

Hence, airlines compensate for an increase in the number of connecting A-E passengers by decreasing the number of local passengers. This makes the alliance experience a different object from the hub-and-spoke one. The hub structure of the industry has increased the availability of indirect routes for connecting passengers. This has been argued to have increased ticket prices for local passengers. However, as Bamberger & Carlton (1993) document, on markets to hub airports, flight frequencies rose sufficently that the total number of local passengers on these markets has actually increased. The analysis in this paper suggests that the benefits from increased flight frequency do not materialize with regards to alliances. In fact, this paper finds little, if anything, to recommend about alliances between major US airlines (from a discussion based strictly on the code-sharing aspect of alliances).

6.3.2. Future work

The marketing simulations do point to directions for future work. Allainces raise the number of connecting A-E passengers an airline serves. I then find that airlines *overly* compensate for the increase in connecting passengers by substantially decreasing the number of local passengers they serve. Some of the magnitude of this effect can, no doubt, be attributed to the estimation results.

First, the point estimates for the slopes in the inverse market demand for local B-D passengers are low, thereby mitigating some of the impact of quantity changes on full prices and ticket prices. Second, the high frequency delay valuation for connecting A-E passengers, relative to local B-D passengers, certainly contributes to this over-reaction.

Having said this, these results have shown much robustness to trial estimation and prediction runs over time. The point is that, to my belief, they have shown sufficient robustness that I feel comfortable with the paper's main findings on alliances. It remains that more work is needed to understand the nature of the externalities among local and connecting passengers, to model demand systems for local passengers (Berry, Carnall & Spiller (1995) is a step in that direction), and to characterize the deciion problem of airlines with regards to connecting passengers.

7. Tables

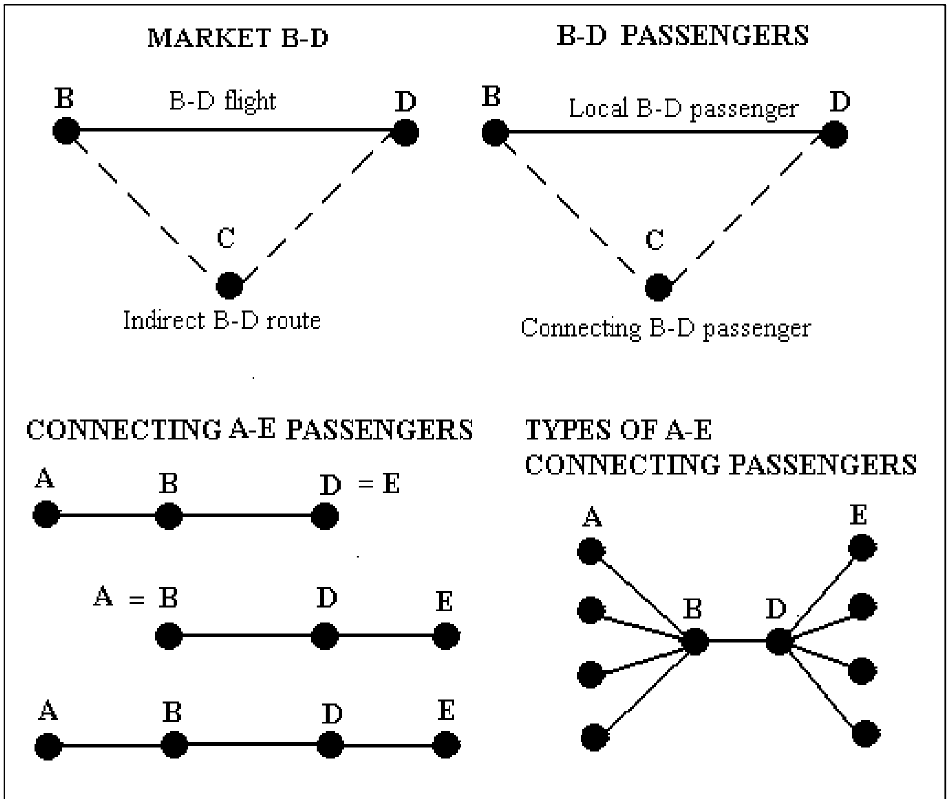

Figure 1.

Table 1. Two-letter codes for U.S. airlines and number of sample markets (with 1, 2 airline(s) with flights) with each airline.

Code	AA	CO	DL	NW	UA	US
Airline	American	Continental	Delta	Northwest	United	US Air(ways
# of markets with						
1 airline	71	45	96	107	49	188
2 airlines	93	44	72	18	72	39

Note: The sample also includes markets with flights by these airlines: Midwest Express, Kiwi, and America West Airlines.

Table 3. Probit results for entry phase. Mean parameter values and associated t-statistics. Dependent variable is dummy denoting whether airline 2 (as compared to airline 1) has flights on the market.

	Mean	T-stat.		Mean	T-stat.
Variables in x*:					
CONSTANT	-12.120	-1.43	SHHUB1	2.542	0.59
MILES	-0.351	-0.29	SHHUB2	17.363	2.79
ln(MILES)	29.727	2.81	CASM	8.368	2.38
ln(POP)	-0.786	-1.43	LNCSBH	-51.006	-2.18
AREA	0.369	2.14	MAINT	-0.957	-1.08
INCF	2.108	4.12	FUEL	0.357	0.43
POPSQK	0.153	0.36	PIMIL	0.035	0.29
MAJHUB	0.433	1.76	ICON	1.082	3.80
HUBAIR	1.265	0.41	IMILES	-2.089	-2.03
SLOT	13.320	0.42	IRATE	1.908	2.39
HFL	-2.958	-3.84			
AIRP	-0.038	-0.31			
ln(ROUT1)	-12.433	-3.71	***Variables in y*:***		
ln(ROUT2)	0.678	0.26			
RMILES1	-0.118	-0.15	CPRICE	0.01028	2.57
RMILES2	-0.590	-1.04	PRICE	-0.00799	-2.45
ln(ROUTD1)	-2.130	-0.82	INDIR	-0.00007	-1.28
ln(ROUTD2)	14.558	5.76	HDIR	-0.00010	-3.24
HUB1	1.901	3.68	HCON	0.00002	1.08
HUB2	1.022	2.51	HFREQ	0.00210	1.07

-2 log Likelihood = -287.5
Maddala's pseudo R-square: 0.515
Mc Fadden's pseudi R-square:0.666

Note: The integers 1 and 2 following the name of a variable denote the airline to which the variable applies. This probit regression examines the likelihood of entry of airline 2.

Table 4. Mean (in $) cost of delay, frequency delay and full price valuation to a local B-D customer.

Mean cost of delay and full price valuation (in $)

Sample market with:	1 airline	2 airlines:	1st airline	2nd airline
Ticket price	183.15		196.64	190.44
Cost of stochastic delay	9.10		9.20	10.03
Cost of frequency dela	90.55		78.20	83.51
Full price	282.80		284.04	283.98

Cost of frequency delay valuation per daily flight (in $):

# of flights :	1	2	3	4	5	6
Cost:	142.75	100.94	82.42	71.38	63.84	58.28
Change =		-41.81	-18.52	-11.04	-7.54	-5.56

Change = worth (in $) of an extra daily flight to a local B-D customer

Note: The mean dollar valuations are the sample means for markets with 1 and 2 airlines with flight(s). Sample means are computed from the mean values of estimated parameters for the cost of delay function.

Table 5. Mean and standard deviaton (S.d.) of estimated parameter values for variables in profit-maximization model.

Stochastic delay parameters / **Frequency delay parameters**

Param.	α_1	α_2	γ^l	γ^c
Var.	***AA, DL***	***Others***	***local***	***connecting***
Mean	0.0053	0.0083	1354	4410
S.d.	0.0014*	0.0014*	99*	338*

Estimates for the inverse local B-D demand

Param.	β_0	β_1	β_2	β_3	β_4	β^l	β^i	β_5^m
Var.	***constant***	***ln(POP)***	***POPSQK***	***INC***	***MILES***	***slope***	***slope***	***AA***
Mean	1.7810	0.1358	-0.0664	0.0660	0.6621	0.00056	0.00285	0.1866
S.d.	0.1594*	0.0357*	0.0294*	0.0452	0.0391*	0.00014*	0.00091*	0.0859*

Estimates for the connecting A-E demand schedule

Param.	δ_0	δ_1	δ_2	δ_3	δ_4	δ	δ_5
Var.	***constant***	***ln(ROUT)***	***RMILES***	***HUB***	***ln(MILES)***	***slope***	***DL***
Mean	8.5592	3.5360	-0.5586	0.7131	5.6156	0.0086	0.6962
S.d.	0.2763*	0.4101*	0.0852*	0.0614*	0.4125*	0.00062*	0.0601*

Estimates for the connecting B-D reaction function

Param.	λ_0	λ_1	λ_2	λ_3	λ_4	λ
Var.	***constant***	***AREA***	***ICON***	***ln(MILES)***	***IRATE***	***slope***
Mean	-0.9419	0.0235	0.1453	0.0761	0.0517	25.42
S.d.	0.1614*	0.0050*	0.0121*	0.2997	0.0256*	3.88*

Estimates for the marginal cost per flight

Param.	κ_0	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6	κ_7
Var.	***constant***	***FUEL***	***CASM***	***HUB***	***MAJHUB***	***SLOT***	***ln(MILES)***	***DL***
Mean	7.2782	0.0578	-0.5081	0.4112	-0.1916	5.5586	3.2931	0.4968
S.d.	0.1895*	0.0421	0.0996*	0.0424*	0.0211*	2.5066*	0.2238*	0.0443*

Estimates for the marginal cost per local passenger

Param.	ω_0	ω_1	ω_2	ω_3	ω_4	ω_5
Var.	***constant***	***ln(CSBH)***	***MAINT***	***MAJHUB***	***MILES***	***HUBxMILES***
Mean	0.6962	2.3931	-0.1422	0.1007	0.3377	0.1477
S.d.	0.2133*	0.7660*	0.0432*	0.0196*	0.0317*	0.0253*

Estimates for the marginal cost per connecting passenger

Param.	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4
Var.	***constant***	***TOTMIL***	***ln(CSBH)***	***MAJHUB***	***HUB***
Mean	0.4124	0.0494	4.3829	0.1156	-0.2313
S.d.	0.2299	0.0028*	0.7735*	0.0210*	0.0303*

Note: Parameter values in each of the specifications for the inverse local B-D demand, the marginal cost per local passenger, and the marginal cost per connecting passenger have been divided by 100.

Note: In S.d row, * denotes significance at a 5% level.

REFERENCES

BAMBERGER, G. AND CARLTON, D. (1993). “Airline Networks and Fares,” *Mimeo, University of Chicago.*

BERECHMAN, J. AND SHY, O. (1993). “Airline Deregulation and the Choice Networks.” *Mimeo, Tel-Aviv University.*

BERRY, S. (1990). “Airport Presence as Product Differentiation.” *American Economic Review, Papers and Proceedings*, Vol. 80, pp. 394-399.

________ . (1992). “Estimation of a Model of Entry in the Airline Industry.” *Econometrica*, Vol. 60, pp. 889-917.

BERRY, S., CARNALL, M., AND SPILLER, P. (1996). “Airline Hubs: Costs, N and the Implications of Customer Heterogeneity.” *NBER. Working Paper*.

BORENSTEIN, S. (1989). “Hubs and High fares: Dominance and Market Powe U.S. Airline Industry.” *RAND Journal of Economics*, Vol. 20, pp. 344-365.

______________ . (1991). “The Dominant-Firm Advantage in Multi-Product In Evidence from the U.S. Airline Industry.” *Quarterly Journal of Economics*, Vol. 10, pp. 1237-1266.

______________ . (1992). “The Evolution of the U.S. Airline Industry.” *Journal of Economic Perspectives*, Vol. 6, pp. 45-73.

BRANDER, J. AND ZHANG, A. (1990). “Market Conduct in the Airline Indus Empirical Investigation.” *RAND Journal of Economics*, Vol. 21, pp. 567-583.

BRESNAHAN, T. (1989). “Empirical Studies of Industries with Market Power. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization*. New-York: North Holland.

BRESNAHAN, T. AND REISS, P. (1990). "Entry in Monopoly Markets." *Review of Economic Studies*, Vol. 57, pp. 531-553.

BRUECKNER, J. AND SPILLER, P. (1991). "Competition and Mergers in Airl Networks." *International Journal of Industrial Economics*, Vol. 9, pp. 323-342.

_____________________________ . (1994). "Economies of Traffic Density in Deregulated Airline Industry." *Journal of Law and Economics*, Vol. 37, pp. 379-415.

BRUECKNER, J., DYER, N., AND SPILLER, P. (1992). "Fare Determination Hub-and-Spoke Networks." *RAND Journal of Economics*, Vol. 23 , pp. 309-333.

CAVES, D., CHRISTENSEN, L., AND THRETEWAY, M. (1984). "Economie Density versus Economies of Scale: Why Trunk and Local Service Airline Costs *RAND Journal of Economics*, Vol. 15, pp. 471-489.

DORMAN, G. (1976). "Airline Competition: A Theoretical and Empirical Anal *Ph.D. Dissertation, University of California at Berkeley.*

DOUGLAS, G. AND MILLER, J.C., III. (1972). *Economic Regulation of Domestic A Transport: Theory and Policy*. Washington, D.C.: Brookings Institution.

GALLANT, A. (1986). *Nonlinear Statistical Models*. New York: John Wiley & Sons.

GASMI, F., LAFFONT, J-J., AND VUONG, Q. (1992). "Econometric Analysis Collusive Behavior in a Soft-Drink Market." *Journal of Economics and Management Strategy*, Vol. 1, pp. 277-311.

KATZ, M. AND SHAPIRO, C. (1985). "Network Externalities, Competition, ar Compatibility." *American Economic Review*, Vol. 75, pp. 424-440.

LEDERER, P. (1993). "A Competitive Network Design Problem with Pricing." *Transportation Science*. Vol. 27, pp. 25-38.

MORRISON, S. AND WINSTON, C. (1995). "*The Evolution of the Airline Industry*." Washington, D.C.: Brookings Institution.

PANZAR, J. (1979). "Equilibrium and Welfare in Unregulated Airline Markets *American Economic Review, Papers and Proceedings*. Vol. 69, pp. 92-95.

REISS P. AND SPILLER, P. (1989). "Competition and Entry in Small Airline Markets." *Journal of Law and Economics*. Vol. 32, pp. S179-s102.

RICHARD O. (1997). "An Empirical Analysis of Pricing and Flight Frequency Airline Routes." *Bradley Policy Research Center*. University of Rochester.

RICHARD O. (1999). "An Empirical Analysis of Airline Flight Schedules." *Bradley Policy Research Center*. University of Rochester.

TIROLE, J. (1990). *The Theory of Industrial Organization*. Cambridge: MIT Press.

8. Appendix A: Descriptive Sample Statistics

Table A1. Descriptive sample statistics for the (605) markets with one airlines with flights.

Variable name	Mean	Std. Dev.	Min.	Max.	Scaling factor
Local B-D price 1	183	43	61	330	
Connecting B-D price 1	206	38	102	457	
Local B-D quantity 1	7853	7064	345	58356	
Connecting A-E quantity 1	13756	12227	258	106064	
Flight frequency 1	284	164	77	1090	
Connecting B-D quantity	1896	2255	40	18910	
MILES	0.841	0.533	0.076	2.724	/1000
ln(POP)	3.931	0.542	2.092	5.264	
AREA	3.539	1.495	0.732	8.834	/1000
INC	1.642	0.279	0.980	2.708	/1000
POPSQK	1.306	0.336	0.707	3.716	/1000
CASM 1	0.478	0.117	0.247	1.034	/10
FUEL 1	0.899	0.219	0.617	2.339	/1000
MAINT 1	0.446	0.216	0.129	2.229	/10
ln(CSBH 1)	0.273	0.016	0.224	0.354	/10
MAJHUB	0.266	0.450	0.000	2.000	
HUB 1	0.884	0.320	0.000	1.000	
HUB 2	0.167	0.373	0.000	1.000	
SLOT	0.001	0.004	0.000	0.010	/100
HFL	0.834	0.225	0.287	1.606	
AIRP	2.689	1.121	2.000	7.000	
HUBAIR	0.117	0.060	0.000	3.000	/10
SHORTHUB 1	0.028	0.045	0.000	0.100	/10
SHORTHUB 2	0.000	0.006	0.000	0.100	/10
ln(ROUTD 1)	0.235	0.089	0.000	0.470	/10
ln(ROUTD 2)	0.152	0.086	0.000	0.364	/10
IMILES	1.208	0.581	0.201	3.076	/1000
Estimation sample:					
ln(ROUT 1)	0.420	0.045	0.139	0.512	/10
ln(ROUT 2)	0.309	0.072	0.000	0.481	/10
RMILES 1	1.383	0.142	1.025	1.932	
RMILES 2	1.415	0.201	1.008	2.238	
TOTMIL 1	10.970	4.338	3.633	24.139	/100
ICON	1.295	0.643	0.100	3.400	/10
IRATE	1.591	0.374	1.040	2.992	
Alliance sample:					
ln(ROUT 1)	0.423	0.044	0.139	0.505	/10
RMILES 1	1.373	0.138	1.029	1.951	
TOTMIL 1	10.976	4.358	3.459	24.607	/100
ICON	1.304	0.582	0.100	2.800	/10
IRATE	1.566	0.361	1.040	2.992	
Predicted value for Pr(N = 0 | x*, y*)	0.925	0.164	0.079	1.000	

The numbers 1 and 2 following the name of a variable designate airlines.
Airline 1 is the airline with flights on market B-D
Airline 2 denotes the airline with highest value for ln(ROUT), which has no flights on market B-D (re: probit estimation of entry).

Table A2. Descriptive sample statistics for the (185) markets with two airlines with flights.

Variable	Mean	Std. Dev.	Min.	Max.	Scaling factor
Local B-D price 1	197	60	65	370	
Local B-D price 2	190	55	67	356	
Connecting B-D price 1	217	46	110	370	
Connecting B-D price 2	225	46	121	383	
Local B-D quantity 1	12057	8911	1083	49347	
Local B-D quantity 2	10233	9318	665	61999	
Connecting A-E quantity 1	19208	15164	360	68823	
Connecting A-E quantity 2	15507	14495	227	102634	
Flight frequency 1	400	227	86	1052	
Flight frequency 2	330	182	89	1302	
Connecting B-D quantity	2578	2602	120	16300	
MILES	0.954	0.548	0.067	2.704	/1000
ln(POP)	4.369	0.458	2.673	5.298	
AREA	4.763	1.508	1.320	9.567	/1000
INC	1.782	0.259	1.252	2.595	/1000
POPSQK	1.371	0.317	0.864	2.573	/1000
CASM 1	0.422	0.097	0.247	0.737	/10
CASM 2	0.478	0.100	0.247	0.900	/10
FUEL 1	0.966	0.238	0.617	2.185	/1000
FUEL 2	0.940	0.281	0.677	2.668	/1000
MAINT 1	0.434	0.196	0.129	1.508	/10
MAINT 2	0.478	0.288	0.129	2.271	/10
ln(CSBH 1)	0.264	0.013	0.231	0.303	/10
ln(CSBH 2)	0.275	0.015	0.231	0.354	/10
MAJHUB	0.660	0.559	0.000	2.000	
HUB 1	0.914	0.282	0.000	1.000	
HUB 2	0.843	0.365	0.000	1.000	
SLOT	0.003	0.005	0.000	0.020	/100
HFL	0.777	0.237	0.309	1.436	
AIRP	2.951	1.190	2.000	6.000	
HUBAIR	0.190	0.064	0.000	3.000	/10
SHORTHUB 1	0.020	0.040	0.000	0.100	/10
SHORTHUB 2	0.017	0.037	0.000	0.100	/10
ln(ROUTD 1)	0.273	0.078	0.000	0.428	/10
ln(ROUTD 2)	0.273	0.075	0.069	0.449	/10
IMILES	1.437	0.584	0.221	3.105	/1000
Estimation sample:					
ln(ROUT 1)	0.411	0.054	0.139	0.498	/10
ln(ROUT 2)	0.414	0.054	0.161	0.504	/10
RMILES 1	1.409	0.159	1.085	2.099	
RMILES 2	1.416	0.166	1.117	2.056	
TOTMIL 1	12.375	4.274	5.317	24.615	/100
TOTMIL 2	12.022	4.527	4.522	25.345	/100
ICON	1.711	0.779	0.100	3.600	/10
IRATE	1.683	0.428	1.058	3.299	
Alliance sample:					
ln(ROUT 1)	0.416	0.052	0.139	0.502	/10
ln(ROUT 2)	0.417	0.054	0.161	0.502	/10
RMILES 1	1.376	0.150	1.081	1.860	
RMILES 2	1.392	0.157	1.068	1.881	
TOTMIL 1	12.383	4.330	5.304	24.654	/100
TOTMIL 2	12.087	4.538	5.295	26.169	/100
ICON	1.716	0.623	0.100	2.700	/10
IRATE	1.607	0.424	1.033	3.299	
Predicted value for Pr(N = 1 | x*, y*)	0.757	0.280	0.003	1.000	

The numbers 1 and 2 following the name of a variable designate airlines.

$$POPUL^{*} = \frac{\sqrt{\text{product of population at cities B and D}}}{\sqrt{100,000}}$$

$$INC = \frac{(\text{sum of annual income levels at cities B and D})^2}{100,000}$$

$$POPSQK = \frac{(\text{ sum of population at cities B and D})}{1000 \times AREA}$$

Note: * Maximal population figures (for the very large cities) were capped at 8,000,000.

9. Appendix B: The Data

9.1. Local vs. Connecting Tickets and Passengers

Databank 1A is a 10% random sample of all tickets sold each yearly quarter to passengers in the USA. Fare, fare class, airline, and itinerary flown are included but no time-of-day or day-of-week information is provided. There may be several passengers listed on a given ticket. The data are from the third quarter of 1993.

The ticket information in Databank 1A is reported on a flight segment basis. Flight segments are associated with flight numbers and may consist in consecutive local flights with same flight numbers. In this latter case, no information is provided on the intermediate city. I associate a flight segment from airport B to airport D with a local flight if the airline issuing the ticket flies local from B to D. If the airline does not, the ticket information is disregarded (counting the passengers on these tickets as connecting/indirect passengers does not modify the qualitative results of the paper).

Databank 1A does not classify tickets as one-way or roundtrip and only the full-itinerary fare is reported. A two-segment ticket is counted as a roundtrip if the origin city lies within 50 miles of the destination city. Roundtrip tickets are split into two (directional) one-way tickets, and the full-itinerary fare is divided by two to yield the one-way fares. All other two-segment tickets are one-way tickets. If a one-way ticket between airports B and D contains only one flight segment, it is said to represent a local B-D passenger. If a one-way ticket between airports B and D contains two flight segments, there are two possibilities. Let C be the intermediate city on this one-way ticket. If both flight segments are with the same airline, the ticket is said to represent an connecting B-D passenger, and a connecting passenger on each of market B-C and C-D. If the flight segments are on different airlines (i.e., the passenger switched airlines at airport C), the ticket is said to represent a local B-C and a local C-D passenger. The one-way fare is, in this case, divided on the basis of the relative mileage of each flight segment. These steps are repeated for tickets with up to 4 flight segments. One-way tickets with 4 flight segments on the same airline are deleted. One-way tickets between two airports B and D with total mileage 3.5 greater than the mileage of the shortest available itinerary between B and D are deleted.

For each airline on market B-D, I have tickets for three categories of passengers: local B-D, connecting A-E, and connecting B-D passengers. These tickets are split across two fare classes: coach and first/business.

9.2. Passenger Volumes

The passenger information is aggregated across tickets for market B-D to yield the number of: local B-D coach passengers (DIR), local B-D business passengers ($DIRF$), connecting A-E coach passengers (CON), connecting A-E business passengers ($CONF$), connecting B-D coach passengers (IND), and connecting B-D business passengers ($DIRF$). These numbers, computed from Databank 1A, represent a 10% random sample of all passengers per airline on market B-D in the second quarter of 1993. Databank DS T-100 provides the total number of passengers per airline on market B-D (TPASS) for the second quarter of 1993. I reconcile the data across both databases. For the purpose of this paper, the number of local B-D coach passengers is equal to $TDIR = \left(\frac{DIR}{DIR+CON+DIRF+CONF}\right) \times TPASS$. The number of connecting A-E coach passengers is equal to $TCON = \left(\frac{CON}{DIR+CON+DIRF+CONF}\right) \times TPASS$. The number of connecting B-D passengers is equal to $TIND = 10 * IND$ (i.e., Databank 1A is a 10% random sample of ticket prices) for lack of a better measure.

9.3. Fare Data

Databank 1A is known to contain some excessively high and low fares. I only consider tickets with a fare of at least \$10 and at most \$1,400 per flight. For an itinerary with total mileage M, the ticket price on a per mile basis has to lie within $\left(0.01 + \frac{10}{M}\right)$and $\left(0.75 + \frac{500}{M}\right)$. These price-per-mile cutoff values eliminate prohibitive fares on short-mileage markets. Local and connecting fares are averaged across, respectively, Local and connecting tickets. Since a ticket may list several passengers, the fares are weighted by passengers when averaged.

9.4. Construction of Potential Connecting A-E Itineraries

These data are from the *first* quarter of 1993. Potential connecting A-E itineraries for an airline k on market B-D are created by appending the markets with flights from airline k to one another. Markets are appended once, creating indirect routes with one intermediate stop. The values for variables $\ln(ROUT_k)$, $\ln(ROUTD_k)$, $RMILES_k$, ICON, IMILES, TOTMIL_k, and CONMIL_k are created from these indirect routes.

To ensure consistency with the way I created the passenger volumes for connecting passengers from Databank 1A, I only keep those potential A-E itineraries with total mileage less than 3.5 times the mileage of the shortest possible (nonstop or indirect) itinerary.

9.5. The Sample and Estimation Details

The sample data, used for estimation purposes, contain 605 markets with one airline and 185 markets with two airlines.

The sample does not include: [1] markets with flights listed in the Official Airline Guide North American editions for July 1993, but not included in Databank DS T-100; [2] markets with no reported price or passenger data; [3] markets where an airline has fewer than 60 flights over the quarter; and [4] markets with no cost data. The cost data are obtained from the first quarter of 1993 Aircraft Operating Costs and Statistics (AQOCS) periodical. This is the last quarter for which these data are available since AVMARK Inc. stopped its publication at that time. I construct this cost data by reconciling the aircraft equipment information in Databank T-100 with the AQOCS information. The implication of [4] is that I only get cost data if an airline was one the market in the first quarter of 1993. This eliminates an additional 15 sample markets from the data.

The sample does not also include markets with Southwest and Alaska Airlines. The data for Southwest Airlines is reported to Databank 1A in a fashion which does not allow me to distinguish between local and connecting passengers. Alaska Airlines's network structure is fundamentally different from that of other US airlines.

9.6. List of the variables in x^* (continued from text)

There are three variables in x^* which are not discussed in the text: $IMILES$, $SHORTHUB_k$, and $\ln(ROUTD_k)$.

The variable $IMILES$ denotes the average mileage of indirect routes on market B-D (variable $IRATE = IMILES/MILES$). $SHORTHUB_k$ is a dummy variable denoting the presence of a hub airport for airline k on a market of 500 miles or less. In its 1/10/94 issue, Airline Business magazine reports that the hub-and-spoke structure increases costs sharply on short haul routes (mileage ≤ 500) "because it underutilizes both aircraft and labor." $\ln(ROUTD_k)$ is the (natural log of) the number of potential connecting A-E markets airline k may draw passengers from. Both $SHORTHUB_k$ and $\ln(ROUTD_k)$ are constructed for two airlines as for the other network characteristics in x^* (see part [2] in discussion of x^* in the text).

All three variables were initially part of the specifications for the profit-maximization models. However, I could not reject the null hypotesis that their parameter values were equal to zero at a 10% significance level. The variables were dropped from the specifications for parsimony in the number of estimated parameter values.