An Empirical Model of Entry across Airline Routes with Incomplete Information and Demand Synergies.

Olivier Armantier* and Oliver Richard†

January 2000

Preliminary; please do not quote

Abstract

We propose a model of simultaneous entry decisions for N symmetric firms across M markets with demand synergies and incomplete information on marginal costs of production.

We develop an algorithm, based upon Monte Carlo simulations, to determine numerically the Nash Equilibrium. We also apply the inference method proposed by Florens, Protopopescu and Richard (1997) to estimate the distribution of costs.

Results provide threshold cost values for entry decisions, probabilities of entry, expected quantities and prices as well as estimated costs for each of the M markets.

Keywords: Entry, Incomplete Information, Structural Estimation, Network, Airline Industry

JEL Classifications : L11, D82, C15, C51, L93, R41

* Dept of Economics, S621 SBS Building, SUNY Stony Brook, Stony Brook NY, 11794. E-mail: olivier.armantier@sunysb.edu.

† University of Rochester, Rochester, NY 14627; E-mail: richard@ssb.rochester.edu

1. Introduction

The few existing competitive models of entry for multimarket industries (i.e., Berry (1992), Gelfand and Spiller (1987), Reiss and Spiller (1989)) make (at least one of) two fundamental hypotheses. First, they consider settings of perfect and complete information. Second, they consider entry decisions on a partial equilibrium basis. Namely, even if they *happen* to account for externalities across markets, they model the entry decisions of firms on a market conditional to the firms' entry decisions across the other markets in the industry (unless they happen to consider a simple two-market setting). Both these hypotheses are empirically questionable. Firms rarely know, or observe accurately, some of their rivals' costs or demand parameters. Besides, in multimarket settings, entry or production decisions on a single market typically affect the state of the other markets within the industry. Namely, these settings require models of simultaneous, global decision-making across markets. It is at these two levels (i.e., incomplete information and global decision-making) that lies the contribution of our article to the empirical literature on entry.

We propose a model of simultaneous entry decisions for N symmetric firms across M heterogenous markets. There are externalities across markets in that a firm's revenues on a market are a function of entry decisions across the M-1 other markets. In this model, each firm is endowed with a vector of private marginal cost signals. Namely, a firm does not observe its competitors' costs at the time it selects its entry strategies across the M markets. Cost signals are drawn independently and identically across firms and markets from a joint probability density function which is common knowledged among firms. Hence, a firm decides simultaneously whether or not to enter and, if it enters, how much to produce on each market based upon its costs vector and the distribution of its rival costs.

This model defines a complex optimization problem since a firm's optimal strategy depends need not only upon the characteristics of the M markets, but also upon the uncertainty regarding its rivals' costs. Firms have to commit to quantity choices on a market without knowing ex-ante how many firms will compete on that market. The model is, in fact, analytically untractable and we propose an algorithm based upon Monte Carlo simulations to determine numerically the Nash Equilibrium solution.

We first solve the first-order conditions to determine systems of equations defining a treshold cost value above which a firm does not to enter a market and, pending entry, the optimal quantity to produce on that market. We then proceed to simulate repeatedly the game in order to approximate the expected quantity produced by a firm on each market. This numerical algorithm provides the estimated probability of entry on each market. Finally, we apply the inference method recently proposed by Florens et al. (1997) to endogenously estimate the parameter θ of the cost density function.

As an application, we consider the entry decisions of American Airlines and United Airlines at their Chicago O'Hare hub airport. These two airlines not only share similar cost structures at the network level, but they have similar brand images and market structures at O'Hare. They also dominate the Chicago O'Hare market. Both airlines are taken to simultaneously select their seat capacity (on nonstop flights) across a sample of airport-pair markets from Chicago O'Hare. The sample data, from the third quarter of 1993, consists in 85 markets with flight service from at least one of American Airlines or United Airlines, and in 15 markets with no nonstop flight service.

The results yield expected seat capacity and revenue levels on a market. These closely match observed mean values on the sample markets. We also estimate a mean value for the cost per available seat mile (CASM) of $0.127 for American Airlines and United Airlines at Chicago O'Hare. CASM's are the standard measure of costs in the airline industry and trade publications provide mean values, at the network level, ranging from $0.10 to $0.13 for both airlines.

We further determine treshold marginal cost values above which an airline does not enter on a market. From these tresholds, we obtain the estimated probability that an airline offers flights on a market. A panel data comparison of entry decisions and seat capacity levels for each sample market reveals that markets with the lowest estimated probability of entry have, in fact, experienced greater relative fluctuations in the number of entrants and in seat capacity levels over time.

The structure of the paper is as follows. Section 2 presents the theoretical model and section 3 proposes a numerical algorithm to determine its Nash Equilibrium solution. Section 4 discusses the application to the airline industry while section 5 outlines the estimation of revenue schedules for airline markets. The details on the econometric method to estimate games of incomplete information are in section 6. Section 7 describes the sample data and the estimation results. We conclude in section 8.

2. The Theoretical Model

We consider a Cournot model with M markets $(m = 1, ..., M)$ and N symmetric firms $(i = 1, ..., N)$. Firms are endowed with a vector of private signals $c_i = (c_{i1}, ..., c_{iM})$, where c_{im} represents firm i's constant marginal cost on market m. The signals c_{im} are drawn independently and identically across firms and markets from a probability density function (hereafter p.d.f.) $f(c_{im}/\theta)$, where θ denotes a vector of parameters. The p.d.f. $f(.)$ and the parameter θ are common knowledge to the firms. Rival firms private signals c_{-i} are not observed at the time firm i selects its strategy. The economist observes only the density function $f(.)$.

The inverse demand specification for firm i on market m is linear, symmetric across both firms, and it allows for externalities in demand across markets. Namely, firm i's price on market m, P_{im}, is a function of firm i's quantity choices across all M markets:[1]

$$P_{im} = \alpha_m + \beta_m \sum_{m' \neq m}^{M} q_{im'} - \gamma_m \sum_{j \neq i}^{N} q_{jm} - \delta_m q_{im} \tag{2.1}$$

where q_{im} is the quantity produced by firm i on market m and the values of the parameter vector $(\alpha_m, \beta_m, \gamma, \delta)$ are common knowledge. Setting fixed costs at zero, the profit of

[1] From a conceptual point of view, we could have allowed for externalities in the cost function as well. The demand approach is more appropriate to an airline setting.

firm i on market m is:

$$\pi_{im} = [P_{im} - c_{im}]\, q_{im} I_{\{q_{im}>0\}} \tag{2.2}$$

where $I_{\{q_{im}>0\}}$ is the indicator function defined as

$$I_{\{x>0\}} = \begin{Bmatrix} 1 & when\ x > 0 \\ 0 & otherwise \end{Bmatrix} \tag{2.3}$$

Given their private cost c_i, firms simultaneously select a vector of non negative quantities $q_i^* = (q_{i1}^*, .., q_{iM}^*)$ $(\forall i = 1, ..., N)$ as to maximize total expected profits over the network of M markets:

$$\begin{aligned} q_i^* &= \varphi(c_i, \theta) = \underset{\{q_{im}\}_{m=1,\dots M}}{Arg\max} \sum_{m=1}^{M} E_{c_{-i}}[\pi_{im}/\theta] \qquad (2.4) \\ \text{subject to} \qquad q_{im} &\geq 0 \qquad \forall m = 1, ..., M \end{aligned}$$

where $\varphi(c_i, \theta)$ is the equilibrium strategy function. Since firms are ex-ante symmetric we have $\forall j \neq i$

$$E_{c_{-i}}[q_{jm}/\theta] = E_{c_{-i'}}[q_{j'm}/\theta] = E[q_m/\theta] \qquad \forall i \neq i'\ or\ \forall j \neq j' \tag{2.5}$$

The optimization problem can then be written as:

$$\begin{aligned} q_i^* = \varphi(c_i, \theta) = \underset{\{q_{im}\}_{m=1,\dots M}}{Arg\max} \textstyle\sum_{m=1}^{M} \quad & [\alpha_m + \beta_m \sum_{m' \neq m}^{M} q_{im'} \\ & - (N-1)\gamma_m E[q_m/\theta] - \delta_m q_{im} q_{im} I_{\{q_{im}>0\}} \qquad (2.6) \end{aligned}$$

$$\text{subject to} \qquad q_{im} \geq 0 \qquad \forall m = 1, ..., M$$

Equilibrium strategies $\varphi(c_i, \theta)$ are, therefore, symmetric across firms.

3. Computing the Nash Equilibrium solution

The theoretical model presents three key characteristics. First, it is a game of incomplete information since rivals marginal costs are unknown at the time of decision. Second, firms decide simultaneously whether to enter and how much to produce on each market. Third, the quantity produced on market m creates an externality affecting profits, and consequently entry and quantity decisions, on all other markets $m' \neq m$. To determine the Nash Equilibrium strategy a firm must consider jointly all markets and the uncertainty on its rivals costs. Such strategy cannot be determined analytically. Instead we propose an algorithm to calculate the Nash Equilibrium solution numerically.

First, we remark that the constraint $q_{im} \geq 0 \; \forall m = 1, ..., M$ can be redefined equivalently by

$$q_{im} \geq 0 \quad \Leftrightarrow \quad c_{im} \leq \overline{c}_{im} \qquad \forall m = 1, ..., M \tag{3.1}$$

where $\overline{c}_{im}$ is a treshold cost defined such that firm i decides to enter market m only when its marginal cost is sufficiently low ($c_{im} \leq \overline{c}_{im}$). Note that $\overline{c}_{im}$ is determined only at the equilibrium and it is a function of firm i marginal costs on every market (c_i).

The first-order conditions of the optimization model (2.6) can be written $\forall m = 1, ...M$,

$$q_{im} I_{\{c_{im} \leq \overline{c}_{im}\}} = \frac{1}{2\delta_m} \left[\alpha_m + 2\beta_m \sum_{m' \neq m}^{M} q_{im'} I_{\{c_{im'} \leq \overline{c}_{im'}\}} - (N-1)\,\gamma_m E\left[q_m/\theta\right] - c_{im} \right] I_{\{c_{im} \leq \overline{c}_{im}\}} \tag{3.2}$$

When $c_{im} = \overline{c}_{im}$ we have $q_{im} = 0$ and the previous equation becomes

$$0 = \alpha_m + 2\beta_m \sum_{m' \neq m}^{M} q_{im'} I_{\{c_{im'} \leq \overline{c}_{im'}\}} - (N-1)\,\gamma_m E\left[q_m/\theta\right] - \overline{c}_{im} \qquad \forall m = 1, ...M \tag{3.3}$$

or equivalently,

$$\sum_{m' \neq m}^{M} q_{im'} I_{\{c_{im'} \leq \overline{c}_{im'}\}} = \frac{-\alpha_m + (N-1)\,\gamma_m E\left[q_m/\theta\right] + \overline{c}_{im}}{2\beta_m} \qquad \forall m = 1, ...M \tag{3.4}$$

If we insert equation (3.4) in equation (3.2) we get

$$q_{im} I_{\{c_{im} \leq \overline{c}_{im}\}} = \frac{\overline{c}_{im} - c_{im}}{2\delta_m} I_{\{c_{im} \leq \overline{c}_{im}\}} \qquad \forall m = 1, ...M \tag{3.5}$$

Finally let us insert equation (3.5) into equation (3.3):

$$0 = \alpha_m + \beta_m \sum_{m' \neq m}^{M} \frac{(\overline{c}_{im'} - c_{im'})}{\delta_{m'}} I_{\{c_{im'} \leq \overline{c}_{im'}\}} - (N-1)\gamma_m E[q_m/\theta] - \overline{c}_{im} \qquad \forall m = 1, ...M \tag{3.6}$$

If the expected quantity produced by a firm on each market $E[q_m/\theta]$ ($\forall m = 1, ...M$) were known and provided a vector of costs $\{c_{im}\}_{m=1,...M}$ we could solve the system of equations (3.6) to obtain the vector $\{\overline{c}_{im}\}_{m=1,...M}$. Then, using the set of equations (3.5), we could calculate the equilibrium strategies $\{q^*_{im}\}_{m=1,...M}$. However, there is no analytically tractable way to calculate $E[q_m/\theta]$. We propose to replace $E[q_m/\theta]$ by a Monte Carlo approximation $\widehat{E}[q_m/\theta]$.

The idea of the algorithm is to consider a value for $E[q_m/\theta]$, say ε_m, then for a given set of simulated costs we can solve the systems of equations (3.5) and (3.6) to obtain some simulated quantities for player i. Finally, we take advantage of the symmetry of the model to compare the empirical mean of the simulated quantities of player i and ε_m. If ε_m is a reasonable approximation of the expected quantity $E[q_m/\theta]$ then it should be close to the average simulated quantity. The approximation $\widehat{E}[q_m/\theta]$ is then defined as

$$\widehat{E}[q_m/\theta] = Arg\min_{\varepsilon_m} \left[\varepsilon_m - \frac{1}{MC} \sum_{l=1}^{MC} q^l_{im}\left(\widetilde{c}^l_{im}, \overline{c}^l_{im}, \varepsilon_m\right)\right]^2 \qquad \forall m = 1, ...M \tag{3.7}$$

where $\widetilde{c}^l_{im}$ ($\forall m = 1, ...M$ and $\forall l = 1, ..., MC$) is a Monte Carlo simulated cost drawn from $f(./\theta)$, MC is the size of the Monte Carlo simulation, and, $\overline{c}^l_{im}$ (respectively $q^l_{im}\left(\widetilde{c}^l_{im}, \overline{c}^l_{im}, \varepsilon_m\right)$) is solution of the system of equations (3.6) (respectively (3.5)), provided $\left\{\widetilde{c}^l_{im}\right\}_{m=1,...M}$ and $E[q_m/\theta] = \varepsilon_m$.

The determination of $\widehat{E}[q_m/\theta]$ may be time consuming but it is not computationally challenging. Indeed, the algorithm requires to estimate $\widehat{E}[q_m/\theta]$ for any value of θ which requires to solve numerically the system of equations 3.6 for each of the N Monte Carlo simulations. However, these equations are linear up to an indicator function and there exist numerous numerical procedures to solve these systems in a matter of seconds. Once $\widehat{E}\left[q_m/\theta\right]$ has been determined we can calculate $\{q^*_{im}, \overline{c}_{im}\}_{m=1,...M}$ for a given cost vector $\{c_{im}\}_{m=1,...M}$, or symmetrically we can invert the strategy function $\varphi^{-1}\left(q^*_{im};\theta\right)$ and calculate $\{c_{im}, \overline{c}_{im}\}_{m=1,...M}$ for a vector of observed optimal quantities $\{q^*_{im}\}_{m=1,...M}$. As we shall see in the next section, the econometric technique subsequently used in the empirical application to the airline industry, requires the inversion of the equilibrium strategy.

4. An Airline Application: American and United at Chicago O'Hare

This section details the application of the theoretical model to an airline environment. We describe, the airline setting of interest, and provide an empirical basis for the maintained hypotheses of the theoretical model.

4.1. Firms, Markets

American Airlines and United Airlines are two major US carriers, with similar cost structure (see table 1), sharing a primary hub airport at Chicago O'Hare. They not only have similar brand images in Chicago, but also serve a similar route structure from that airport. For example, United has nonstop flights on all seventy-two Chicago-markets with (nonstop) flights from two or more airlines, American on 69. Based upon the Databank DS T-100 data, they control 89.55% of passenger enplanements at O'Hare. As a comparison, Delta Airlines, the *third* largest carrier at O'Hare, has only 3.3% of passenger enplanements and offers flights on just 8 of the 124 Chicago-markets. The market structure at Chicago is also very stable over time. Over a period stretching from the third quarter of 1992 to the fourth quarter of 1993, there is little new entry/exit on Chicago-markets and almost all of them are the result of decisions from either of

American and United.

Therefore, following Brander and Zhang (1990, 1992), we assume that American Airlines and United Airlines are two symmetric firms in duopoly competition at Chicago O'Hare.

4.2. Decision Variables

We model entry decisions of American Airlines and United Airlines on markets from Chicago O'Hare airport. The decision variables are the seat capacity on (nonstop) flights, q_{im}, each of these airlines offers on Chicago markets. The choice of seat capacity, over price or passenger volume, is prompted by the nature of the industry. Flight schedules, which detail departure times and aircraft type, change little over time once published and, if anything, at a much slower rate than prices. The preference of seat capacity over passenger volume relates to the cost structure at the market level. Modeling passenger choice would demand, in this entry model, a representation of flight frequency choices in order to properly characterize the fixed costs of entry since these would include both aircraft and airport-specific fixed costs. This would require a model with two-decision variables and greatly complicate the current structure of the model.

4.3. Marginal Costs

Our model calls for a representation of costs at the market level. The standard characterization of market-specific costs in the airline literature is one of a fixed cost per flight plus a constant marginal cost per passenger (references). This implicitly defines, graphically, a step function between the total aircraft costs and the seat capacity on a market. A linear approximation to this step function seems reasonable since seat capacity are modelled quarterly. This leads us to assume, at the market-level, a linear relation between total aircraft costs and seat capacity.

Fixed costs then represent the airport-specific fixed costs of offering flights on a market. At Chicago O'Hare, the number of markets with flights from American Airlines and United Airlines varies little around the period of our sample data (table). In addition, both airlines have dedicated airport facilities at O'Hare, allocated under 30-year leases, which underwent major updates in the early 1990's. This leads us to treat all airport-specific fixed costs as being sunked prior to our sample period.

We need a representation of an airline i's marginal cost per available seat on a market. Following Brander and Zhang (1990) and Morrison and Winston (1995), we propose the following specification for airline i's marginal cost per available seat on a Chicago market m:

$$\begin{aligned} & \text{marginal cost per available seat}_{im} \\ = \quad & CASM_i \times \sqrt{AVGLGHTCHI} \times \sqrt{MILES_m} \end{aligned}$$

where $CASM_i$ is the marginal cost per available seat-mile for airline i on Chicago markets, $AVGLGHTCHI$ is the average mileage length of a Chicago-market, and $MILES_m$ is the mileage of market m. $CASM$'s are the standard measure of costs for airlines and they are only reported on a network-wide or aircraft-type basis. For American Airlines and United Airlines in the third quarter of 1993, we find, across publications, $CASM$ figures ranging from 10 cents to 13 cents depending upon which particular categories of costs are accounted for.

Our article differs from existing airline studies, such as the aforementioned, in that we endogenously estimate mean $CASM$ values. Namely, we have that airline i on market m draws a private marginal cost signal $c_{im} = CASM_{im}$ from a p.d.f. $f(.|\theta)$ where values for the parameter vector θ are estimated endogenously. The marginal cost signals are drawn independently and identically across airlines and markets from the pd.f. $f(.|\theta)$. We feel comfortable with the assumption of an identical distribution for American Airlines and United Airlines at Chicago O'Hare airport. Chicago is a major hub in each airline's network and, as Brander and Zhang (1990) note, it plays a similar role in each airline's network. The airline trade literature also reports similar CASM figures for both airlines (table). We will look to relax the independence hypothesis in later work.

In summary, the total cost of q_{im} seats for airline i on market m is written as:

$$Cost\left(q_{im}\right) = CASM_i \times \sqrt{AVGLGHTCHI} \times \sqrt{MILES_m} \times q_{im}$$

5. Estimation of Revenue Schedules

5.1. Local and Connecting Passengers

Airline 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 (nonstop) 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 indirect route is a path made up of flights which links two airports and requires at least one stop at an intermediate airport.

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. This means that an airline may sell seats on market B-D to both local B-D and connecting A_i-E_j passengers. It is the presence of these connecting A_i-E_j passengers which leads to demand synergies across entry decisions on connected markets. Connecting A_i-E_j passengers account for, on average, 58% of an airline's passenger volume on a B-D market in the third quarter of 1993.

5.2. An Aggregated Specification

We are looking for an aggregated representation of revenues per seat at the market-level; i.e., a revenue schedule function for seat capacity. There are few, if any, references in the literature. Both Reiss and Spiller (1989) and Richard (1999) estimate aggregated market-level demand functions but the context of their analyses bears little comparisons with the present article. Berry (1992), meanwhile, specifies revenues on an airline market only in terms of market characteristics and the number of competitors.

We could model demand on an origin and destination basis and model a passenger's choice of an itinerary between two airports B and D through a logit specification (not unlike Lederer). There are two main problems with this approach. First, while logit models make for a rich demand specification, such specifications are not analytically tractable within this paper's framework. Second, the presence of connecting passengers means that, in our analysis of entry across M markets, we would have to account for demand across all pairwise combinations of M cities. To provide some measure of scale, we note that, in the third quarter of 1993, an airline on market B-D draws, on average, connecting A_i-E_j passengers from 57 indirect routes with one intermediate stop and 54 indirect routes with two stops.

We propose, rather, to talk of a revenue schedule for airline i's seat capacity on a market m. This schedule describes revenues per seat for airline i on market m as a function of seat capacity choices across connected markets. It is defined as the sum of two revenue schedule functions, one for each of local and connecting passengers, since both local and connecting passengers contribute to revenues at the market-level[2].

This approach requires, however, to assume an allocation rule for connecting ticket prices. Indeed, a connecting A_i-E_j passenger for a market B-D pays one ticket price for the indirect A_i-E_j route. To determine revenues from connecting passengers A_i-E_j on a market B-D requires allocating the price paid by these passengers among the various flights on their indirect routes. While a choice of allocation rule has some arbitrariness, mileage is not only the variable most highly correlated with airline prices, but also the primary cost variable. Hence, we allocate the ticket price of a connecting passenger to the various flights on a mileage basis.

Finally, unlike traditional models where demand and price are jointly determined, firms, in our framework, select their seat capacity, then subsequently observe the actual realization of revenues. Hence, estimation of the revenue schedules is by ordinary least-

[2]Cargo shipments can be a source of revenues for airlines on some markets. We unfortunately have no data on cargo shipments or value.

squares.

5.3. Sample Data for the Estimation of the Revenue Schedules

The data are compiled from three databases for the third quarter of 1993[3]. All three databases report data on a per airline, per airport-pair market basis (i.e., unit of observation is airline i on a market from airport B to airport D). Databank 1A, from the Department of Transportation (DOT), is a 10% random sample of all airline tickets sold each quarter. It yields, for this analysis, the data on ticket prices and on the number of local and connecting passengers. Databank DS T-100, another DOT database, provides monthly data on seat capacity and passenger volumes. Only the major U.S. airlines and their (directly-owned) subsidiaries (e.g., shuttle, commuter airlines) report to Databank DS T-100. We turn to the OAG North American editions for a complete listing of all scheduled flight operations for the third quarter of 1993. Given Databank 1A data are quarterly data, revenues and seat capacities are, for this paper, defined on a quarterly basis. Exogenous market characteristics are obtained from Census data.

For the estimation of the revenue schedules, we consider a sample of 919 airport-pair markets for the third quarter of 1993. The sample consists in all markets for which we have a complete set of third quarter of 1993 Databank 1A and Databank DS T-100 data (see appendix B for ampler details). Airport-pair markets in the sample are non-directional in that we have averaged the data for each market across both directional flows. Namely, the unit of observations is a market B-D rather than a market from B to D or from D to B.

5.4. A Revenue Schedule for Local Passengers

Sample markets have from one to four airlines with flights. An airline's share of passenger enplanements at the origin airport on a directional market has been argued to significantly determine the distribution of local passengers across competitors on that market

[3]Ampler details on the data, and construction of the variables, are provided in appendix A.

(Borenstein (1990)). In our sample, 84% of markets with two or more airlines include at least one hub airport for each competitor. By looking at markets on a non-directional basis, we are, therefore, evening out most of the discrepancies in airport presence on a market. This leads us to treat products (seats) on a market as homogenous across airlines with regards to local passengers[4].

For a market m with N airlines with flights, the revenue schedule specification is linear[5]:

$$\begin{aligned} REV_m^l = \left[\sum_{i=1}^{N} REV_{im}^l\right] / \left[\sum_{i=1}^{N} q_{im}\right] &= \alpha_0 + \alpha_1 MILES_m + \alpha_2 POPUL_m + \alpha_3 INCOME_m \\ &\quad + \alpha_4 \ln(INCOME_m) + \alpha_5 AREA_m - \delta^l \sum_i q_{im} \\ &= \alpha_m^* - \delta^l \sum_i q_{im} \end{aligned}$$

where REV_{im}^l is airline i's revenues from local passengers (across coach and first class) on market m and q_{im} is airline i's seat capacity (on nonstop flights) on market m. The explanatory variables are the mileage of market m ($MILES_m$), the population ($POPUL_m$), income level (INC_m), and square miles area ($AREA_m$) for both cities including the airports on the market.

Estimation results are provided in table 1[6].

[4]While we have data on an airline's airport presence, we have no data on airline-specific exogenous variables which could measure the demand from local passengers for a particular airline.

[5]The specification does not account for connecting B-D passengers. We do not have data on the number of seats allocated to connecting B-D passengers across the various flights on these passengers' indirect paths. We only observe the number of connecting B-D passengers per market which makes it difficult to incorporate these passengers in our specification. Richard (1999) also provides some evidence that local and connecting passengers may sufficiently differ in their flight frequency valuation, hence valuation of time, that we could consider the number of connecting B-D passengers on a market to be exogenous.

[6]With regards to these results, we note that the high implied (from estimated value of δ^l) elasticity of seat capacity with regard to revenues per seat from local passengers can be explained. One of the primary sources of dispersion in seat capacity across markets are varying levels of connecting passengers. Connecting passengers account, on average, for 58% of the passenger volume on a flight. In fact, connecting passenger volumes have higher correlation with seat capacity levels than local passenger volumes.

Table 1.
Estimated mean parameter values (standard deviations) for the revenue schedule function for local passengers
Dependent variable is REV_m^l

Cst	POPUL$_m$	MILES$_m$	INC$_m$	ln(INC$_m$)	AREA$_m$	δ^l slope
986.44	0.1463	0.0186	0.1268	-159.94	0.2556	-1.1×10^{-4}
(179.39)	(0.030)	(0.0011)	(0.0168)	(27.98)	(0.0594)	(1.1×10^{-5})

5.5. Connecting demand

The fact that an airline may sell seats on market B-D to connecting A_i-E_j passengers creates demand synergies between entry decisions across markets. In essence, the greater the number of markets with flights at each of airports B and D, the greater the scope of A_i-E_j markets an airline may draw passengers from for travel on B-D.

In our framework, we determine, for each airline i on a market m, that airline's total seat capacity across all markets j, $j \neq m$, at each of airport B and D; that is, $\sum_{j\neq m} q_{ij}^B$ at airport B and $\sum_{j\neq m} q_{ij}^D$ at airport D. The size of the revenue schedule for that airline's connecting passengers is then assumed proportional to $\sum_{j\neq m} q_{ij} = \max\left\{\sum_{j\neq m} q_{ij}^B, \sum_{j\neq m} q_{ij}^D\right\}$. This restriction to one of airports B or D is necessary to limit the scope of our analysis to entry decisions at a particular airport. It remains that on the 84% of sample markets where an airline has a single hub airport, the chosen airport (B or D) is always the hub airport. On these markets, almost all of an airline's connecting A_i-E_j passengers happen to transit through that hub airport.

An examination of the data also reveals that an airline draws rather similar numbers of connecting passengers from each A_i-E_j market since each such market contributes, on average, 1.68% (median is 0.89) of the airline's total connecting A-E passenger volume. Hence, we assume that any increase in capacity across markets j, j≠m, generates a constant increase in the number of connecting passengers on a market m. We do allow, nevertheless, for that increase to vary with market m's mileage.

Finally, if a connecting passenger switches airlines on an indirect route, that passenger is said to interline. Morrison and Winston (1995) document that, by 1994, fewer than 1% of all connecting passengers interline. We thus specify that the scope of synergies for an airline on market B-D is only a function of its own seat capacity choices and not its competitors'.

We have this linear specification for revenues per mile from connecting A_i-E_j passengers for airline i on market m:

$$\begin{aligned}\frac{REV_{im}^c}{MILES_m} &= \delta_0 + \delta_1 \ln(MILES_m) + \delta_2 MILES_m + \delta_3 HU1B_{im} \\ &\quad + \delta_4 HU2B_{im} + \delta_5 MAJHUB_m + \beta \sum_{\substack{j=1 \\ j \neq m}}^{M} q_{ij} - \delta^c \frac{q_{im}}{MILES_m} \\ &= \delta_{im}^* + \beta \sum_{\substack{j=1 \\ j \neq m}}^{M} q_{ij} - \delta^c \frac{q_{im}}{MILES_m}\end{aligned}$$

where $\frac{REV_{im}^c}{MILES}$ are airline i's revenues per mile from connecting A_i-E_j passengers; $HU1B_{im}$ is a dummy variable equal to 1 if airline i has one hub airport on market m; $HU2B_{im}$ is a dummy variable equal to 1 if airline i has two hub airports on market m; and $MAJHUB_m$ is a dummy variable equal to 1 if at least one of the airports on market m is one of Atlanta, Dallas Forth-Worth, Houston IAH, Los Angeles LAX, Miami, New York JFK, and Chicago O'Hare.

Estimation results are provided in table 2[7].

[7] A comment is warranted with regards to the low estimated value for the slope coefficient δ^c. Airlines draw connecting passengers from a wide range of markets. For example, in the third quarter of 1993, an airline on market B-D draws, on average, connecting A_i-E_j passengers from 57 indirect routes with one intermediate stop and 54 indirect routes with two stops. As mentioned in the text, they also draw similar numbers from each market. Hence, we do not expect the revenues per seat from connecting passengers to vary much with *total* seat capacity levels.

Table 2. Estimated mean parameter values (standard deviations) for the revenue schedule function for connecting passengers. Dependent variable is $REV^c_{im}/MILES_m$							
Cst	$\ln(\text{MILES}_m)$	MILES_m	HU1B_{im}	HU2B_{im}	MAJHUB_m	$\sum_{j\neq m}^{M} q_{ij}$	δ^c slope
0.303	-0.0446	0.0234	0.0259	0.0278	-0.0106	0.4350	-4.3×10^{-5}
(0.017)	(0.0028)	(0.0031)	(0.0021)	(0.0038)	(0.0013)	(0.0420)	(0.9×10^{-5})

5.6. The Revenue Schedule for Seat Capacity

The revenue schedule for airline i's seat capacity on a market m is defined as the sum of the revenue schedules for each of local and connecting passengers:

$$REV^l_m + REV^c_{im} = \alpha^*_m + \delta^*_{im} MILES_m + \beta MILES_m \sum_{\substack{j=1 \\ j\neq m}}^{M} q_{ij} - \left(\delta^c + \delta^l\right) q_{im}$$

6. Inference in Game Theoretic Models

The estimation of the private signals distribution necessitates non standard econometric techniques. Indeed, a key component of any game theoretic model with incomplete information is that unobserved private signals are transformed into observed actions $(q_i = \varphi(c_i, \theta))$. Besides, the strategic nature of games of incomplete information translates into the fundamental property that these strategies depend upon the underlying probability distribution of types. Consequently, one cannot estimate jointly the functional form of players' strategies and the distribution of types from the sole observation of actions, and one cannot directly estimates the distribution of unobserved types. The specification problem is traditionally solved by imposing that strategies are Nash Equilibrium solutions of the game. To estimate the distribution of unobserved types, we adopt the generic estimation principle recently proposed by Florens et al. (1997). Within this estimation framework, one initially selects an 'unfeasible' estimator $\widetilde{\theta}(c)$, whereby one could estimate θ if the cost of all firms on every markets $(c = (c_1, ..., c_N))$ were known. The corresponding 'feasible' estimator $\left(\widehat{\theta}(q), \widehat{c}(q)\right)$ $(q = (q_1, ..., q_N))$ of (θ, c) is defined as the fixed point solution[8]

$$\widehat{\theta}(q) = \widetilde{\theta}(\widehat{c}(q)) \quad , \tag{6.1}$$

$$\widehat{c}(q) = \varphi^{-1}\left(q; \widehat{\theta}(q)\right), \qquad i : 1 \to N \quad . \tag{6.2}$$

where $\varphi^{-1}\left(q; \widehat{\theta}(q)\right)$ is the inverse strategy function calculated with the algorithm introduced in section 3. In practice, the computation of such a fixed point solution necessitates iterating between equations (6) and (7) until convergence obtains. See Armantier and Richard (1998) for additional numerical considerations.

Private signals are assumed to have a truncated normal distribution on $]0, \infty[$

$$f(c_{im} \mid \mu, \sigma) = \frac{1}{1 - F(0 \mid \mu, \sigma)} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(c_{im} - \mu)^2}{2\sigma^2}\right] \tag{6.3}$$

where $F(0 \mid \mu, \sigma)$ is the cumulative distribution function of a normal distribution $N(\mu, \sigma)$.

We estimate $\theta' = (\mu, \sigma) \in \Re \times]0, \infty[$ with the inference method developed by Florens and al. (1997). The unfeasible estimator is given by the censored Maximum Likelihood estimator:

$$L(\mu, \sigma / \widehat{c}(q)) = \prod_{i,m} \left(1 - \frac{F(\widehat{c}_{im}(q) \mid \mu, \sigma)}{1 - F(0 \mid \mu, \sigma)}\right)^{I_{\{q_{im}=0\}}} \left[f(\widehat{c}_{im}(q) \mid \mu, \sigma)\right]^{I_{\{q_{im}>0\}}} . \tag{6.4}$$

Computing is of the order of 285 minutes of CPU time on a 7 years old DEC workstation, for a Monte Carlo simulation size of $MC = 1000$.

[8] Conditions for the local identification of θ from the sole observation of q and for the existence and (local) unicity of a fixed joint solution are found in Florens et al. (1997), together with characterizations of the asymptotic distributions of $\widehat{\theta}$ and $\widehat{c}$.

7. Structural Estimation of the Airline Model

7.1. The Sample Data

We are now in a position to define the sample data for our analyses. There are, in the third quarter of 1993, 124 Chicago markets with flights. For six of these markets, no data are reported in Databank DS T-100. For 33 of these markets, there are either asymmetries in the number of hub airports (i.e., in the values for $TWOHUB_{im}$ in the revenue schedule specification) between American Airlines and United Airlines, or there is another airline (i.e., Delta, etc.) with flights. On the remaining 85 Chicago markets, American Airlines and/or United Airlines are the only airlines with flights. In fact, we do not observe any other airline with flights on these markets over a time period stretching from the third quarter of 1992 to the fourth quarter of 1993. We include these 85 markets in our sample.

To this sample of 85 markets, we add 15 markets with no flights from any airline (e.g., the 15 markets with highest estimated value for the intercept of the revenue schedule function, $\alpha_m^* + \delta_{im}^* MILES_m$). Hence, our sample includes markets with zero, one, or two airlines and contains, overall, 100 different Chicago markets.

The six markets with missing data link Chicago to a small city and have, at best, minimal passenger volumes. These markets are ignored. For the 33 markets with asymmetries and/or additional competition, we take the seat capacities for American Airlines and United Airlines on these routes as a given. Given the stable environment at Chicago O'Hare in terms of entry/exit decisions, these seat capacities vary little over a time period stretching from the third quarter of 1992 to the fourth quarter of 1993 (see table). In the third quarter of 1993, American Airlines' seat capacity across these 33 markets amounts to 2,396,654, while United Airlines' is 2,705,356. To maintain the symmetry of the model, we average these capacities across both airlines and use this average value, 2,551,005, as the starting value for $\sum_{j \neq m}^{M} q_{ij}$ in the revenue schedule specification.

7.2. Estimation Results

Note to the reader: We are in the final stages of estimating our model over our full sample of 100 Chicago markets. This means that our final results are not yet available.

Nevertheless, before proceeding with a full sample estimation, we isolated a sample of 15 markets: two randomly selected monopoly markets and 13 randomly selected duopoly ones.

Here are the data (table 1) and primary results (table 2) for these 15 markets:

Market	3-letter airport code	miles	Observed q_{1m}	Observed q_{2m}
1	ABQ	1118	0	46122
2	BMI	116	0	35210
3	CMI	135	39734	45300
4	DBQ	147	47984	21066
5	ELP	1236	18618	28064
6	EVV	273	64500	82716
7	FAR	557	65850	56704
8	FNT	223	21248	50537
9	ISP	776	50356	68884
10	LAF	119	15924	12624
11	LSE	215	44370	54612
12	MKG	118	117896	208482
13	RFD	63	35989	67297
14	RST	268	17848	19838
15	TOL	214	21850	24386
Average		305.2	37477.8	54789.46
Standard Deviation		319.63	30302.87	47133.63

Table 7.1: Data

Tables 1 indicates that these routes offer a large variation of miles per routes, and a large variation of quantities across routes and across companies.

The estimated parameters are $(\widehat{\mu}, \widehat{\sigma}) = (0.126, 7.167E-04)$ with a standard deviation calculated with a Monte Carlo simulation of $(0.011, 1.125E-05)$. This corresponds to an estimated mean and standard deviation for the marginal cost of $(0.127, 2.661E-02)$. In other words, the marginal cost of a seat per miles is in average 0.127\$. This compares with $CASM$ figures ranging, across trade publications, from 10 cents to 13 cents depending upon which particular categories of costs these are accounted for. The main estimation results are summarized in Table 2.

Market	$\widehat{E}_i\,(q_{im})$	$\widehat{E}_i\,(\overline{c}_{im})$ (in \$)	Expected Cost per passenger(in \$)	Percentage of Entry
1	37337.24 (11765.37)	43.58 (17.5)	36.20 (12.6)	64.0
2	32452.48 (25545.56)	59.18 (35.2)	60.13 (31.8)	43.2
3	42427.88 (21879.61)	84.81 (29.0)	91.53 (30.7)	65.4
4	34040.28 (23600.15)	101.79 (59.6)	123.05 (58.4)	49.7
5	24326.44 (12338.51)	24.49 (10.2)	15.44 (8.7)	92.8
6	78751.27 (31673.43)	97.73 (29.7)	99.12 (26.1)	85.6
7	62512.71 (36195.88)	64.93 (33.7)	59.87 (28.6)	63.2
8	34333.99 (11341.31)	31.36 (17.4)	24.81 (12.9)	89.0
9	66938.26 (19676.40)	48.41 (21.3)	37.47 (19.4)	91.4
10	15222.85 (9711.07)	30.68 (17.5)	26.96 (14.4)	70.2
11	44641.60 (17491.55)	40.44 (22.4)	30.38 (17.0)	93.2
12	21266.98 (16777.48)	95.20 (31.8)	77.47 (26.5)	86.0
13	43446.38 (26510.79)	43.57 (26.6)	37.85 (22.1)	75.8
14	18042.52 (17743.24	53.33 (40.1)	58.48 (39.5)	35.9
15	24522.82 (13289.84)	27.23 (9.6)	19.87 (8.3)	95.6

Estimated parameter values with standard deviations in parentheses.

Table 7.2: Estimated Model

Overall the observed quantities (Table 1) are close to the expected quantities estimated by the econometric model (Table 2) and well within one standard deviation. We are also able to estimate the average treshold cost $\widehat{E}_i\,(\overline{c}_{im})$ above which a company do not enter the market, the average total cost and the percentage of entry per period on a given route.

8. Conclusion

[to be completed]

References

- ARMANTIER O., and J-F. RICHARD (1997). "Inference and Constrained Equilibrium in Game Theoretic Models: Numerical Aspects", invited by the Journal of Computational Economics for a special issue based on the 1997 Conference on Computing in Economics and Finance.

- BRANDER, J. AND ZHANG, A. (1990). "Market Conduct in the Airline Industry: An Empirical Investigation." RAND Journal of Economics. Vol. 21, pp. 567-583.

- BRANDER, J. AND ZHANG, A. (1992). "Dynamic Oligopoly Behaviour in the Airline Industry." Mimeo. University of British Columbia.

- BERRY, S. (1992). "Estimation of a Model of Entry in the Airline Industry." Econometrica. Vol. 60, pp. 889-917.

- BORENSTEIN, S. (1991). "The Dominant-Firm Advantage in Multi-Product Industries: Evidence from the U.S. Airline Industry." Quarterly Journal of Economics. Vol. 10, pp. 1237-1266.

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

- BRESNAHAN, T. AND REISS, P. (1991). "Entry and Competition in Concentrated Markets." Journal of Political Economy. Vol. 99, #5, pp. 977-1009.

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

- FLORENS, J-P., PROTOPOPESCU C. and J-F. RICHARD (1997). "Inference in a Class of Game Theoretic Models." Mimeo. University of Pittsburgh.

- GELFAND, M. AND P. SPILLER (1987). “Entry Barriers and Multiproduct Oligopolies: Do They Forebear or Spoil?.” International Journal of Industrial Organization. Vol. 5, pp. 101-113.

- KADIYALI, V. (1996). “Entry, its Deterrence, and its Accomodation: A Study of the US Photographic Film Industry.” RAND Journal of Economics. Vol. 27, #3, pp. 452-478.

- KATZ, M. AND SHAPIRO, C. (1985). ”Network Externalities, Competition, and 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.

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

- RICHARD, J-F. (1997). “Simulation Techniques”, ch 23 in the Econometrics of Panel Data, a Handbook of the Theory with Applications”, edited by L. Matyas and P. Sevrestre, 2nd Edition, Kluiver, Boston.

- RICHARD, O. (2000). “An Empirical Analysis of Marketing Alliances between Major US Airlines.” Bradley Policy Research Center. University of Rochester.

- SUTTON, J. (1991). “Sunk Costs and Market Structure.” Cambridge, Mass.: MIT Press.