Estimation of A Dynamic Oligopoly Entry Game in the US Airline Industry: Hubs, and LCC

Moshe Cohen*†
Columbia Business School

February 21, 2011

Abstract

Airlines choose the domestic markets – city pairs – they serve and the prices they charge given the structure of their network and the networks of rival airlines. I cast this choice into a dynamic oligopoly entry game to recover the fixed and variable operating costs, entry costs, and profits, using a panel of 20 quarters of DB1B and T-100 Domestic Segment Data. These estimates are then used to analyze the strategic and cost saving effects of hubs, and LCC. I find that while hubs produce benefits consumers value which translate into higher variable profits, but when including fixed costs their desirability is much less clear even in hub markets. LCC, and especially Southwest and JetBlue are especially attractive to consumers, have lower marginal costs and have a strong negative impact on the profits on the incumbents in the markets they serve.

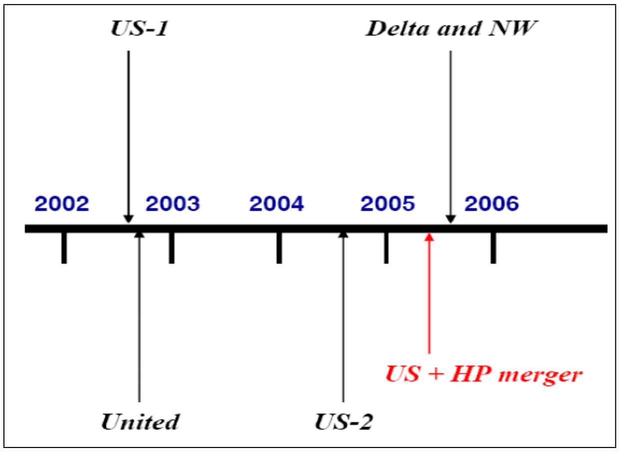
1 Introduction

The recovery and analysis of airline profits and their determinants have long been elusive. Airlines, by and large, have lost money since the invention of the first planes by the Wright brothers in 1903. As shown in figure 1(a), several of the large legacy carriers have been in and out of Chapter 11 in recent years (compare Borenstein and Rose 2007). However, the number of passengers has increased steadily over this time period[1], and as shown in figure 1(b) the operating revenues in the industry have risen significantly as well (following the post September 11th decline). In this paper I develop and estimate a dynamic model of airline competition, with entry and exit, and recover the supply and demand sides of the domestic US airline market. The

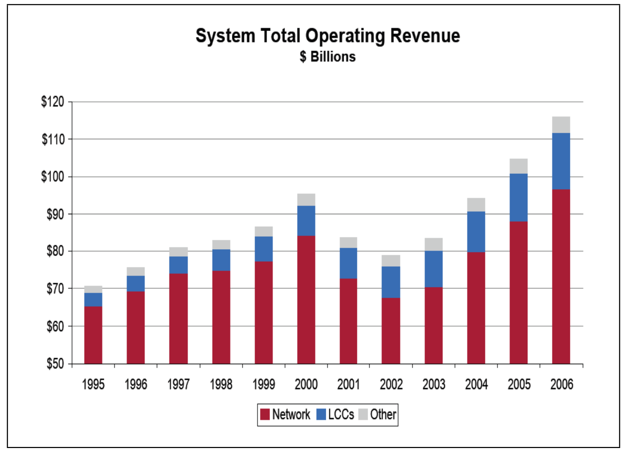
*E-mail: mac2310@columbia.edu / Land-mail: Columbia Business School, 3022 Broadway, Room 819, New York, NY, 10027.

†I thank my advisors: Glenn Ellison, Whitney Newey and Victor Chernozhukov for their valuable input throughout the thesis writing process. I also thank Daniel Gottlieb, Jerry Hausman, Konrad Menzel, Paul Schrimpf, Nancy Rose, and Stephen Ryan for useful comments, discussions and suggestions.

[1] The number of domestic passengers has gone from 551M in 2002, to 679M in 2007. There is some decline in 2008 to under 650M.



(a) Recent Bankruptcies



(b) Operating Revenues

Figure 1: Recent Trends

model exploits the information embedded in the decisions of consumers in choosing between the many airline product offerings in each quarter, regarding consumer preferences, demand, and marginal costs, as well as the information embedded in the quarterly decisions of entry and exit, which are assumed to reveal each airline's belief of the (expected present discounted) profitability of the markets it chooses to serve.

As is common in the dynamic game literature, this paper makes the simplifying assumptions of airlines maximizing the profits from each market separately and not taking into account the added benefits to the entire network; of the transitions between states following a first order Markov process where the payoff relevant variables are only the market specific variables; of the individual airline transition probabilities being independent conditional on the state space; and while the model can have multiple equilibria, this paper assumes that the data is generated by one (and only one) of them. It also specifies a relatively simple nested logit demand model over the quarterly tickets sales, which is used as an input to the dynamic game. However, this paper extends previous applied work on dynamic games by allowing for firm identities to matter, while accounting for the impact this has on the size of the state space by structuring the model in a manner which facilitates the use of state-of-the-art mathematical solvers. This allows for the exploration and exploitation of the highest level of richness in airline costs and profits possible under the current computational optimization technology.

The recovery of the costs of serving the heterogenous US markets - and consequently a better understanding of the profit structure - allows for an analysis of many of the key questions that have been at the focus of the vast airline literature both within and outside of economics. This paper addresses two of them: The desirability of hub networks, and the impact of the increased presence of low cost carriers (LCC) on the costs and profits of incumbent airlines, which I detail below:

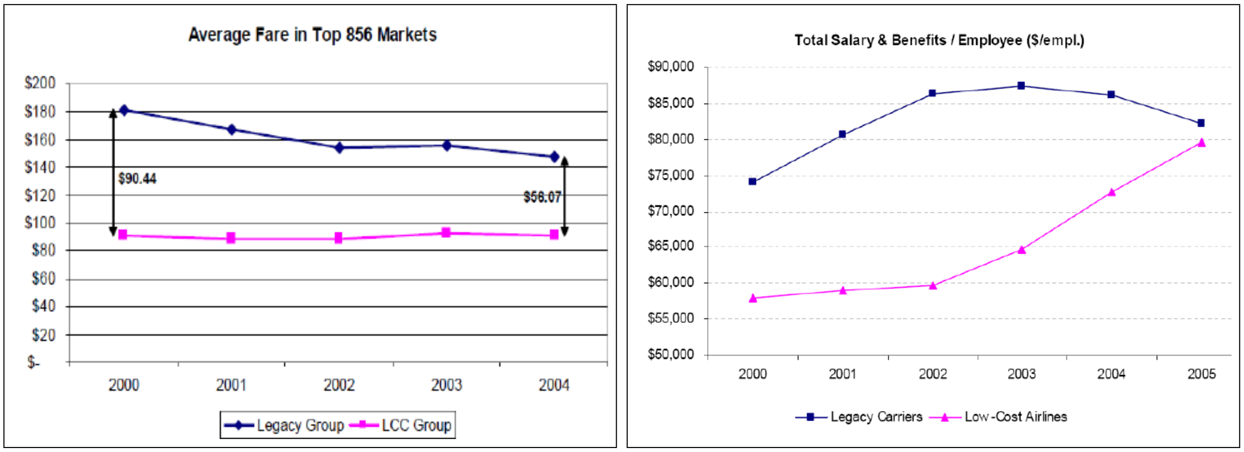
Figure 2: Hubs in 2008

1.1 The Case for Hubs

Following the deregulation of airlines in the late 1970s, many of the "legacy carriers" chose to concentrate a large portion of their operations in certain airports (the "hubs") and to connect the other cities served (the "spokes") to these hubs by non stop flights. Figure 2 shows the distribution of hubs across the US. The rationale for hubs is that there are significant benefits or returns to scale from having a large presence in the hub airport that outweigh, in certain situations, the additional costs (and inconvenience to passengers) of having the many additional connecting flights, and travelling larger distances when serving two spoke end points (compare Hendricks, Piccione and Tan, 1995) . Hubs facilitate the use of larger planes which reduce the cost per passenger and allow for a reduction in the number of direct connections. They also allow for economies of scope in having a concentration of manpower in the hub, and may also allow for more bargaining power with the airports. They have been claimed to also be attractive to consumers since they offer more variety and frequency of flights (at the hub), and more expertise[2]. There are thus potentially both supply (cost savings) and demand (revenue increases) advantages to hubs. Finally, there are claims that hubs serve as an entry deterrent, given the complementarities in profits between different routes (any city added connects the entire network to that city).

Given the many puzzles surrounding the airline industry, these claims require empirical support, by examining the determinants of profits in general and fixed costs in particular. I

[2] Separating variety and frequency is difficult since both measures, but frequency in particular, are largely demand driven.



(a) Fares

(b) Salaries

Figure 3: Convergence in Operating Procedures: Legacy and LCC

find that hub carriers have higher variable profits that their competitors in their own hub markets but lower than average profits in non-hub markets. Consumers prefer to travel with the hub carrier in hub markets and have a significant distaste for flying with other carriers in these markets. They also prefer to travel with airlines that have more destinations from the origin, or more flexibility. However, the static game estimates further reveal a distaste for connecting flights, or a preference for nonstop flights, which can offset the preference for using the hub airlines. Furthermore, the profits garnered by carriers in their hub markets are significantly reduced when including fixed costs and they do not increase the costs of entry for other carriers. These estimates further question the desirability and overall profitability of the hub structure, even in the hub markets themselves.

1.2 The Impact of Low Cost Carriers (LCC) on Rival Costs and Profits

The **second** question analyzed is the heterogenous effects of the different carriers on their rivals. There is a natural interaction between the six legacy carriers. However, a key feature has been the growing market share of low cost carriers, leader amongst which is Southwest Airlines, followed, more recently, by JetBlue. The differing products offered by these carriers induce a response by the actual or potential other players in the market. There seems to be a convergence in operating procedures (such as less food on flights), and as figure 2 shows, a convergence in fares and labor costs as well. Previous work, including Berry, 1992, and, more recently, Ciliberto and Tamer 2009 examined the effects of firm heterogeneity on entry into airline markets in a static framework. This paper extends this work to a dynamic framework, allowing for firm identities, and thus for heterogeneous effects of the legacy carriers and low cost carriers on the profits of their rivals, accounting for the heterogeneity in markets. The interactions between players are decomposed into their differing effects on rival variable profits and marginal costs (both part of a static pricing game), as well as on the costs of entry and the fixed operating costs. I find variable profits to generally be higher for the LCC. Consumers have a significant preference for these airlines in general (accounting for observable features of the products offered), but especially for Southwest and JetBlue, upon which I therefore focus. This likely reflects some unobservable features of the services provided (or the characteristics of their products). Marginal costs are also significantly lower for these airlines, suggesting that they are doing something better on the cost side. Furthermore, there are large strategic effects between airlines, but the (negative) impact of the presence of LCC on the profits of legacy competitors is most pronounced. The preliminary results for the entry costs and fixed costs suggest that there too there are significant strategic effects, where the presence of all major airlines increases the costs associated with entry and the fixed costs, but the LCC play a special role.

The paper is organized as follows: in section 2 I briefly survey the relevant literature. Section 3 discusses the datasets used and the construction of the sample. Section 4 sets up the general framework and the model. Section 5 then discusses the estimation, separating the variable profits and marginal costs recovered from a static pricing game, and the entry and fixed cost parameters recovered from the dynamic game. Section 6 collects and discusses the results. Finally, section 7 concludes and suggests some of the many extensions and future work that can be done using the framework in the paper.

2 Related Literature

As to the methodology, there are a number of dynamic game applications using one of two methods. The first is a simulation of moment inequalities approach, where the value functions are simulated forward using policy functions and transition probabilities estimated from the data, and the estimates are those under which value functions of the policies chosen are greater than the value functions from any alternative policies (as developed by BBL 2007). The second is a maximum likelihood with best response equality constraints approach (as developed by AM 2007), which will be discussed at greater length below. Given the complexity of these methods, previous applied work, including Ryan 2009 (cement, BBL), Collar-Wexler 2005 (ready-mix concrete, AM), Sweeting 2007 (radio, BBL), Beresteanu and Ellickson 2006 (supermarket products, BBL) and Macieira 2006 (supercomputers, BBL), generally assumes symmetric equilibria. This is a difficult assumption to justify in most settings. This paper is closer to the AM approach, but allows for individual airline heterogeneity and a state space visible to all players which transitions consistent with the equilibrium probabilities thus allowing for the treatment of the effect of the airlines' own and rival market features (such as hubs), by exploiting state-of-the-art optimization techniques.

The application, the study of airlines, has received much treatment in a voluminous literature spread across many fields, for which there are now many useful surveys (see for example Borenstein and Rose 2007 and 2008). I briefly sample that most relevant to the analysis here.

For entry into airlines markets, Berry (1992) which builds on Bresnahan and Reiss (1991) analyses entry as a static game of discrete choice. The profit function is restricted to ensure the uniqueness of the number of players, by assuming firm characteristics only affect the fixed costs, and a symmetric post entry game. Ciliberto and Tamer (2009) use a similar framework, but rather than requiring a unique number of entrants (and restricting the profit function to ensure this), they consider multiple equilibria and allow for a different number of entrants and different selection mechanisms in different markets. Their estimation builds upon the set estimation procedures in Chernozhukov Hong and Tamer (2007), and thus they allow for multiple equilibria within a set constrained by the requirement of airlines earning positive profits in markets. The identified estimated parameters are those for which there is a selection function such that the predicted choice probabilities in the model match the empirical choice probabilities in the data. Both papers find heterogeneity in the manner in which competitors' profits are affected by the presence of their rivals and an important role played by airport presence. This paper is similar in its identifying assumption of airlines operating in city pair markets when they produce positive profits, however rather than looking at static single period profits I require the difference between the value function of operating and of not operating in the period to be positive. The dynamic framework relies on the assumption that the equilibrium estimated is the one most likely given the data. This framework allows for an estimation of fixed and entry costs from the moments of entry and exit, and for the study of both the strategic effects airlines impose on one another and the effects of hubs. However, it requires a full specification of the state space today and in all possible future periods and of the transitions to and from all elements of the state space.

Aguirregabiria and Ho (2009) analyze hubs in a dynamic framework as well, and as such are closer to this paper. Their markets are city pairs, but their incumbency in a market is defined by operating non-stop flights in the market. The analysis is restricted to 2004. The restrictions they impose on the state space require binning all states with four or more incumbents together and discretizing the variable profits to a grid of 11 points. They also do not include the characteristics[3] and identity of rival airlines in the state space observed by each airline. Their paper thus focuses on the effect of hubs on airlines' own profits. They find much lower fixed costs, entry costs and variable profits than those found here. In contrast, this paper looks at a larger time period and specifically includes the identity and characteristics of all incumbent airlines, which, in the dynamic model treats all 6 legacy carriers separately and bins the low cost carriers together. Equilibrium transition probabilities are estimated to and from all elements of the state space. This allows for the study of the effects of airline characteristics - most notably hubs - and identities - such as being a low cost carrier - on own and rival profits and costs.

Finally, in preliminary work, motivated by the BBL approach, Benkard, Bodoh-Creed and Lazarev (2008) estimate and project simple probit entry probability functions to simulate the effect of mergers. This paper estimates similar activity probabilities, but these are used as initial values in the search for activity probabilities which represent a MPE of the full dynamic model.

[3] They have a measure of the "mean value of hub size for the incumbents".

For the demand side, Berry, Carnall and Spiller (2006), followed by Berry and Jia (2008) use a characteristic based model of demand, which is a simplification of the now canonical BLP (1995) framework to a bimodal distribution of tastes, to estimate (variable) costs and markups, defining products as unique combinations of airline-fare-itineraries. This paper specifies a pricing game at the quarterly level - consistent with the data driven time periods in the dynamic model - and thus uses a simple nested logit to estimate demand, which is then projected on the state space. The demand model used here is closer to that used by Peters (2006), who looked at data from 1985 and found static demand models to not predict post merger prices well. There are also a host of reduced form studies. For example, Borenstein (1989, 1991) finds that flights on airlines with hubs at end points command higher prices. However, recently (Borenstein 2005) he finds these premiums to have declined. Goolsbee and Syverson (2008) finds preemptive price cuts in expectation of Southwest entry (which is generally into markets in which there is a Southwest presence at one of the endpoints). However, are but a few of the many studies focusing the importance of hubs and the effects of LCC[4].

3 Data Construction

The main datasets used are two of the three datasets (merged by ticket id) from the Origin and Destination Survey (DB1B, hereinafter "the survey"), which is a 10% random sample of all domestic US tickets aggregated up to the quarter and the aggregate information in the T-100. These are public and commonly used (for example Ciliberto and Tamer (2009), Berry, 1992, and Borenstein, 1989). I use the 20 quarters from $2002 - 2006$[5]. The Coupon dataset has coupon specific information for each domestic itinerary in the survey, including the operating carrier, number of coupons, origin and destination airports, trip break code, number of passengers, fare class, and distance. Each coupon is a separate observation and represents a city pair trip (these may be pieces of the same itinerary). An itinerary is the entire trip and may contain many coupons (a round-trip contains at least 2). The Ticket dataset has the number of coupons, the origin airport, round-trip indicator, reporting carrier, a credibility indicator, the itinerary fare, the number of passengers, and distance and miles flown[6]. These are merged (by operating carrier) with the T-100 Domestic Segment Dataset, which includes all (100% of the data rather than just a sample) of domestic market data by air carriers, and origin and destination airports for passengers enplaned, including load factors, number of passengers and flights, etc.[7]. Tickets not in the T-100, or that are not provided on a regular basis (at least once a week) are dropped.

[4] See, for example, Borenstein and Rose (2007,2008) for more.

[5] Future versions of the paper will examine the effect of dropping 2002 which following September 11th, was an atypical year.

[6] The third dataset, the Market dataset, has directional market characteristics of each domestic itinerary in the survey, with a seperate observation for each market (defined as an airport pair), in the itinerary. It is not used due to inconsistencies in the market definition.

[7] There is also a T-100 Market database, which again has inconsistent definitions of a market and thus is not used. For example, a carrier change is defined as serving a different market.

For the entry decisions, a market is a (nondirectional) city pair. Airlines decide which cities to connect and in doing so are "in the market" for itineraries involving both cities as both origins and destinations. On the demand side, the products are the tickets sold based on the origin and destinations of consumers (compare Aguirregabiria and Ho 2009 that look at non-stop itineraries). The numbers of stops are a product characteristic. Time periods are a quarter as dictated by the data. I think of the airlines as supplying these products in different ways: some with more direct connections, and some with complex hub structures. The effective seller is the ticketing carrier[8].

This data has many dimensions and its reliability is not perfect. Accordingly, following the previous airline literature using this data, the sample is reduced by dropping tickets with more than two stops, multiple ticket carriers (per directional trip)[9], credibility questioned by the Department of Transportation, segments of international trips or non-contiguous domestic travel with Hawaii, Alaska and Territories; less than 120 passengers per quarter, and particularly high (over 2000 dollars) prices[10]. I keep all classes of tickets, including one-way tickets since the objective is to determine the total profitability of the route. Airports in the same MSA are joined (to reflect the competition induced by the multiple airports in a given city), and the size of the market is seen as the geometric mean of the population of the endpoint cities[11].

Airports commonly seen as hubs are coded as such. I thus have the following cities as hubs: Atlanta (Delta), Chicago (American, United), Charlotte (US), Cincinnati (Delta), Dallas (American), Denver (United), Detroit (Northwest), Houston (Continental), Memphis (Northwest), Minneapolis (Northwest), Philadelphia (US), Pittsburgh (US), Salt Lake City (Delta).

This data is also merged with a time series of jet fuel prices from the United States Department of Energy's Energy information Administration (to be used as cost shifters), aggregated to the quarterly level. As can be seen in figure 4 the spending on fuel has increased dramatically. Airlines explicitly cite fuel costs as a reason for the increase in prices and these prices do indeed work well at explaining ticket prices, as discussed below.

4 General Framework

I adapt the general structure proposed by AM 2007, which is amenable to the use of the computational techniques I employ. To set notation, assume there are N airlines $i \in I = \{1, 2...N\}$, potentially operating in M markets, where markets are combinations of the D different US cities[12]. These markets are not directional in that we assume that for the LA-Boston combi-

[8] The ticketing carrier sells the tickets. The operating carrier is determined by the airlines (not necessarily the owner of the plane or in any other fixed definition). The reporting carrier is seen as pretty meaningless by the DOT.

[9] This represents less that one percent of the data.

[10] These tickets are dropped due to suspected reporting error. In further versions I plan to explore the impact of this cut.

[11] Data were available at http://www.census.gov/popest/metro/CBSA-est2006-annual.html. Note that I only looked at markets that had at least one ticket carrier at some point in the sample.

[12] More precisely, these are metropolitan statistical area combinations, as I will explain below.

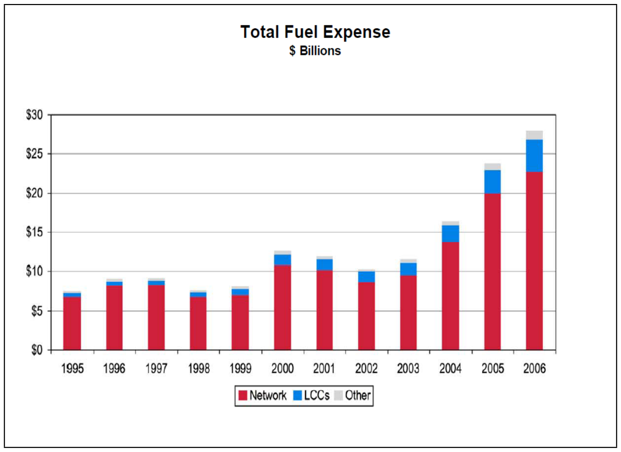


Figure 4: Spending on Fuel

nation, for example, airlines in the market sell (one way and round-trip) tickets originating and culminating in both cities. Thus for D origin and destination cities there are $M = D(D-1)/2$ markets. Airline choice sets have to be made to both reflect the data limitations, as well as to provide a tractable framework for the dynamic analysis. We assume the following timing for airlines decisions, where each quarter is a time period:

4.1 Timing

1. Each airline observes the state space at the beginning of the period (which is determined by the actions of the previous period).

2. Airlines observe their private productivity shock.

3. Airlines choose an action $a_{imt} \in \{0, 1\}$, where a 1 corresponds to being active in the market. This choice, of course, may require entry or exit, depending on the state in the previous period. Airlines know whether they enter or exit in the period, but form expectations over which of the other airlines will be in the market, given the state space of the previous period. Based on these beliefs airlines choose capacity for the market and the characteristics of the tickets that they offer, and play a price-competition game with the other airlines that chose to be in the market for the period. These latter choices will not be modeled but rather will be assumed to shape the variable profits which we estimate as a function of the evolving state space, which ensues from the entry and exit decisions.

4. Consumers choose the ticket with the characteristics that maximize their utility. Airline payoffs (the sum of the prices they collected for the tickets sold) are realized.

4.2 State Space

The state space is driven by both data limitation and the feasibility of the computation. The structure of this problem - the study of hubs and the heterogeneous effects of competition -

requires treating markets and airlines heterogeneously. However, many of the market level variables do not evolve and so the state space can be made relatively rich by allowing for one market specific state. Markets include the following variables (which remain fixed in the estimation[13]): hub variables, the nonstop distance between cities, the size of the market, the density of passengers in the market (taken from the first quarter of 2002 and held fixed throughout the sample) and whether the market is a tourist market. The variables that evolve are the number (and identity) of the incumbents in the market[14].

Given that identities matter, there are 2^N states associated with N players in a given market. There will also be N different value functions for each element of the state space in each market. The analysis of hubs requires including all 6 legacy carriers: American Airlines, United Airlines, Continental Airlines, US Airways, Delta and Northwest Airlines. The analysis of the effects of LCC requires having at least one more player. I thus chose the most parsimonious player state space with $N = 7$. Given the prominence of Southwest Airlines and, more recently, JetBlue, the seventh player is either one of these two carriers.

Following previous work, and particularly CT (2009), I order markets by the geometric mean of the city populations. I begin with all markets between the 50 largest cities (and show below that these are not a bad approximation to all US markets). This gives me a total of 1225 markets. I further eliminate 14 markets between cities that are very close geographically, leaving me with a total of 1211 markets.

As discussed below, I am thus left with $1211 \cdot 2^7 = 155,008$ states. This represents the richness (and computational burden) of the model.

4.3 Profit function

Airlines' per period profit function from all markets is:

$$\Pi_{it}(a_{it}, s_t, \varepsilon_{it}) = \sum_m \pi_{im}(s_t) - \sum_m a_{im} C_{im}(s_t, \varepsilon_{imt})$$

where π_{im} is the variable profits from each market, and $C_{im}(.)$ are the fixed and entry costs incurred by serving the market:

$$C_{im}(s_{mt}, \varepsilon_{imt}) = FC_{imt} + \varepsilon_{imt} + (1 - a_{imt-1}) EC_{imt}.$$

Note that given the timing assumption, the state space at time t, s_{mt}, represents the identity and number of firms that were in the market at time $t-1$. a_{imt} is the action taken at time t. Simple put, airlines incur an operational fixed cost if they are active, and an additional cost

[13] This assumption simply implies that the airlines *beliefs* are such that they do not expect an evolution, on average, in the future.

[14] In future verisons I plan to include the sum of the number of destinations flown by the carrier from the two connecting cities which is a deterministic function of the airlines in the market (assuming we limit it to the select number of markets).

for entering the market (i.e. becoming active after a quarter in which they were inactive)[15]. We can think of the payoff from not being in the market as $\mu_i + \varepsilon_{it}(0)$ but since this payoff is not separately identified from the fixed cost we redefine fixed costs as net of this opportunity cost. Exit costs can also not be separately identified, since identification comes from the (two) moments of entry and exit. The structural fixed and entry cost parameters are modeled as:

$$
\begin{aligned}
FC_{imt}(\boldsymbol{\gamma}) &= \gamma_1 + \gamma_{2i} + \gamma_3 X_m + \gamma_4 MyHUBS_{im} + \gamma_5 HUBS_m + \gamma_6 Legacies_i + \gamma_7 LCC_i \\
EC_{imt}(\boldsymbol{\varphi}) &= \varphi_1 + \varphi_{2i} + \varphi_3 X_m + \varphi_4 MyHUBS_{im} + \varphi_5 HUBS_m + \varphi_6 Legacies_i + \varphi_7 LCC_i.
\end{aligned}
$$

The costs are decomposed into a constant and carrier specific fixed effects, X_m - market level variables (which I begin by having as just the distance between cities), $MyHUBS$, the number of cities which are the airline's hub, $HUBS$, the number of hubs for other carriers[16], $Legacies$ - the number of (other) legacy carriers in the market, and LCC - whether a low cost carrier (here Southwest or JetBlue) is in the market (other than the airline itself, and so this is zero for the LCCs).

Finally, $\varepsilon_{it} = \{\varepsilon_{imt} : m = 1, 2...M\}$ are the private information idiosyncratic shocks incurred by each airline in each market m. We assume they are i.i.d. over airlines, markets and time[17] with an extreme value CDF G_ε.

4.4 The Dynamic Entry Game

This game has the standard markov-structure: Airlines maximize the expected present discounted value of profits, taking into account all payoff relevant variables. Denote the strategy functions by $\sigma = \{\sigma_i(s_t, \varepsilon_i), i \in I\}$. This gives a value function for each airline i over the states:

$$V_i^\sigma(s_t, \varepsilon_{it}) = \max_{a_{it}} \{\Pi_{it}(a_{it}, s_t, \varepsilon_{it}) + \beta E[V_i^\sigma(s_{t+1}, \varepsilon_{it+1}) | s_t, a_{it}]\}$$

which takes as given the strategies of the other airlines (belonging to σ) and chooses a_{it} as a best response - maximizing the expected discounted profits. The MPE (markov perfect equilibrium) implies that: $\forall \sigma_i \in \sigma$ we have:

$$\sigma_i(s_t, \varepsilon_{it}) = \arg\max_{a_{it}} \{\Pi_{it}(a_{it}, s_t, \varepsilon_{it}) + \beta E[V_i^\sigma(s_{t+1}, \varepsilon_{it+1}) | s_t, a_{it}]\}$$

[15] Natually, other definitions of inactivity could be used, exploring the possibility of seasonality in the service of some markets, etc. For entry I require an airline to active for at least two consecutive periods, following inactivity. Similarly, for exit, I require an airline to be inactive for at least two periods following activity. This is expected to alleviate some of the possible errors in the data collection, which results from an (imperfect) 10% survey.

[16] I chose to count up all hubs for other carriers to allow for a difference between having one and more than one carrier with a hub in a given city.

[17] This assumption, which is common in the dynamic game literature, may be a strong one. It is possible for certain markets or airline-market combinations to have serially correlated shocks. Such would be the case if airlines were reluctant to exit certain markets even if they were unprofitable (for example due to network benefits not captured in the model). I plan to explore this important extension in future work.

or that all airlines are best responding to each other. As in the AM framework, players' strategies depend on one another only through the conditional choice probabilities, i.e. the probabilities that airlines choose a_{imt} given the state space. These integrate the strategy functions over the private information shocks.

$$P_i^{\sigma}(a_i|s) \equiv \Pr(\sigma_i(s,\varepsilon_i) = a_i) = \int I\{\sigma_i(s,\varepsilon_i) = a_i\} g_i(\varepsilon_i) d\varepsilon_i.$$

This gives the equilibrium condition in terms of probabilities, which essentially will form the constraints in the dynamic optimization problem. In order for the conditional choice probabilities to represent an MPE they must satisfy:

$$P_i^{\sigma}(a_i|s) = \int I\{\sigma_i(s_t,\varepsilon_{it}) = \arg\max_{a_{it}}\{\Pi_{it}(a_{it}, s_t, \varepsilon_{it}) + \beta E[V_i^{\sigma}(s_{t+1}, \varepsilon_{it+1})|s_t, a_{it}]\}\} dG_{\varepsilon}(\varepsilon_{it}).$$

Equilibrium existence follows the proofs in AM and Doraszelski and Satterthwaite (2007) for any absolutely continuous (with respect to the Lebesgue measure) density function.

Airlines can and should jointly optimize their entire network. However, for tractability we make some simplifying assumptions:

First, for simplicity we treat each market separately. We assume that each regional airline manager maximizes the expected present discounted profits from each market and does not consider the private shocks or decisions that the airline makes regarding other markets. In other words, although there is some commonality in the matter in which consumers respond to airlines (for example through the fixed effect or brand effect and through the total number of destination served), and entry into one market may now enable consumers to fly between other cities that become connected in the airline's network, airline managers do not take this into account.

Second, we simplify the structure of the transition probabilities. We assume a first order markov process, where:

$$\Pr(s_{mt+1}|s_{mt}, a_{imt}, s_t; P_{-im}) = \Pr(s_{mt+1}|s_{mt,} a_{imt}; P_{-im}).$$

In other words, rather than considering the state space of all markets considered (including the airline's own state in these other markets) the payoff relevant variables are the market specific state variables. This assumption extends the previous one in that airlines do not consider the entire state space for all markets even with regards to the profits in their own market. Thus, airlines consider how the this being a hub market for themselves and for their competitors affects profits in the market, but not how the profits from being in this market are affected by the identity and characteristics of the airlines in related markets. For example, serving both LA-Boston and LA-NY may be something valued by consumers in adding to the flexibility the airline offers, and thus the profitability of entering LA-Boston may depend on whether the airline is in LA-NY, but here we assume the airline does not consider this. The conditional transition probability is thus assumed to be independent of the state space in other markets (which is the assumption we relied on above in specifying the size of the state space).

Third, following what is standard in this literature, we assume that the individual probabilities are independent conditional on the state space and so:

$$\Pr(s_{mt+1}|s_{mt,}) = \prod_{i=1}^{N} \Pr_i(a_{imt}|s_{mt}).$$

Fourth, while the model may have multiple equilibria, we assume the data are generated by one MPE, which players expect to be played into the future.

These assumptions allow for a redefinition of the equilibrium (note the addition of m subscripts), where an airline chooses to be active depending on the value function from each market iff:

$$V_{imt}(s_{mt}, 1) - V_{imt}(s_{mt}, 0) \geq 0$$

$$\Longleftrightarrow$$

$$E[\Pi_{imt}(a_{imt}, s_{mt}, \varepsilon_{it})] + \beta[E[V^P_{im,t+1}|s_{mt}, 1] - E[V^P_{im,t+1}|s_{mt}, 0]] \geq \varepsilon_{imt}.$$

The first expectation is the expected profits in the market given the state space at the beginning of the period,

$$E[\Pi_{imt}(\cdot)] = -a_{imt}(1-a_{imt-1})EC_{im}(s_{mt}) + \sum_{s'_m \in S_m} [\pi_{im}(s'_m) - FC_{im}(s'_m)]FP(s'^{*}_m|\sigma_{im}(s_{mt}, \varepsilon_{it}) = a_{imt}, s_{imt})]$$

where entry costs are incurred when an inactive firm becomes active, and the variable period profits depend on which firms decide to be active in the period. Thus, there is an expectation taken using the transition probability matrix, $FP(\cdot)$, which is a function of the true equilibrium probabilities. The second is the expectation of the value functions from next period onward (once again using the transition probability matrix).

This implies equilibrium probabilities of firm i being in market m at time t, of the form:

$$P_{im}(a_{it} = 1|s_{mt}) = G_\varepsilon(E[\Pi_{it}(a_{it}, s_t, \varepsilon_{it})|s_{m,t-1}] + \beta[E[V^P_{im,t+1}|s_{mt}, 1] - E[V^P_{im,t+1}|s_{mt}, 0]]).$$

Now, we can think of the per period profit function as:

$$\Pi_{imt} = (1 - a_{imt})[z_{imt}(0)'\theta] + a_{imt}[z_{imt}(1)'\theta + \varepsilon_{imt}]$$

where: $z_{imt}(0, s_{mt})$ is a vector of zeros, $z_{imt}(1, s_{mt}) \equiv (E[\hat{\pi}_i(s_m)|s_{mt}], E[FC_i(s_m)|s_{mt}], a_{imt}(1 - a_{imt-1})EC_i(s_{mt})$ and $\theta \equiv \{1, \boldsymbol{\gamma}, \boldsymbol{\varphi}\}$.

Denote the variance of ε by σ_ε. We then have a MPE being a vector of $P = \{P_{im}(s)\}$ such that for all (i, m, s_{imt}):

$$P_{im}(a_{imt} = 1|s_{mt}) = \Psi(\tilde{z}^P_{imt}`\frac{\theta}{\sigma_\varepsilon} + \tilde{e}^P_{imt}) \qquad \text{((EQ))}$$

where $\Psi(\cdot)$ is the extreme value CDF (and θ is all of the parameters in the model), and we have the infinite sums of:

$\tilde{z}^P_{imt} = z_{imt}(1, s_{mt}) - z_{imt}(0, s_{mt}) +$
$\sum_{j=1}^{\infty} \beta^j [$
$\sum_{s' \in S}\{E\{(P_{im}(a_{im,t+j} = 0|s'_{m,t+j}))z_{im,t+j}(0, s'_{m,t+j}) + P_{im}(a_{im,t+j} = 1|s'_{m,t+j})z_{im,t+j}(1, s'_{m,t+j})\}|a_{imt} = 1, s_{im,t+j-1}\}\} -$
$\sum_{s' \in S}\{E\{(P_{im}(a_{im,t+j} = 0|s'_{m,t+j}))z_{im,t+j}(0, s'_{m,t+j}) + P_{im}(a_{im,t+j} = 1|s'_{m,t+j})z_{im,t+j}(1, s'_{m,t+j})\}|a_{imt} = 0, s_{im,t+j-1}\}]$

$\tilde{e}^P_{imt} = \sum_{j=1}^{\infty} \beta^j [\sum_{s' \in S} E\{P_{im}(a_{im,t+j} = 1|s'_{m,t+j})(Euler - \ln P_{im}(a_{im,t+j} = 1|s'_{m,t+j}))\}|a_{imt} = 1, s_{m,t+j-1}\} -$
$\sum_{s' \in S}\{\{\{P_{im}(a_{im,t+j} = 1|s_{m,t+j})(Euler - \ln P_{im}(a_{im,t+j} = 1|s_{m,t+j}))|a_{imt} = 0, s_{m,t+j-1}\}]$.

Thus airlines compare the value of being active with that of not being active in terms of the payoff today and in all future period paths beginning at being active or not being active today respectively. Thus includes the effect of the activity status today on future entry and exit through the equilibrium probabilities. Once again, expectations are taken using the transition probability matrix which provides a probability for every possible transition from a given state. The derivation of $e^P_{imt} \equiv E(\varepsilon_{imt}(a_{imt})|s_{mt}, \sigma^*_i(s_t, \varepsilon_i) = a_i)$, the expectation of the error conditional on it rendering a_i the optimal action, using the Euler-Mascheroni constant was shown by Hotz-Miller (1987?). As is standard, the infinite sums are solved by solving a system of linear equations.

Equation (EQ) represents the conditions required for their to be an equilibrium in the model, namely that each firm is best responding to the actions of all other firms. Given the assumptions above, an equilibrium exists (see AM 2007). There are as many best response constraints as the dimension of the state space, multiplied by the number of carriers.

4.5 The Transition Probability Matrix

Note above that the transitions are the product of the individual choice probabilities. This still however leads to a prohibitively large number of transition, since we have for each market $(7 \cdot 2^7)^2$ possible transitions. After looking at the data, I find that in 99% of the data, there are no more than 2 movements (changes of status - entry or exit) per period. I thus constrain all transitions involving more than two transitions to be zero (and scale the permissible transitions accordingly). To give a sense of the data, table 2 lists the transitions (by number of players, not identities).

Table 2

Transitions

	0	1	2	3	4	5	6	7	**Total**
0	**74**	10	0	0	0	0	0	0	84
1	10	**608**	91	12	1	0	0	0	722
2	2	63	**896**	194	40	6	4	0	1205
3	0	12	128	**1275**	363	57	14	12	1851
4	0	0	6	242	**1830**	545	97	10	2730
5	0	0	0	21	345	**2848**	735	62	4011
6	0	0	0	1	25	425	**5158**	613	6222
7	0	0	0	0	1	24	336	**5823**	6184
Total	86	693	1211	1745	2605	3905	6344	6510	23009

5 Estimation

5.1 General Strategy

The estimation strategy is a result of both the complexity inherent in the airlines' optimization problem, as well as the feasibility constraints imposed by the state-of-the-art solvers available today. Over time, as solvers improve and computational power increases larger problems (and richer state spaces) will become feasible.

As mentioned, airlines are interested in maximizing profits, which are the difference between the revenues they can garner from ticket purchasing consumers and their costs. Airlines make many choices, including the choice and allocation of their fleet, the structure of their networks, and so on. For simplicity, I think of the airlines as solving this complex profit optimization problem using backwards induction. This implies that when playing the pricing game at the ticket level, airlines take capacity (as well as product characteristics and their fleet and network) as given. Consequently, at this level, the marginal costs of serving an additional customer are just the costs of filling a seat, conditional on capacity, as well as some probability of incurring a cost for adjusting capacity. These latter costs likely involve compensating consumers for overbooked flights (adding more capacity at the "last minute" is much less common) and are not likely to be very large in the overall calculus of firm costs. The difference between the revenues from an optimally chosen pricing structure and these marginal costs represent the variable profits from the route. These actual (or potential) variable profits then go into the calculus of the decision regarding the airline's activity status in each route, in forming the expected present discounted profits.

The model allows for an estimation of variable profits from the dynamic game as well. However, this would add many more parameters to be fit, and ignore the abundance of ticket level data which is available. Consequently, my approach is to estimate all parameters that can be estimated from the static game and then construct projections of these variable profits, for each market, on the state space and use them as inputs into the value function and the dynamic model. Specifically, I specify a static pricing game to obtain variable profits, and estimate them with all of the cross sectional data described above. Note that, when estimating the static game markets are directional: consumer choose from all available tickets beginning at their desired origin. I thus combine the profits from each direction in computing the total *market* profits. I then move to estimate the structural parameters of the entry and fixed costs specified above.

5.2 Variable Profits

5.2.1 Approach

In modeling the demand, choices have to be made to both reflect the data limitations, as well as to provide a tractable framework for the dynamic analysis. On the data side, the time of purchase (and, consequently, the choice set available to each consumer) is not observed; many of the ticket restrictions, which are key determinants of the price, are unobservable as well. Indeed, a disaggregate analysis, like a BLP 1995 approach (treating each ticket separately), requires identifying products as any tickets with a different price (or any other characteristic); and thus, products are rarely repeated. Each quarter will have tens of thousands of products, yielding a huge amount of product shares. In my model, entry is into a market in a particular quarter, and so, in the model, airlines consider the quarterly profits. Thus, instead of estimating the per product variable profit and aggregating to the quarterly level, I simplify the analysis by thinking of a (reduced form) game over the entire quarterly traffic. This eliminates the ability to have consumer specific coefficients (or even the simple two-type bimodal distribution in BCS and Berry and Jia 2009). It does, however, simplify the instrumental variable methodology. I define the price of a product as the average price paid by each passenger to the carrier in the market. If p_{km} is the price of each ticket sold by airline k in market m, redefine p_j as:

$$p_j = \frac{\sum_{k \in Km} p_{km}}{s_j MS},$$

where MS is the size of the market, the geometric mean of the MSAs of the endpoint cities.

5.2.2 Static Pricing Game and Demand

I use the common discrete choice framework (for example Berry 1994), where consumers have the following utility function:

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \zeta_{igt}(\sigma) + (1-\sigma)\varepsilon_{ijt}.$$

x_j is a vector of characteristics of the product and market characteristics. p_j is the product price, β and α are the vector of tastes for the product characteristics and price respectively; ξ_j are the unobserved ticket features; ε_{ijt} is an i.i.d. (across all products and consumers) logit error, ζ is common (for consumer i) to all products in group g and has a distribution function that depends on σ, $\sigma \in [0,1]$,where $\zeta_{ig}(\sigma) + (1-\sigma)\varepsilon_{ijt}$ is still distributed logit. As σ goes to 1 there is no i.i.d error (within group correlation of utility levels goes to one). With $\sigma \in (0,1)$ product shares have the common nested logit form:

$$s^i_{j/g} = \frac{\exp[(x_{jt}\beta - \alpha p_j + \xi_{jt})/(1-\sigma)]}{D_g} \tag{1}$$

where

$$D_g = \sum_{j \in J_g} \exp[(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})/(1-\sigma)].$$

Define

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

as the mean utility of product j. Assume the outside good has a zero mean utility, and so the share of consumers who decide to fly (and purchase a ticket) follows the same logit form (substituting in for the outside good) :

$$\bar{s} = s^i_{j \in J} = \frac{D_g^{(1-\sigma)}}{1 + D_g^{(1-\sigma)}}. \tag{2}$$

The (unconditional) share of product j is the product of (1) and (2):

$$s^i_j = s^i_{j/g} \cdot s^i_{j \in J} = \frac{\exp[(x_{jt}\beta - \alpha p_j + \xi_{jt})/(1-\sigma)]}{D_g^{\sigma}(1 + D_g^{(1-\sigma)})} \tag{3}$$

Specifically, the share of the outside good is:

$$s^i_0 = \frac{1}{1 + D_g^{(1-\sigma)}}.$$

This framework is convenient. Taking logs (and suppressing the index for each individual), we have

$$\ln(s_j) - \ln(s_0) = \frac{1}{1-\sigma}\delta_j - \sigma \ln(D_g)$$

and taking logs of (2) we have $\frac{\ln(\bar{s}) - \ln(s_0)}{1-\sigma} = \ln(D_g)$ and so:

$$\ln(s_j) - \ln(s_0) = \frac{1}{1-\sigma}\delta_j - \sigma\frac{\ln(\bar{s}) - \ln(s_0)}{1-\sigma}$$

and rearranging we have:

$$\begin{aligned}
\delta_j &= (1-\sigma)\ln(s_j) - \ln(s_0) + \sigma \ln(\bar{s}) \\
&= (1-\sigma)\ln(s_j) - \ln(s_0) + \sigma[\ln(s_j) - \ln(\bar{s})] \\
&= \ln(s_{j)} - \ln(s_0) - \sigma \ln(\bar{s})
\end{aligned}$$

where the second equality comes from substituting in (3). This yields the following:

$$\ln(s^i_{j)}) - \ln(s^i_0) = x_{jt}\beta - \alpha p_{jt} + \sigma \ln(\bar{s}) + \xi_{jt}. \tag{4}$$

Using this relationship, we can obtain the coefficients β, α and σ using a linear instrumental variable approach, we where know that, at minimum, p_j and $\bar{s}$ are endogenous with respect to ξ_j. Finally for the supply side of this last stage game, we can assume a standard Bertrand-Nash game and then have:

$$p_j = c_j + \frac{s_j}{\frac{\partial s_j}{\partial p_j}}$$

and so differentiating (3) we have that:

$$p_j = c_j + [\frac{\frac{(1-\sigma)}{\alpha}}{1 - \sigma\bar{s} - (1-\sigma)s_j}]$$

and so we can form

$$\hat{c}_j = p_j - [\frac{\frac{(1-\hat{\sigma})}{\hat{\alpha}}}{1 - \hat{\sigma}\bar{s} - (1-\hat{\sigma})s_j}]$$

and obtain measures of the variable profits in the market. $p_{jm} - \hat{c}_{jm}$.

We can project the marginal cost on product characteristics and estimate:

$$\hat{c}_j = w_j\gamma + \omega_j$$

where w_j are the product characteristics affecting the marginal cost, which, once again simplifies to this linear form.

Instrumental Variables are needed, as mentioned above to account for the simultaneity of the determinations of prices and quantities. A good cost shifter for the average prices are the fuel costs and so I use a 2SLS strategy of instrumenting for the price and internal share with fuel costs and, what are commonly known as the BLP instruments, the sum of the characteristics of the other products - which are the sum of the average characteristics of the other airlines in the market[18]. The latter group of instruments stem from the assumption that firms play a pricing game where the characteristics of all other products affect the prices they can charge and the overall share of consumers choosing to fly, but that airlines are not adjusting the product characteristics (and specifically the ξ_{jt}) jointly with the other players[19].

For comparison[20] I also replace these BLP instruments with the set of instruments used by Hausman (1996). The identifying assumption behind these instruments - which exploit the panel

[18] These instruments provide reasonably large first stage R-squares, and highly significant parameters.

[19] This assumption, which is questionable in many industries, is questionable here as well. It may be unreasonable to assume that airlines are not adjusting their ticket features, although this is difficult within a time period given the complexity in coordinating the entire network.

[20] In this version of the paper, I use the first version of estimates for as inputs in the dynamic game.

structure of the data - is that, given the controls, market-specific valuations are independent across markets (but potentially correlated within a market). This allows for the use of ticket prices in other markets as valid IVs. These prices are thus assumed to be correlated across markets due to common marginal costs but, given the mentioned assumption, not due to market specific valuations. All prices in all markets and all quarters could potentially be used as instruments. Following Nevo 2001, I use the average price in all markets (excluding the market being instrumented for) in a given quarter[21].

5.3 Estimation of the Dynamic Game

5.3.1 The Optimization Problem

Given the estimation of the bottom node of the game, we estimate the parameters of entry and fixed costs from the entry and exit decisions/moments. We therefore track the activity status of each airline and construct the following constrained maximum[22] pseudo likelihood estimator[23]. We maximize ($\theta \in \Theta$):

$$\tilde{L}(\Theta, P) = \sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{t=1}^{T} \mathbf{1}(a_{imt}=1)\ln(\Psi(a_{imt}|s_{mt};P,\Theta) + (1-\mathbf{1}(a_{imt}=1))\ln(1-\Psi(a_{imt}|s_{mt};P,\Theta))$$

subject to the equilibrium condition (EQ) above of:

$$P_{im}(a_{imt}=1|s^*_{imt}) = \Psi(\tilde{z}^{P}_{imt}`\frac{\theta}{\sigma_\varepsilon} + \tilde{e}^{P}_{imt}).$$

Simply put, we are maximizing the likelihood of observing the activity patterns in the data, given the model, subject to all actions representing best responses to P, or being consistent with the most likely equilibrium given the data, in the game specified above. We will call the solution to this problem the **"Full Maximum Likelihood Estimator"**.

5.3.2 Feasibility and Computational Methodology

This problem is computationally challenging. Thus, while the objective function is smooth, there is a (non-linear) equilibrium constraint for each element of the state space, for each player. There are thus $7 \cdot 1211 \cdot 2^7 = 1,085,056$ probabilities, for each of which there are all the transitions described above. The value function has to be solved for each player and each element of the state space as well, by inverting a system of linear equations, where:

$$V_{im}(s) = \sum_{a\in A} P_{im}(a)E(\Pi(s)\cdot\beta\sum_{s'\in S} V_{im}(s')TP(s,s').$$

[21] There is no claim being made here regarding the optimality of this choice of instruments. I also include the hub characterization of the other produces here, since these are, by construction, not adjustable.

[22] More precisely, it is the supremum of the maximization of the pseudo likelihood.

[23] The term "pseudo" comes from these probabilities not necessary representing the equilibrium probabilities, but rather best responses to an arbitrary vector P. See below and in AM for more on this.

This estimator is consistent, and efficient (see AM). To ease this computation burden, previous work has simplified the state space by assuming that players are symmetric and by discretizing the state space. AM note that "this estimator can be impractical if the dimension of P is relatively large....this is the case in models with heterogeneous players...even when the number of players is not too large". It is not possible to assume symmetry here, since the purpose of this study is to understand the differential effects that the differing network features have on the players themselves, and the heterogenous effects they have on their actual and potential competition.

When the population probabilities P^0 are known, the equilibrium constraints are not needed and the estimator is root-M consistent. When a $\sqrt{M}$ non-parametric estimator of $P^0, \hat{P}^0$ is available (as is the case for example with a frequency estimator or a kernel method, when there are no unobservable market characteristics), estimates of θ, resulting from the maximization of the likelihood: $\tilde{L}(\Theta, \hat{P}^0)$, or from what we will call the **"Two Step Estimator"**, are consistent as well (see full details in AM 2007). It is difficult to determine and establish consistency of the estimators of the probability methods in most applications. The use of market fixed effects when feasible, is helpful, but there still could of course be market-time specific unobservables corrupting the estimates. AM propose a Nested Pseudo Likelihood method: The prescription set forth by this methodology is that (potentially non-consistent) estimates of $\hat{P}^0_k$ are formed, $\hat{\theta}_k$s are obtained from the maximization of $\tilde{L}(\Theta, \hat{P}^0_k)$, $\hat{P}^0_{k+1}(\hat{\theta}_k)$ are formed, using the equilibrium constraints with $\hat{\theta}_k$, new $\hat{\theta}_{k+1}$ are formed from the maximization of $\tilde{L}(\Theta, \hat{P}^0_{k+1})$ and so on until:

$$(\hat{P}^0_{k+t} - \hat{P}^0_{k+t-1}) \leq r,$$

where r represents the stopping rule. This sequence is well defined when there is a unique value of θ that maximizes the pseudo-likelihood function for each value of P, which is assumed in all applications using this method. When this sequence of $\{\hat{P}^0_k, \hat{\theta}_k\}$ converges (if it converges), its limit represents the maximum of the constrained problem. This is what they call the "**Nested Fixed Point**" estimator. In the Monte-Carlo examples presented in the AM paper convergence is acheived. Interestingly, the "two-step" estimates provided very similar results, suggesting that in any case, the estimated probabilities should be used in initializing the solution algorithm for this problem.

To ease the computational burden, I follow the MPEC (Mathematical Programming with Equilibrium Constraints) approach, advocated by Judd and Su (2008) and Dube, Fox and Su (2009). The MPEC structure of the problem essentially relies on the "augmented likelihood function", $\pounds(\theta; \sigma, X)$, presented above, which explicitly expresses the dependence of the likelihood on σ. θ and σ do not need to be consistent with the conditions of the equilibrium of the model[24]; however, when adding the equilibrium conditions as constraints, the solution to the problem will be a θ which represents the most likely equilibrium. Given this formulation, I can

[24] Compare to the discussion of the "pseudo" maximum likelihood above.

use solvers which rely on quadratically convergent constrained optimization methods, based on Newton's method (see Schmedders 2008 for a review of optimization methods)[25]. The solvers do not solve a fixed point, or require the specification of an algorithm for solving the equilibrium conditions, and the augmented likelihood uses single valued functions. Furthermore, the constraints need only to be solved at the point of the optimal solution; an LU decomposition is computed, and backsolving is used (rather than inverting matrices); derivatives are computed using automatic differentiation[26], which "eliminate[s] this as a serious problem"[27]; and the sparsity in the Jacobians and Hessians is exploited.

I begin by estimating smooth functions for the activity probabilities. For each player, I estimate the logit probability:

$$P_i(a, s_m) = G(X\beta)$$

where $G(\cdot)$ is the standard logistic CDF. X includes the activity status of each of the players in the previous period, a quadratic function of the distance, the geometric mean of the population, the passenger density, whether it is a tourist market, the number of hubs for the airline itself in the market, and the number of hubs for the other legacies. I cluster by market to flexibly allow for serial correlation[28]. The results are presented in table 5 below. As can be seen, the strongest predictor of being active in a market is last period's activity status, reflecting the stability of decisions over time and potentially the large role for entry and fixed costs. The other coefficients are consistent with the variable profit results presented below. Note that these are merely activity probabilities and not structural parameters. However, these results can be compared to Benkard et al (2008) who use similar probit probabilities to simulate the effects of Delta-Northwest merger and find them to have much predictive power[29].

I use the predicted probabilities generated by these regression to form the initial probabilities for all elements of the state space in every market. In other words, this initializes the values of all of the equilibrium constraint probabilities discussed above. In addition, as is standard, β is treated as a parameter. Given that these are quarters, I chose $\beta = 0.98$.

Finally, I chose the knitro solver, which is one of the most powerful solvers today, designed to handle linear and nonlinear problems with dimensions running into the hundreds of thousands. It has the versatility of three different algorithms which it can choose between, including direct and conjugate gradient interior point methods, as well an active set algorithm to rapidly solve binding constraints using linear programming. The main advantage of the AMPL language is that, once the problem is transformed into the form above, the code is straightforward and

[25] This as opposed to Guass-Siedel methods, commonly used in past work, which have at best linear convergence, and (even local) convergence for which is difficult to prove with nonlinear equations.

[26] This refers to methods that computed analytic gradients and Hessians efficiently and use the chain rule of differentiation to build a sequence of simple operations. Languages such as AMPL and GAMS aid in computing these gradients and Hessians using insights incorporated in symbolic software.

[27] See Judd and Su 2008.

[28] In future versions I plan to explore more flexible specifications of these activity probabilities.

[29] I note that one of the key predictors, especially for Southwest entry was the amount of passenger traffic that could be added to the network. This variable could be tracked in further versions of the paper, along with the total number of destinations from the cities discussed above.

the communication with a multitude of available solvers is made easy. AMPL prepares the problem for the solver, and, when the presolve option is used, it transforms the problem into an equivalent smaller problem which is easier to solve. It removes unnecessary constraints and applies useful transformations. However, AMPL, can require a large amount of memory and even the most advanced solvers are not without limitations. In this application, the richness of the model is largely dependent on the size of the state space that can be estimated, where each element of the state space essentially adds an equilibrium constraint for each player. For the full model specified here, there are over a million constraints. This is generally above the limit of what can be done with the best solvers and so the problem has to be estimated in parts (or, equivalently, the number of markets that can be used in the estimation has to be reduced). Furthermore, both AMPL and the solver demand a lot of RAM memory. A 64 bit operating system (and AMPL version) is required to process a problem using over 4 GB of RAM, which in my experience is the case for most non-trivial applications. The largest server I was able to use has 65 GB of (shared RAM), and approximately another 30 GB of (slower) swap memory. These represent the computational constraints with which I was faced[30].

The difference between the full maximum likelihood and the two step estimator, given this formulation, hinges on whether the initial estimates are treated as starting values, or are held fixed in the estimation[31]. Note that even for the two step estimator, each player's value functions have to be solved for, using the transition probabilities (as functions of the initial values). I have found that 50 market can solved for in a matter of minutes, while 300 markets requires about a day. For the full model, even 50 markets can take a matter of days. To get initial estimates I estimated the markets in groups of 50 and in groups of 300 (while averaging the scaled coefficients). The results vary between groups and between the averages of the groups.

6 Results

This section collects the results of all stages of estimation. We begin by discussing the results from the static demand: the demand parameters and the marginal costs. We then discuss the projection of these parameters on the state space, which feed into the value function. Finally we move to discuss the parameters of the entry costs and fixed costs from the estimation of the full dynamic game.

6.1 Demand Parameters

In order to track the impact of the necessary simplifications of the state space, I present several specifications, using all relevant (and available) variables. My approach is to specify as rich a demand model as possible and then project the resulting profits onto as rich a state space

[30] These were the limits when writing the first draft in June of 2009. I hope and expect these limit to soon be seen as laughable.

[31] In the code, the difference between the two-step and the full maximum likelihoos estimates is two comment characters.

as possible. Summary statistics for all filters considered (all markets, all markets between the 100 largest MSAs, and all markets between the 50 largest MSAs) are presented in table 1. The results are presented in tables 3(a) and 3(b). Once again I present the results for the three filters considered, which suggest that the limits made on the subset of markets are reasonable.

I cluster by market to allow for serial correlation, but, as can be seen, all coefficients are highly significant and of the expected sign. I include fixed effects for all major carriers (the six legacy carriers, and six of the biggest LCC including Southwest, Jetblue, Frontier, AirTran, Spirit Airlines and ATA Airlines), and for each time period.

The demand is downward sloping in price. Distance (measured in miles) is positive but non monotone, reflecting an inverted U-like relationship where air travel becomes more attractive as it crowds out other modes of travel, but at further distances travel is needed less and is consumed less. The number of destination cities may reflect more convenient gate access and expertise as well as flexibility in rerouting passengers and thus is positive. The tourist dummy captures travel to or from Las Vegas or Florida and fits the high level of traffic to these cities not captured by the other variables in the model. Travellers prefer less stops and more direct flights. They also prefer travel to and from the hub of the ticket carrier they are using, and have a negative preference for airlines other than the hub airline at hub airports. Within this model, the airline dummies are positive and large for the main LCC: JetBlue and Southwest ("B6" and "WN" respectively), but not for the legacies, capturing the features of the LCC service not captured by the relatively parsimonious specifications possible with the data. Finally, as expected, travelers prefer round-trip tickets. To get a sense of the monetary value of the characteristics it is useful to use the value of a marginal dollar of price as a scale. Doing this suggests for example that on average, passengers would pay \$143 dollars to travel an extra 1000 miles, and \$241 less for a connecting flight. They are willing to pay \$53 extra to travel with a carrier that has a hub at their destination airport, and about \$55 more for a tourist spot. In comparison, when using the Hausman instruments in table 3(b), the magnitudes change: on average, passengers would pay \$500 to travel another 1000 miles, and \$957 less for a connecting flight. The hub premium increases as well to \$93.5 dollars, as does that for tourist destinations.

6.2 Marginal Costs

For the marginal costs recall that these are those in the pricing equation, where price is defined as the average price for a given airline's product in a given market and a given quarter. Here, we find the tourist variable coefficient to be extremely low. This likely reflects much lower prices for these markets, which is likely due to the higher elasticity of travel to these destinations. In other words this is a control for the parsimonious model which assumes the same Bertrand-Nash pricing game in all markets. The other coefficients are as expected. The more round-trip tickets, the higher the cost. Similarly, a larger number of connections increases the cost (per ticket), because there is more travel. Note that this is the final level costs. The choice of more connections is helpful in the aggregate analysis of the airline, since it eases constraints on the rest of the network. The more speciality in terms of destinations from the origin, the lower the cost. Costs follow an inverted U shape with respect to distance where there are savings in costs at much higher distances. The large coefficients on the hub variables are surprising and may, once again, caution the simplicity of this analysis, but suggest that costs are higher at hub airports to the hub carriers themselves. As expected, all LCC have much lower marginal costs. In other specifications, the number of markets is reduced. This helps reduce the size of the data, but also, potentially, allows for a more homogenous group of markets. This homogeneity has advantages in fitting the (parsimonious) model, but omits some of the information embedded in markets in which at least one of the endpoints is in a small MSA. Popular vacation spots would be an obvious example of this. However the second and third specifications, corresponding to CT and excluding all markets not between cities in the largest 100 MSA (55% of the markets) and 50 MSAs (15% of the markets, for a total of 1211) respectively, have roughly similar results. Similarly, the results using the alternative set of instruments in table 3(b) are qualatatively similar[32].

The average number of connections and the percent of round-trips is likely to be endogenously determined, and thus the assumption made in these specifications is that product characteristics are fixed at the beginning of the period. However, specifications without these variables yield similar results. These variables are not part of the state space for the dynamic game and are used to get the best fit for the variable profits where are used below. Taken as a whole, these results can be related to the two main questions of the paper, hubs, and LCC. Regarding hubs, the results from the static estimation suggest that while consumers prefer flying with hubs or with airlines offering more connections from the origin, more generally, they have a very high distaste for flights with connections (stops). Thus, the benefits from having a large hub presence comes at a high cost if indeed this requires many more connecting flights. It is difficult to comment on the magnitude without accounting for the benefits of the whole network, but these results do suggest the potential for segregation by different airlines in offering different products that meet either the flexibility features, or the nonstop features respectively. As to the LCC, it is clear that consumers have a high preference for the LCC brand and especially for Jetblue and Southwest. This suggest that there are some unobservable characteristics of their products which consumers like. Marginal costs are also lower for these airlines and overall variable profits are thus higher. The increased number of nonstop flights offered by these carriers increases consumers' willingness to pay as well.

6.3 Structural Profit Parameters

Now, for the purposes of the dynamic model we need a simple projection of the variable profits (revenues minus marginal costs), estimated above, on the state space. The results are presented

[32] As mentioned, there are some differences in magnitudes which I plan to explore. There is also a difference in sign in the effect of the HubDest parameter on the marginal cost, but these estimates were not of the same sign in table 3(a) either.

in table 4 below. As can be seen the results from this parsimonious regression reflect those from the full (static) variable profit estimation. The fixed effects are negative for all airlines, relative to the omitted category of Southwest and JetBlue. The distance follows a U shape, the population, passenger densities and additional profits earned from tourist markets are all positive. Variable profits are higher in the airlines' own hub markets, but lower in markets in which their competitors have hubs. The LCC, Delta and Northwest all have significant negative effects on their rivals, when they are in the market. Interestingly, the strategic effects of the players on rival profits suggest that Southwest and Jetblue significantly and to a large degree, reduce the profits of their competitors[33].

The constant is 6.7 million dollars of quarterly variable profits. This reflects the baseline profit of the omitted category (the LCC). Baseline profits (fixed effects) are lower for all other carriers (relative to the omitted category) and are generally three to four million dollars. The variables distance, passenger density and population are scaled down (and so coefficients are per 1000 miles, passengers or residents respectively).

As mentioned above, these are seen to be the structural parameters of variable profits. Thus, in the estimation of the dynamic game we project these parameters on the state space of every market in each time period and input these projected profits into the value function. This both exploits the DB1B and T-100 data available, and reduces the burden on the estimation of the dynamic game below[34].

6.4 Entry Costs and Fixed Costs

I present the results (from the estimation of 300 market blocks) in table 6 below. Column 1 represents the average from all groups, while columns 2 and 3 represent the averages from the first two and second two blocks of 600 markets respectively. As can be seen, the estimates are of the expected magnitude, but vary between blocks of markets. Markets are ordered by the geometric mean of the city populations (not by the passenger traffic or profitability), but this likely still introduces some systematic differences (and so perhaps choosing the subset of markets randomly may be better). The preliminary results here suggest that entry costs are the equivalent (for the LCC omitted category) of slightly less than 3 years of variable profits, and so can be quite high, but not surprising. Fixed costs are about 60% of the variable profits, and so are considerable as well. Both costs vary by player, where American, Northwest and United seem to have lower costs, while the costs for Continental, Delta and US Airways are higher. Distance (measured here and in the profit projections in thousands of miles) is negative for some of the blocks, which is unlikely. This result seems to be sensitive (it was not the case when I averaged over market blocks of fifty markets). Thus, this issue may be alleviated by added a quadratic in distance, as in the static results, as well as increasing the size of the blocks

[33] This projection, although clearly much more parsimonious than the estimation of variable profits above, has a relatively high R-square of 40%.

[34] In future versions I plan to explore more flexible specifications of these projections, given the large sample size.

(both of which I plan to explore). Entry costs are much lower in markets in which the carrier has hubs, and interestingly (although to a smaller degree) for hub markets more generally. Entry costs are higher the more legacy carriers incumbent in the market, and, to a smaller degree when there are LCC incumbents. Fixed costs, are higher in hub markets. This echoes the results of the marginal costs being higher in airlines' own hub markets as well. If this is indeed the case this results suggest that the added benefits of hubs are smaller than what is commonly perceived. In the specification presented here legacy carriers and especially LCC, being in the market, raises fixed costs[35]. These results were also not stable across blocks and are puzzling. I will revisit these as well, as the scale of the estimation increases.

Returning the our motivation, we find evidence that while hubs increase consumers willingness to pay, as do an increased number of destinations from the origin airport, they come with considerable added marginal and fixed costs which may outweight the benefits even in the hub markets themselves. The increase in fixed costs is higher for the hub airline in its own hub airports than for competitors. The additional connections required are both more costly to the airlines and also can garner less from consumers, especially when competitors offer nonstop flights. Similarly, I do not find that hub airports are more expensive for competitors to enter. Taken as a whole, these results do not paint an optimistic picture for the use of major hub airports.

As mentioned, low cost carriers are found to be more profitable in the static demand and to significantly reduce the variable profits of the legacy airlines. In the estimates in table 4[36] we find that on average the presence of a LCC in the market is associated with a reduction of over $\$660,000$. These effects are much smaller than the effects of legacy carriers on themselves and on the LCC. These airlines are more appealing to consumers and to have lower marginal costs. However, surprisingly, LCC impose less of a deterrent to the entry of legacy carriers in the airports they serve, as compared to that which the legacy carriers impose on other legacy carriers and on the LCC. In addition, the results presented here do not show an impact of LCC persence in a given market on the fixed costs of the other incumbents. Thus, while there are huge benefits to LCC in both desirability and costs, future work is needed to understand the source of the convergence in operating procedures and fares shown above, and the to explore the potential for the LCC benefits to be emulated by other carriers.

There are many caveats to these preliminary estimates. Firstly, the structure of the model is such that, conditional on the parameters of the model, all randomness comes from the extreme value errors. This is a serious limitation with which these types of dynamic models are faced. From a computational point of view the solver has to fit the σ and the entry and fixed costs. Entry and exit can come from either the draws of the error or these costs. This is likely a reason for the high sensitivity of the results. It may (and does for some of the blocks) even produce negative costs. Furthermore, as mentioned, we cannot identify exit costs separately and thus

[35] This may represent, for example, higher advertising and spending in these markets to combat the LCC presence.

[36] There results are from the projections of the variable profits on the state space as discussed above.

the fixed costs are essentially net of exit costs (which are not estimated), as well. Secondly, larger blocks need to be estimated. I am currently estimating the markets in two blocks, but ideally they should be estimated together. Thirdly, the full maximum likelihood estimates (for the full number of markets) should be used. Unfortunately, these take time, and will likely only be feasible for smaller blocks of markets. In their absence, the accuracy of the results hinges on the consistency of the initial probability estimates and their small sample bias, and so richer specifications of these initial probabilities should be explored. Fourthly, correct standard errors need to be obtained. Given the complexity in deriving analytical standard error for this multi-step process, a bootstrap methodology is more feasible. This methodology requires drawing from the data (with replacement), accounting for market clusters, and completing all stages of the estimation for each draw[37]. This is time intensive and somewhat difficult (and thus often not done in the applied dynamic game papers), given that solvers can, in some instances fail, and may need to be run several times to find an optimum. These caveats will hopefully be addressed in future versions of the paper.

7 Conclusion and Future Work

In this paper I applied a dynamic entry game model to the complex airline industry, in an effort to recover market specific profitability and its determinants. Specifically, I focused on the desirability of hubs and the strategic effects of the heterogeneous players. The model used allowed for the exploitation of the benefits of state-of-the-art mathematical solvers and optimization software, which have recently been strongly promoted.

The results in this paper have important implications: I found that while hubs offer benefits that consumers desire and higher profits to hub airlines in their own hub markets, this network may also come at a cost in only offering many more flights with connections which are much less desirable to consumers. In addition, the preliminary estimates of fixed costs suggest that costs may be higher in the hubs as well further dampening their desirability and that they do not deter entry. As expected, Southwest and JetBlue are more profitable than legacy carriers, and their brands offer benefits not captured by the limited product characteristics in the data. Their presence in the market imposes particularly large strategic effects for the other carriers, in lowering variable profits and in raising fixed costs. More work is needed to understand the nature of the impact of LCC on their rivals and the potential for their advantages to be emulated.

As mentioned throughout, while this paper takes a further step in employing and enriching what can be done with dynamic models, there are important limitations both to the overall use of dynamic entry games in this application and in the preliminary results presented above. The estimation can be improved by increasing the number of markets in each block and in the full estimation, but memory limits and the computational capacity of solvers will inevitably constrain

[37] There are many distributions to think about. For example, when projecting the profit parameters we take the point estimates, even though these can lie in large regions and are sometimes insignificant.

the size of the state space - and, consequently, the richness of the model that can be estimated, and thus final judgement will have to be made in the interpretation of the results. First and foremost, the simplifications which essentially abstract out of the joint network optimization made by airlines may indeed be deemed unreasonable.

The bridge between applied econometric work and state-of-the-art computational software is an important one. Economic models in general and dynamic models in particular, test the boundary of these tools and greatly benefit from increases in computational power. Any flexibility granted by using better methods, offers more room to develop better representative economic models.

The extensions to this paper are immediate. More has to be done to ensure that we indeed at the limit in capturing the richest profit function possible. As more of the ticket transactions move online, better data is also becoming available. Then, with this profit function at hand, many counterfactual experiments may be estimated. For example, the preliminary work by Benkard et al, regarding the simulation of mergers can be extended to exploit the full structural model. Given that identity specific value functions are estimated, specific mergers can be explored, once choices are made regarding how we view the new merged entity. I plan to explore some of these extensions in future work.

References

[1] Aguirregabiria, V., and P. Mira (2007), "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, $1 - 53$.

[2] Aguirregabiria, Victor; Ho, Chun-Yu (2009), "A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments, mimeo.

[3] Bajari, P., L. Benkard and J. Levin (2007), "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, $1331 - 1370$.

[4] Benkard, C. Lanier, Aaron Bodoh-Creed and John Lazarev (2008), "Simulating the Dynamic Effects of Horizontal Mergers: U.S. Airlines", mimeo.

[5] Beresteanu, Arie and Paul B. Ellickson (2006), "The Dynamics of Retail Oligopoly", mimeo.

[6] Berry, S. (1992), "Estimation of a model of entry in the airline industry", *Econometrica*, 60(4), $889 - 917$.

[7] Berry, S. (1994), "Estimating Discrete-choice Models of Product Differentiation" *Rand Journal of Economics*, 25, $242 - 262$.

[8] Berry, S, J. Levinsohn, and A. Pakes (1995) (BLP), "Automobile Prices in Market Equilibrium." *Econometrica*, 63, $841 - 890$.

[9] Berry, S., M. Carnall, and P. Spiller (2006), “Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity,” in *Advances in Airline Economics*, *Vol. 1: Competition Policy and Antitrust*, (D. Lee, ed.) Elsevier Press. 183 – 214.

[10] Berry, Steven, and Panle Jia (2009), "Tracing the Woes: An Empirical Analysis of the Airline Industry", mimeo.

[11] Boguslaski, C. H. Ito, and D. Lee (2004), "Entry Patterns in the Southwest Airlines Route System" Review of Industrial Organization, 25(3), 317 – 350.

[12] Borenstein, S. (1989), “Hubs and high fares: Dominance and market power in the US Airline industry”, *Rand Journal of Economics*, 20(3), 344 – 365.

[13] Borenstein, S. (1991), "The Dominant Firm-Advantage in Multiproduct Industries: Evidence from the U.S. Airlines", *Quarterly Journal of Economics*, 106(4), 1237 – 1266.

[14] Borenstein, S. (1992), "The Evolution of U.S. Airline Competition." Journal of Economic Perspectives, Vol. 6.

[15] Borenstein, Severin (2005), "U.S. Domestic Airline Pricing, 1995 – 2004", *U.C. Berkeley Competition Policy Center Working Paper*, CPC05 – 48.

[16] Borenstein, Severin, and Nancy Rose (2007), National Bureau of Economic Research working paper #13452, forthcoming in N. Rose ed., *Economic Regulation and Its Reform: What Have We Learned?*, University of Chicago Press.

[17] Borenstein, Severin, and Nancy Rose (2008), forthcoming in N. Rose ed., *Economic Regulation and Its Reform: What Have We Learned? (chapter 2)*, University of Chicago Press.

[18] Bresnahan, T., and P. Reiss (1993), “Measuring the importance of sunk costs”, *Annales D’Economie Et De Statistique*, 31, 181 – 217.

[19] Chernozhukov, V. H. Hong and E. Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica,* 75(5), 1243 – 1284.

[20] Ciliberto, F. and E. Tamer (2009) (CF), "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, forthcoming.

[21] Collard-Wexler, Allan, (2005), "Plant Turnover and Demand Fluctuations in the Ready-Mix Concrete Industry", mimeo.

[22] Doraszelski, U., and Satterthwaite, M. (2007), “Computable Markov-Perfect Industry Dynamics: Existence, Purification, and Multiplicity,” manuscript. Department of Economics. Harvard University.

[23] Dube, Jean-Pierre H., Jeremy T. Fox, and Che-Lin Su, (2009), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficient Demand Estimation", mimeo.

[24] Evens, W.N. and I.N. Kessides (1993), "Localized Market Power in the U.S. Airline Industry", The Review of Economic Studies, 62, $53-82$.

[25] Fourer, R. D.M. Gay, and B.W. Kernighan (2003), *AMPL: A Modeling Language for Mathematical Programming*, Brooks/Cole - Thomson Learning, 2nd Edition.

[26] Gayle, K. (2004), "Does Price Matter? Price or Non-Price Competition in the Airline Industry, mimeo, Department of Economics, Kansas State University.

[27] Goolsbee, Austan and Chad Syverson, "How do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines", *Quarterly Journal of Economics* , 123(4), $1611-1633$.

[28] Hausman, Jerry (1996), "Valuation for New Goods Under Perfect and Imperfect Competition", in *The Economics of New Goods*, Studies in Income and Wealth Vol. 58, ed. by T. Bresnahan and R. Gordon. Chicago: National Bureau of Economic Research.

[29] Hendricks, K. M. Piccione, and G. Tan (1995), "The Economics of Hubs: The Case of Monopoly", *The Review of Economics and Statistics,* $62(1), 83-99$.

[30] Hendricks, K., M. Piccione, and G. Tan (1997), "Entry and exit in hub-spoke networks," *Rand Journal of Economics*, 28(2), $291-303$.

[31] Hendricks, K., M. Piccione, and G. Tan (1999), "Equilibria in Networks," *Econometrica*, 67, $1407-1434$.

[32] Hotz, V. Joseph, and Robert A. Miller, (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, $61(2), 265-289$.

[33] Judd, Kenneth L., and Che-Lin Su, (2008), "Constrained Optimization Approaches to Estimation of Structural Models", mimeo.

[34] Januszewski, J. and M. Lederman (2003), "Entry Patterns of Low-Cost Airlines ," manuscript, Rotman School of Management, University of Toronto.

[35] Kasahara, H., and K. Shimotsu (2008), "Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," *Journal of Econometrics*, 146, $92-106$.

[36] Lederman, M. (2008), "Are Frequent-Flyer Programs a Cause of the Hub Premium?" *Journal of Economics and Management Strategy*, 17, $35-66$.

[37] Nevo, Aviv, (2000), "A Practitioner's Guide to Estimation of Random Coefficient Logit Models of Demand", *Journal of Economics and Management Strategy,* 9(4),$513-548$.

[38] Nevo, Aviv, (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69(2), $307-342$.

[39] Macieira, Joao, (2006), "Measuring the Effects of Innovation Subsidies: The Case of Supercomputers", mimeo.

[40] Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *Rand Journal of Economics*, 38(2), $373-399$.

[41] Peters, Craig, (2006), "Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry", *Journal of Law and Economics, XLIX,* $627-649$.

[42] Reiss, P.C. and P.T. Spiller, (1989), "Competition and Entry in Small Airline Markets", *Journal of Law and Economics*, 32, $179-202$.

[43] Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, 55, $999-1033$.

[44] Ryan, Stephen, (2009), "The Costs of environmental Regulation in a Concentrated Industry", mimeo.

[45] Schmedders, K. (2008), "Numerical Optimization Methods in Economics. In S.N. Durlauf and L.E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd Edition, Palgrave Macmillian.

[46] Sweeting, Andrew, (2007), "Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry", *NBER Working Paper* 13522.

8 Tables:

Table 1: Summary statistics

Variable	Mean	Std.	N	Mean	Std.	N	Mean	Std.	N
Price	362.442	108.73	814066	352.05	105.176	612319	331.017	95.827	282724
Fuel	126.411	43.375	814066	125.89	43.38	612319	124.98	43.419	282724
Roundtrips	0.891	0.112	814066	0.885	0.115	612319	0.872	0.121	282724
Connections	0.911	0.315	814066	0.888	0.334	612319	0.806	0.391	282724
Destinations	63.077	26.504	814066	64.107	26.129	612319	65.579	26.485	282724
Frommyhub	0.055	0.228	814066	0.058	0.233	612319	0.076	0.265	282724
Tomyhub	0.053	0.224	814066	0.057	0.233	612319	0.076	0.265	282724
Distance	1148.841	621.705	813925	1181.412	630.388	612302	1271.623	666.848	282724

These are the main variables used in the demand estimation. The first three columns are for all markets, the second three is for the first filter which keeps the 100 largest MSAs and the third are for the main sample used which keeps the markets between the 50 largest MSAs. As can be seen, most of the variables have roughly the same summary statistics.

Table 3(a)

Utility Function and Marginal Cost Parameters (BLP Instruments)

	allmkts	filter-1	filter-2	mc-1	mc-2	mc-3
	(1)	(2)	(3)	(1)	(2)	(3)
Price	-.007***	-.007***	-.006***	-	-	-
	(.0002)	(.0002)	(.0003)			
Lwshare	.279***	.274***	.249***	-	-	-
	(.011)	(.013)	(.022)			
Tourist	.390***	.397***	.406***	-53.043***	-53.712***	-55.712***
	(.023)	(.025)	(.032)	(2.369)	(2.587)	(3.263)
HubOrigin	-.190***	-.195***	-.148***	5.511***	3.784***	9.523***
	(.015)	(.018)	(.026)	(1.305)	(1.425)	(1.894)
HubDest	-.066***	-.059***	-.093***	-2.997**	-1.223	8.373***
	(.014)	(.016)	(.024)	(1.259)	(1.386)	(1.880)
Roundtrips	1.300***	1.436***	1.799***	21.871***	25.323***	-.681
	(.045)	(.055)	(.103)	(2.965)	(3.374)	(5.805)
Connections	-1.687***	-1.783***	-1.958***	35.050***	33.224***	23.240***
	(.024)	(.029)	(.049)	(1.611)	(1.745)	(2.166)
Destinations	.012***	.013***	.016***	-.382***	-.395***	-.226***
	(.0002)	(.0003)	(.0005)	(.023)	(.029)	(.049)
Frommyhub	.368***	.382***	.356***	57.569***	58.998***	53.962***
	(.019)	(.022)	(.031)	(1.593)	(1.780)	(2.331)
Tomyhub	.719***	.754***	.792***	50.874***	52.347***	52.465***
	(.019)	(.021)	(.031)	(1.553)	(1.700)	(2.174)
Distance	.001***	.001***	.001***	.073***	.066***	.078***
	(.00004)	(.00004)	(.00007)	(.004)	(.005)	(.007)
Distancesq	-3.35e-07***	-3.45e-07***	-3.32e-07***	-6.59e-06***	-3.49e-06**	-8.74e-06***
	(1.30e-08)	(1.49e-08)	(2.19e-08)	(1.49e-06)	(1.70e-06)	(2.39e-06)
AA	-.273***	-.306***	-.342***	70.115***	62.321***	44.618***
	(.030)	(.033)	(.051)	(2.108)	(2.412)	(3.970)
CO	-.430***	-.516***	-.604***	67.590***	59.979***	45.512***
	(.030)	(.033)	(.049)	(2.041)	(2.312)	(3.735)
DL	-.281***	-.344***	-.534***	92.407***	85.830***	52.335***
	(.031)	(.035)	(.054)	(2.386)	(2.875)	(4.922)
NW	-.524***	-.610***	-.803***	63.022***	55.912***	39.528***
	(.029)	(.033)	(.050)	(2.105)	(2.459)	(4.130)
US	-.184***	-.221***	-.241***	50.863***	46.671***	31.646***
	(.029)	(.032)	(.049)	(2.059)	(2.359)	(3.884)
UA	-.298***	-.359***	-.464***	74.510***	72.253***	56.097***
	(.030)	(.034)	(.050)	(2.069)	(2.414)	(3.979)
B6	.475***	.516***	.656***	-37.678***	-38.238***	-36.205***
	(.066)	(.067)	(.085)	(3.453)	(3.574)	(4.881)
F9	-.134***	-.106***	.129**	-18.519***	-16.865***	-18.183***
	(.034)	(.036)	(.050)	(1.994)	(2.144)	(3.061)
FL	.180***	.308***	.603***	-48.095***	-46.653***	-51.839***
	(.037)	(.041)	(.067)	(1.961)	(2.188)	(3.365)
NK	.699***	.533***	.622***	-74.194***	-69.661***	-79.750***
	(.102)	(.098)	(.109)	(3.667)	(3.689)	(4.604)
TZ	-.052	.024	.387***	-27.351***	-27.376***	-26.417***
	(.037)	(.040)	(.058)	(1.950)	(2.163)	(3.287)
WN	.323***	.292***	.398***	-14.173***	-13.392***	-17.443***
	(.030)	(.033)	(.052)	(1.895)	(2.110)	(3.249)
Obs.	644516	485915	224977	644516	485915	224977

Table 3(b)

Utility Function and Marginal Cost Parameters (Hausman Instruments)

	all-h	f-1-h	f-2-h	mc-a_all	mc-1	mc-2
	(1)	(2)	(3)	(1)	(2)	(3)
price	$-.002^{***}$	$-.002^{***}$	$-.003^{***}$	–	–	–
	(.0002)	(.0002)	(.0003)			
Lwshare	$.176^{***}$	$.210^{***}$	$.295^{***}$	–	–	–
	(.012)	(.016)	(.034)			
Tourist	$.615^{***}$	$.611^{***}$	$.593^{***}$	-53.840^{***}	-54.015^{***}	-54.923^{***}
	(.026)	(.030)	(.041)	(2.308)	(2.499)	(3.059)
HubOrigin	$-.287^{***}$	$-.253^{***}$	$-.130^{***}$	10.545^{***}	9.976^{***}	14.486^{***}
	(.016)	(.021)	(.032)	(1.270)	(1.385)	(1.839)
HubgDest	$-.130^{***}$	$-.095^{***}$	$-.074^{**}$	2.566^{**}	4.880^{***}	13.047^{***}
	(.016)	(.020)	(.031)	(1.235)	(1.354)	(1.830)
Roundtrips	1.361^{***}	1.418^{***}	1.639^{***}	11.933^{***}	13.520^{***}	-13.102^{**}
	(.046)	(.061)	(.124)	(2.894)	(3.298)	(5.703)
Connections	-1.914^{***}	-1.963^{***}	-1.912^{***}	43.043^{***}	45.085^{***}	42.256^{***}
	(.024)	(.029)	(.064)	(1.585)	(1.706)	(2.132)
Destinations	$.013^{***}$	$.014^{***}$	$.016^{***}$	$-.344^{***}$	$-.397^{***}$	$-.281^{***}$
	(.0002)	(.0002)	(.0004)	(.022)	(.027)	(.047)
Frommyhub	$.187^{***}$	$.173^{***}$	$.128^{**}$	50.633^{***}	51.221^{***}	44.369^{***}
	(.025)	(.031)	(.050)	(1.543)	(1.714)	(2.230)
Tomyhub	$.564^{***}$	$.573^{***}$	$.559^{***}$	45.080^{***}	44.653^{***}	41.707^{***}
	(.023)	(.029)	(.053)	(1.506)	(1.639)	(2.078)
Distance	$.001^{***}$	$.001^{***}$	$.001^{***}$	$.080^{***}$	$.073^{***}$	$.083^{***}$
	(.00004)	(.00005)	(.00008)	(.004)	(.004)	(.007)
Distancesq	$-3.11e-07^{***}$	$-3.38e-07^{***}$	$-3.05e-07^{***}$	$-8.30e-06^{***}$	$-5.59e-06^{***}$	$-1.00e-05^{***}$
	(1.34e-08)	(1.56e-08)	(2.39e-08)	(1.44e-06)	(1.63e-06)	(2.31e-06)
AA	$-.520^{***}$	$-.539^{***}$	$-.511^{***}$	65.537^{***}	60.223^{***}	45.528^{***}
	(.039)	(.042)	(.057)	(2.100)	(2.394)	(3.934)
CO	$-.712^{***}$	$-.765^{***}$	$-.748^{***}$	65.945^{***}	61.072^{***}	48.187^{***}
	(.038)	(.040)	(.050)	(2.044)	(2.311)	(3.735)
DL	$-.608^{***}$	$-.671^{***}$	$-.733^{***}$	82.527^{***}	77.059^{***}	51.221^{***}
	(.043)	(.048)	(.062)	(2.355)	(2.813)	(4.814)
NW	$-.758^{***}$	$-.829^{***}$	$-.934^{***}$	58.774^{***}	53.782^{***}	40.776^{***}
	(.038)	(.040)	(.051)	(2.096)	(2.435)	(4.081)
US	$-.355^{***}$	$-.391^{***}$	$-.372^{***}$	47.578^{***}	44.840^{***}	31.465^{***}
	(.036)	(.039)	(.054)	(2.057)	(2.347)	(3.835)
UA	$-.585^{***}$	$-.642^{***}$	$-.646^{***}$	72.230^{***}	72.443^{***}	60.410^{***}
	(.039)	(.042)	(.053)	(2.065)	(2.394)	(3.938)
B6	$.713^{***}$	$.718^{***}$	$.688^{***}$	-39.982^{***}	-41.295^{***}	-39.302^{***}
	(.075)	(.074)	(.092)	(3.535)	(3.674)	(4.974)
F9	$-.067^{*}$	$-.041$	$.153^{***}$	-17.870^{***}	-16.637^{***}	-19.340^{***}
	(.035)	(.035)	(.044)	(1.988)	(2.149)	(3.098)
FL	$.478^{***}$	$.551^{***}$	$.676^{***}$	-50.835^{***}	-49.898^{***}	-55.232^{***}
	(.034)	(.037)	(.063)	(1.987)	(2.220)	(3.386)
NK	1.106^{***}	$.880^{***}$	$.815^{***}$	-79.987^{***}	-69.467^{***}	-74.902^{***}
	(.098)	(.093)	(.105)	(4.447)	(3.567)	(4.511)
TZ	$.112^{***}$	$.162^{***}$	$.423^{***}$	-28.098^{***}	-28.795^{***}	-28.003^{***}
	(.036)	(.038)	(.051)	(1.969)	(2.189)	(3.301)
WN	$.470^{***}$	$.391^{***}$	$.360^{***}$	-19.491^{***}	-18.594^{***}	-25.226^{***}
	(.031)	(.034)	(.061)	(1.943)	(2.170)	(3.348)
Obs.	644476	485875	224937	644516	485915	224977

Table 4

Projection of the Variable Profits on the State Space

Variable Profit Projection	
Constant	6694214.000***
	(855010.000)
AA	-3630095.000***
	(508058.800)
CO	-4367321.000***
	(446053.300)
DL	-3546622.000***
	(471656.600)
NW	-4471127.000***
	(467761.500)
US	-4486935.000***
	(549021.800)
UA	-4848132.000***
	(519141.500)
Distance	-2409303.000***
	(443088.200)
Distancesq	743346.000***
	(139236.800)
Population	454.471***
	(160.204)
Passenger Density	53355.020***
	(6386.904)
Tourist	1128930.000**
	(527532.800)
OwnHubs	6107603.000***
	(385589.100)
OtherHubs	-1627747.000***
	(212550.800)
AA in Market	163744.400
	(265317.300)
CO in Market	-156505.000
	(207198.200)
DL in Market	-1336960.000***
	(293430.900)
NW in Market	-798468.300***
	(188964.000)
US in Market	212201.300*
	(125126.900)
UA in Market	206435.800
	(173316.800)
SWJB in Market	-662852.400***
	(136145.000)
Obs.	127800

Table 5

Estimates of Predicted Activity Probabilities

	AA	CO	DL	NW	US	UA	SWJB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	−5.192***	−4.017***	−3.747***	−3.410***	−3.769***	−3.798***	−4.085***
	(.228)	(.192)	(.157)	(.160)	(.162)	(.167)	(.170)
LastAA	**4.271***	.609***	.086	.409***	−1.166***	.398***	.197*
	(.153)	(.092)	(.095)	(.094)	(.107)	(.105)	(.120)
LastCO	.659***	**3.351***	−.009	.436***	.069	−.241***	.059
	(.100)	(.103)	(.087)	(.086)	(.089)	(.081)	(.097)
LastDL	.140	−.088	**4.184***	.575***	.072	−.460***	.169
	(.113)	(.106)	(.110)	(.093)	(.105)	(.100)	(.114)
LastNW	.325***	.471***	.590***	**4.074***	.142	.226***	−.449***
	(.109)	(.095)	(.084)	(.107)	(.092)	(.085)	(.111)
LastUS	−1.101***	.345***	.208**	.232***	**5.466***	.940***	.282***
	(.109)	(.082)	(.091)	(.086)	(.125)	(.090)	(.094)
LastUA	.414***	−.231***	−.421***	.052	.632***	**4.244***	.263**
	(.097)	(.082)	(.089)	(.082)	(.087)	(.104)	(.102)
LastSWJB	−.075	.261***	.133	−.161**	.017	−.251***	**7.489***
	(.100)	(.071)	(.084)	(.070)	(.067)	(.081)	(.168)
Distance	4.691***	2.970***	3.674***	1.492***	1.543***	3.124***	.430
	(.372)	(.307)	(.285)	(.291)	(.285)	(.315)	(.268)
Distancesq	−1.213***	−.834***	−1.057***	−.384***	−.413***	−.770***	−.008
	(.150)	(.110)	(.100)	(.104)	(.098)	(.114)	(.089)
Population	.0004***	.0002***	.00009*	−.00003	.0004***	.0001**	.0001***
	(.00005)	(.00008)	(.00005)	(.00004)	(.00005)	(.00005)	(.00004)
Pdensity	.002	−.003***	−.0006	−.0001	.004***	.0007	.004***
	(.001)	(.001)	(.0006)	(.0006)	(.001)	(.001)	(.001)
Tourist	1.347***	.406*	−.120	−.402*	.500*	.297	.484**
	(.338)	(.220)	(.219)	(.223)	(.301)	(.273)	(.208)
Hubs-AA	1.540***	–	–	–	–	–	–
	(.258)						
Hubs-CO	–	2.545***	–	–	–	–	–
		(.334)					
Hubs-DL	–	–	1.416***	–	–	–	–
			(.180)				
Hubs-NW	–	–	–	2.776***	–	–	–
				(.409)			
Hubs-US	–	–	–	–	.927***	–	–
					(.158)		
Hubs-UA	–	–	–	–	–	1.095***	–
						(.129)	
Totalhubs-oAA	−.215***	–	–	–	–	–	–
	(.083)						
Totalhubs-oCO	–	−.497***	–	–	–	–	–
		(.074)					
Totalhubs-oDL	–	–	.075	–	–	–	–
			(.077)				
Totalhubs-oNW	–	–	–	.039	–	–	–
				(.074)			
Totalhubs-oUS	–	–	–	–	−.187***	–	–
					(.067)		
Totalhubs-oUA	–	–	–	–	–	.116*	–
						(.069)	
Total Hubs							−.434***
							(.058)
e(N)	23009	23009	23009	23009	23009	23009	23009

Table 6

Estimates of Fixed and Entry Costs

Type of Costs:	Entry (1)	Entry (2)	Entry (3)	Fixed (1)	Fixed (2)	Fixed (3)
Constant	81.5	140	22.7	4.17	8.13	.222
AA	−30.8	−32.8	−28.9	−9.48	−20.3	1.29
CO	3.87	28.4	−20.6	−10.2	−21.1	.636
DL	.441	17.2	−16.3	−9.43	−19.5	.668
NW	−87.2	−164	−10	−7.08	−14.6	.398
US	41.3	116	−33.3	−11.3	−23.9	1.35
UA	−37.0	−54	−19.9	−10.4	−21.4	.624
Distance	−22.5	−46.8	1.8	−.05	−1.05	.047
My Hubs	−33.7	−70.1	2.78	8.75	17.9	−.37
Other Hubs	−15.1	−2.27	−7.51	.095	.415	−.225
Legacies	4.15	8.09	.219	.445	.942	−.052
LCC	2.12	2.62	1.61	4.46	9.88	−.952

This table represents the parameters of the fixed and entry costs. Column 1 is an average over all four groups of markets. Column 2 is an average over the first two groups and column 3 is an average over the last two groups. Estimates are in the millions of dollars.