The London School of Economics and Political Science

Does Collusion among Full-service Airlines Affect the Entry Decision of Low-cost Carriers?

Extended Essay for EC427—The Economics of Industry

Candidate Number: 64653

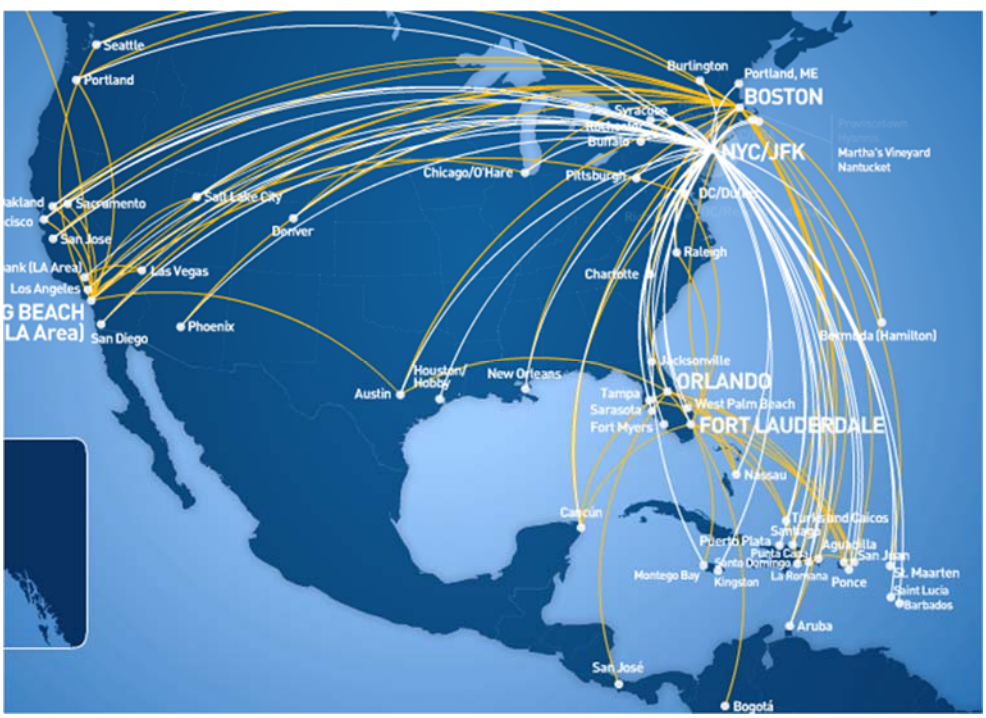


Image Source: JetBlue Airlines

Abstract

I use a cross-section regression to study whether the anti-competitive behavior of incumbent full-service airlines (code-sharing, joining alliance, and merging) affects the entry decision of low-cost carriers. I apply the idea of entry threshold ratio in Bresnahan and Reiss (1991) to analyze the competitiveness of U.S. aviation market, as well as whether new entrants face high entry barriers. I tailor the model in Mazzeo (2002) to examine the effects of product differentiation and incumbent collusion on heterogeneous airlines' profitability and market structure. The results indicate that incumbent collusion reduces the odds of airline entry. US air market is not perfectly competitive and newcomers face high entry barriers. Baseline preference for full-service airlines is high in all markets, and outweighs the effect of demand conditions.

Acknowledgements

I am truly grateful to Dr. Pasquale Schiraldi for his patient guidance and invaluable suggestions which are indispensable for the timely completion of the essay. I myself am responsible for all the errors.

Table of Contents

1. Introduction
2. Background and Literature Review
3. Data, Route and Airline Selections, and Definitions
4. Theoretical Entry Model and Regression
5. Results of the Regression and Interpretation
6. Application of Bresnahan and Reiss (1991) in analyzing incumbent conducts and predicting market equilibria
7. Adaptation of the Empirical Framework in Mazzeo (2002) and Maximum Likelihood Estimation
8. Conclusion and Potential Ways to Improve
9. Appendix
10. References

1. Introduction

When a low-cost carrier (LCC) decides upon whether to enter a new flight route, it takes account of several factors, of which the existing behavior and expected reactions from the full-service airlines (FSA) will be of great importance. FSAs have several options at their disposal to deter LCCs from entering the market, which involve code-sharing, forming alliances with partner airlines[1], and price predation. Such anti-competitive behavior of the established FSA's might greatly diminish the expected profit of a prospective LCC entrant, thereby lowering the probability of entry.

However, it is possible that full-service airlines' effort to prevent a LCC from entering might have very little or no bearing on its entry decision, because FSA and LCC are targeting different groups of customers. LCCs tailor their products to price-sensitive customers like students and budget travelers, who make purchasing decisions mainly based on prices. In contrast, customers who buy tickets from FSAs care more about service quality and on-time performance; business flyers and wealthy travelers are usually willing to pay a premium for a comfortable flight with fewer delays and hassles. Although consumers are free to purchase tickets from any airline companies, and overlaps exist between the price-conscious and quality-sensitive groups, in general the collusive behavior of FSAs (code-sharing, forming alliances, and predatory pricing) will not significantly harm the core customer base of the LCCs. In other words, LCCs are offering a new combination of price and service to the aviation market, and

[1] "Forming alliances" includes 1) joining an aviation alliance and 2) merging with airlines that currently operate in the market to become a more influential player.

this product differentiation effect allows LCCs to enter a route without suffering fierce competition from incumbent FSAs.

This paper is dedicated to study which of the two arguments above is substantiated by the data. Besides regression analyses, I apply the concept of "entry threshold ratio" proposed in Bresnahan and Reiss (1991) and use the obtained ratios to infer whether anti-competitive behavior exist among airlines operating in a route. In addition, Mazzeo (2002) develops a model to assess the impact of product differentiation on market configuration and firm profitability. By imitating his methodology, I devise a framework to evaluate the impact of collusion and the differential effects of heterogeneous incumbents on an LCC's entry decision.

The rest of the paper is structured as follows: in section 2 I review some classic papers on airline entry and the key literatures whose models are exploited in my essay. Section 3 contains the description of data sources, the essential definitions, and the construction of sample. The empirical model is presented in section 4 and the regression results are analyzed in section 5. Section 6 employs B&R's entry ratio model to study incumbent conduct and entrants' influence on competition. Section 7 tailors Mazzeo's model to probe competitors' impact on firm payoffs and market structure. Section 8 concludes and discusses ways to improve the paper. Regression outcomes, as well as facts on airline mergers and code-sharing, are reported in the appendix.

2. Background and Literature Review

Previous study by Goetz and Shapiro (2012) scrutinizes the code-sharing conduct among incumbent airlines as a response to the threat of entry by a low-cost competitor. Using a fixed-effect, lagged-time linear probability model to study the threat's effect on incumbents' code-sharing decisions, they found that the likelihood of code-sharing with partner airlines increases significantly when there is a possibility of LCC entrance. Nevertheless, their analysis utilizes the linear estimation model to get rid of the fixed effects, resulting in a probability space greater than [0, 1]. Moreover, the authors focus on the codeshare decision-making of the incumbent airlines, and it would be beneficial to complement their research by evaluating the entry decision of the low-cost carriers. To avoid reverse causality, I treat the airline entry pattern as a two-stage sequential game, assuming the incumbent FSAs have entered in stage 1 and concentrating on the 2^{nd} stage entry contest by the LCCs, given the presence and behaviors of the FSAs.

Boguslaski, Ito, and Lee (2004) focus on the dynamics of Southwest Airlines' entry strategies in the 1990s. They estimate the size of a federal legislation's unfavorable effect on SW's expansion scheme,[2] and discuss the impact of certain exogenous factors[3] which affect SW's entry pattern. Building upon their study, my research explores the competition impact of FSA's actual behaviors (allying, code-sharing, and presence), and by directly characterizing the endogenous actions from the FSAs, a more concrete and convincing argument can be made regarding what factors affect the route-expansion decision of a LCC.

Bresnahan and Reiss (1991) develop the concept of entry threshold to study entries' effects on incumbents' conducts in markets where one couldn't directly observe prices and

[2] The Wright and Shelby Amendment which limited the number of nonstop flights from DAL airport to certain destinations.
[3] Exogenous factors involve market features like distances and market size, city characteristics like population and income, and information on competition environment such as hub presence and concentration indices.

costs.[4] By examining the relationship between market size (S) and the number of firms supported by this market (N), they find that incumbents' profits and the price level fall sharply with the entry of the second and third firms, and later entrants have little effect on prices and profits. Using data on population and number of airlines operating in a market, I will estimate the market size of different routes and construct entry threshold ratios for the US aviation market, and infer from the ratios whether the incumbent airlines are guilty of deterring entry.

Mazzeo (2002) designs an empirical framework to predict the entry and product supply decisions by heterogeneous firms in a market equilibrium. Because the model allows different types of firms to have distinct effects on firm profitability, it can be applied in the context of airline entry to explore whether the presence of FSAs and LCCs in a route has disparate effect on the profits of new LCC entrants. If empirical evidence shows that an existing LCC has stronger influence on the entry decision of an incoming LCC than an extant FSA does, then it can be reasoned that the entrant competes by offering a differentiated product from the FSA, but not from other LCCs. Mazzeo's model enables quantitative analysis of how product differentiation lessens competition between airlines of different types, and in my paper, I adapt it to investigate the effect of product differentiation and collusion on the profitability of different kinds of airlines. Furthermore, the sequential game presumption can be elaborated by allowing incumbent airlines in stage one to anticipate the behaviors of subsequent entrants, thereby fitting the entry game into a classical Stackelberg model.

3. Data, Route and Airline Selections, and Definitions

[4] Entry threshold is the minimum market size required to support a given number of firms. It is calculated using the zero-profit equilibrium level of demand.

One aim of this paper is to empirically test the impact of full-service airlines' collusive behaviors on an LCC's decision to enter a given flight segment, and as a result it is necessary to find data on FSA's alliance affiliation, code-sharing information, and merging decisions. The U.S. Bureau of Transportation Statistics DB1B Origin and Destination Survey is a collection of 10% of all domestic flight itineraries every quarter. Variables include ticketing and operating carriers for each route, number of passengers on the same itinerary, fare paid, and other route-level characteristics. The raw dataset extends across several dimensions: time, flight segments, airline carriers, etc., and I am only using the flight segment dimension to conduct a cross-sectional study.

I define each domestic flight route as a potential market for entry, and observe in each quarter, in how many routes does a low-cost carrier operate. Since I have insufficient data on the most popular U.S. domestic routes each year, I use the T-100DM dataset to rank the popularity of U.S. domestic airports, based on the number of passengers traveled through the airport every quarter.[5] I limit the sample to the top 45 popular airports, and consider all the nonstop segments among these airports to be potential routes for a LCC to enter.[6] This practice generates 1703 airport-pairs, or potential markets for entry.[7] For all these routes, I consider the entry decisions of 3 major US low-cost carriers—JetBlue, Frontier, and Spirit Airlines. In addition, the DoT dataset also provides statistics on flight distance, number of airlines operating

[5] The bureau only provides data on the top 10 domestic routes (in terms of passengers transported) in 2016, so the alternative method to produce the markets of interest is utilized. See 7) in reference for source information.
[6] Excluding the airports that are serving the same metropolitan area, like JFK and LGA which both serve New York city.
[7] By stating 1703 airport-pairs, I treat TPA-SFO and SFO-TPA as two different pairs.

in a market, and information on hub and focus cities, which are all necessary for the empirical analysis.

To proxy for the market size of different routes, the Pop variable is constructed, using the sum of the population in the metropolitan areas served by the two endpoint airports. The US Census Bureau provides statistics on metropolitan population estimates and changes, and in this essay, data on 2016 is utilized.

4. Theoretical Entry Model and Regression

The LCC entry contest can be modeled in a sequential or a simultaneous game setting. For simplicity, section 4 and 5 assume that full-service airlines have entered a flight market prior to a low-cost carrier does, and the focus of analysis is on the second stage of a sequential entry game where the LCC makes entry decision and the incumbents do nothing. Moreover, under a multi-time-horizon, dynamic game setting, a firm will enter a market if the net present value (NPV) of its expected stream of profits is greater than zero. Since there is not enough information to calculate the NPV of expected profits, I only look at one-period entry decisions by different LCCs.

Because the dependent variable in the regression is a binary (Entry=1 and Non-entry=0), I adopt the nonlinear estimation models (probit) to assess the regressors' impact on each LCC's decision to enter. An airline i will enter a flight segment j if the expected profit from serving this market is positive, or the expected variable profit is greater than the entry cost: $E(\pi_{ij}) - C_{ij} > 0$. The reduced form expression for the expected profit is $E(\pi_{ij}) - C_{ij} = X_{ij}\beta + \varepsilon_{ij}$, where X_{ij} are the explanatory variables and ε_{ij} is an i.i.d. error term.[8] Although neither the

[8] I don't have time t in my subscripts, because I use the dataset to conduct a cross-sectional regression.

expected profit nor the entry cost is directly observable, the actual entry decision of a LCC is observable:

$$e_{ij} = \begin{cases} 1, \text{ if } E(\pi_{ij}) - C_{ij} \geq 0 \text{ and the firm chooses to enter} \\ \\ 0, \text{ if } E(\pi_{ij}) - C_{ij} < 0 \text{ and the firm chooses not to enter} \end{cases}$$

Therefore I estimate the probability of entry by a LCC on a flight segment using a probit model based on the equation: $Prob\ (e_{ij} = 1 \mid X_{ij}) = Prob\ (\varepsilon_{ij} > -X_{ij}\beta)$.

The X_{ij} matrix contains all the explanatory variables, a detailed description of which can be found in the appendix. Here I summarize how the key variables are constructed. Specifically, the regressors are categorized into three types:

- route characteristics (population at endpoint cities, and nonstop flight distance);
- features of the entrants and the incumbents (the hub dummies), and
- the incumbent behavior variables

The first two types serve as controls, and the last category includes the variables of real interest. The incumbent behavior variables consist of a characteristic of the airlines already operating in a flight segment (i.e. the total number of rival operators on a route), and their potentially anti-competitive actions (the ally dummy). The ally dummy stands for conduct like code-sharing and joining an airline alliance: since neither participating in an alliance nor code-sharing with a partner airline happens frequently, I use the following rules to construct the ally dummy: if the full-service airline operating on a flight segment joins an alliance, codeshares with partner airlines, or merges with a partner airline, then this ally variable will have a value of 1; if none of the three actions takes place, then the ally dummy will equal 0. In addition, although price predation is a viable option for incumbent FSAs to deter entry, prices depend upon several different factors (time to buy, purchasing site, bulk fares, award travel, etc.), and it is difficult for FSAs to use price to prey on the entrants. Hence the price variable is not included in this empirical model. The regression model ends up looking like this:

$$P(e_{ij} = 1) = \beta_1 Pop + \beta_2 Dist + \beta_3 Hub + \beta_4 NumRival + \beta_5 Ally + \varepsilon_{ij}$$

5. Results of the Regression and Interpretation

Table 1-3 in appendix 7 report the regression results of JetBlue (B6), Frontier (F9), and Spirit Airlines (NK). It is shown that the ally variable has statistically significant, negative impact on the entry decision of JetBlue (β_5=-0.352; z=-2.36) and Spirit (β_5=-0.241, z=-2.06), but not on Frontier Airlines. In comparison, the number of incumbent operators (numrival) has a weakly significant, negative impact on Frontier's entry decision (β_4=-0.106; z=-1.645), but not on JetBlue and Spirit. Although all three airlines are low-cost carriers, they are distinct in terms of company characteristics,[9] and each company might respond differently to the strategies pursued by rival airlines. Nevertheless, the statistical evidence shows that in general, the features and actions of incumbent airlines do impact the entry decision of low cost carriers.

The empirical outcomes demonstrate that LCCs are prone to incumbent FSAs' anti-competitive behaviors, even though the LCCs are nominally offering a differentiated service from their FSA rivals. One feasible explanation for this phenomenon is that the FSAs are diversifying their product offerings and providing no-frill services like what the LCCs have been doing. Since 2016, the major airlines in the United States (DL, AA, and UA) have successively

[9] JetBlue, Frontier, and Spirit are different from one another in terms of market shares, number of market served, and number of passengers carried annually. See 6) in reference.

introduced their own “basic economy” products, which offer fewer benefits than ordinary economy class service—a move by the FSAs to match and compete with the LCCs in the market and protect their own market shares. In other words, FSAs diversify their service offerings to compete with LCCs and maintain their dominance in domestic market.

The hub dummies in the regression capture an important feature of all the airlines operating domestically in the U.S.: since most U.S. airlines use the hub-and-spoke paradigm[10] to arrange their air traffic, I anticipate that having a hub of its own at one of the two endpoint airports will enhance the LCC’s odds of entry, while having a rival’s hub will diminish the chance of entry. It is evidenced that having an own hub will significantly boost all three low-cost carriers’ probability of entry, but instead of having negative impact, the influence of rivals’ hubs on LCCs’ entry probability seems to be positive, whether significant or not. Possible interpretation of this result is that several of JetBlue, Spirit, and Frontier’s hubs overlap with those of the three major airlines’, and consequently, the effect of hubs on entry decisions might be biased.[11] Nevertheless, the effect of hubs on entry is not what the paper is focusing on, as Sinclair (1995) has shown that hub-and-spoke networks are significant determinants for route entry decisions.

Further examination of the routes served by the three LCCs reveals that they are not intentionally avoiding the flight segments which are already served by FSAs. Although the anti-competitive behaviors from incumbent airlines do have negative impact on LCCs’ entry decisions, LCCs are still inclined to enter markets where FSAs are currently servicing. On the one

[10] Hub-and-spoke paradigm: a network structure in which all traffics move along spokes linked to the hub at the center
[11] See appendix for the list of hubs

hand, most U.S. carriers adopt the hub-and-spoke model in arranging their flights to achieve economies of scale and greater route efficiency. Since LCCs and FSAs sometimes share the same airport as their own hub,[12] it is possible that both a LCC and a FSA operate flights out of the same hub to an identical, popular destination. On the other hand, a route already served by FSAs might be a mature market with stable demand or promising growth prospect, and the entering LCC is attempting to supply a product which is further differentiated from what the incumbents have already offered.

6. Application of Bresnahan and Reiss (1991) in analyzing incumbent conducts and predicting market equilibria

❖ 6.1 Firm Conduct Identification using Entry Threshold Ratios

Bresnahan and Reiss create the idea of entry threshold by first proposing the concept of "zero-profit equilibrium level of demand": by assuming profit is zero in market equilibrium, BR rewrites the profit function to express market size S:

$$\pi(S_N) = (P_N - AVC) \times d \times S_N - F_N$$

$$S_N = \frac{F_N}{(P_N - AVC) \times d} = \frac{F_N}{V_N * d}$$

, where S_N is the market size that supports exactly N firms, V_N stands for the Nth entrant's variable profits (price minus average variable costs), F_N is the Nth firm's fixed costs, and d represents individual demand.[13] The ratios of two market sizes can be written as

$$\frac{s_{N+1}}{s_N} = \frac{V_N}{V_{N+1}} \frac{F_{N+1}}{F_N}$$

[12] See appendix 2 for hub information.
[13] B and R assume representative consumer, meaning that an individual consumer's demand is constant and identical across markets.

, and as N increases, the successive entry threshold ratios will fluctuate.

The entry ratio concept can be applied in my paper to deduce the conducts of incumbent airlines in a market. Traditional oligopoly theory states that higher market demand (S) attracts entrants (N), and as entry occurs, fierce competition will bring down variable profits, and entry ratios will decrease to 1. By observing to what value the sequence of ratios converges, inference on incumbents' conducts can be made. For instance, if the sequence converges to 1 and fixed cost is assumed to be constant for all, then it could be argued that additional entries do not affect levels of variable profit per customer, and the firms in the market might be either collusively sustaining a cartel or engaging in perfect competition. If the sequence converges to a level greater than 1, then it is either $\frac{V_N}{V_{N+1}} > 1$ or $\frac{F_{N+1}}{F_N} > 1$. In the former case, the market is not yet saturated, and additional entry will further bring down variable profits; the latter case might result from entrants using inefficient production technologies, or the incumbent firms are deterring competition by creating entry barriers, thereby raising F_{N+1} above F_N. Since the U.S. aviation market is constantly evolving, it is difficult to conclude whether a route is saturated or not. In section 6.1, I will treat entry ratios above 1 as potential evidence for $\frac{F_{N+1}}{F_N} > 1$, and use it to support the conclusion that the incumbents are deterring competition.

Of the 1703 observations in my dataset, the number of airlines operating in a market (N) ranges from 0 to 6. To model market size, population of metropolitan areas are used. The average size of markets with N firms are denoted as $\overline{S_N}$, and the entry threshold ratios are reported in the table below:

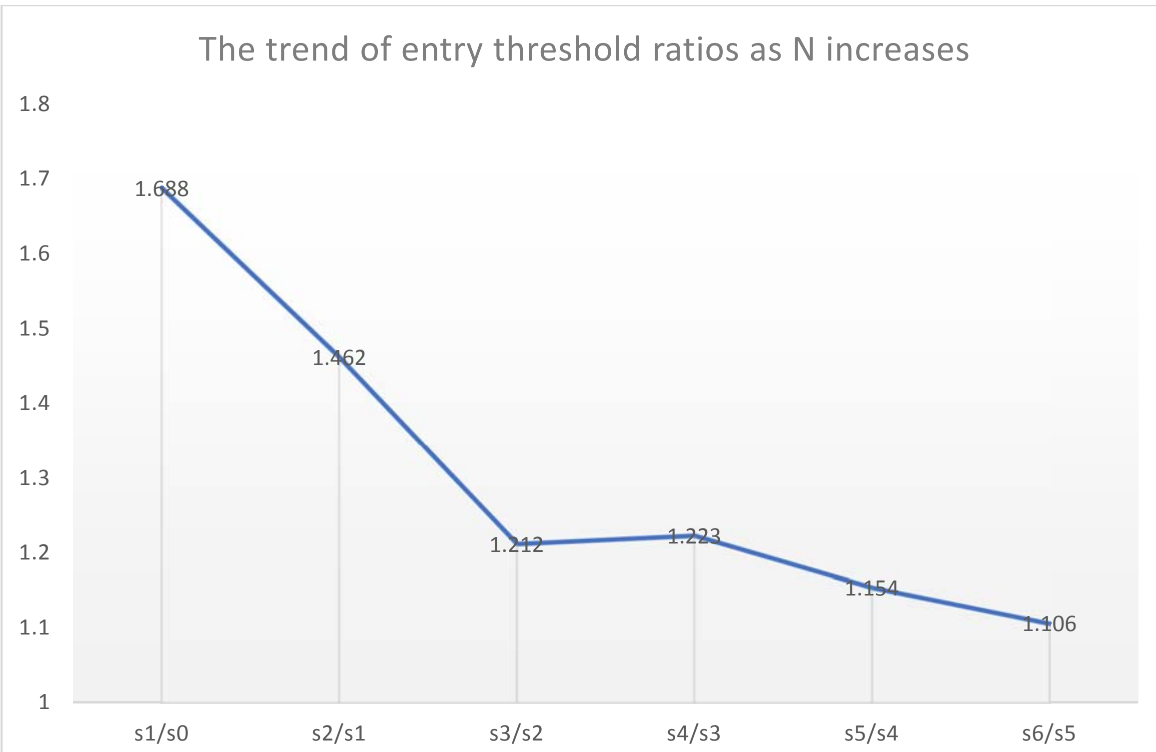
	$\overline{s_0}$	$\overline{s_1}$	$\overline{s_2}$	$\overline{s_3}$	$\overline{s_4}$	$\overline{s_5}$	$\overline{s_6}$
Average market size	7791K	9155K	12393K	14020K	15160K	13445K	14135K
		$\frac{\overline{s_1}}{\overline{s_0}}$	$\frac{\overline{s_2}}{\overline{s_1}}$	$\frac{\overline{s_3}}{\overline{s_2}}$	$\frac{\overline{s_4}}{\overline{s_3}}$	$\frac{\overline{s_5}}{\overline{s_4}}$	$\frac{\overline{s_6}}{\overline{s_5}}$
The ratio		1.175	1.354	1.131	1.081	0.887	1.051

The ratios didn't exhibit monotonicity. One possible reason is that different airlines operate different frequency of flights each day on a certain market, and we cannot assume that the number of airlines strictly increases with market size.[14] In other words, since airlines are rarely constrained by capacity, one company can monopolize a huge market by providing many flights a day. This is different from the case discussed in Bresnahan and Reiss (1991), where a single doctor or plumber can only meet the demand of a certain number of people.

To remedy the case, I combine the airports serving the same metropolitan region into one market, as one airline can hardly monopolize all the airports in a metropolitan area.[15] Owing to the adjustment, the 45 airports in the original sample is transformed into 34 "endpoints", as 9 MSAs are served by more than one airports.[16] The flights among the 34 endpoints are studied, and a market is redefined as route between two endpoints, with the size of each market revised correspondingly.

[14] For instance, United Airlines monopolize the EWR-LAS market, supplying all the 7 daily flights. The flight segment has high market demand, as it connects the Greater New York MSA with Las Vegas. According to Bresnahan and Reiss more firms should enter, but here we have UA monopolization.
[15] The Greater New York MSA is served by 3 airports in the sample: JFK, EWR, and LGA. For flights from NY to Las Vegas, United monopolizes on EWR-LAS, but JFK-LAS and LGA-LAS are served by 4 other airlines including AA, DL, VX, and B6.
[16] The NY-NJ MSA and the Washington DC MSA are served by 3 airports each.

	$\overline{s_0}$	$\overline{s_1}$	$\overline{s_2}$	$\overline{s_3}$	$\overline{s_4}$	$\overline{s_5}$	$\overline{s_6}$
Avg mkt size	5021K	8475K	12393K	15020K	18364K	21199K	23436K
		$\frac{\overline{s_1}}{\overline{s_0}}$	$\frac{\overline{s_2}}{\overline{s_1}}$	$\frac{\overline{s_3}}{\overline{s_2}}$	$\frac{\overline{s_4}}{\overline{s_3}}$	$\frac{\overline{s_5}}{\overline{s_4}}$	$\frac{\overline{s_6}}{\overline{s_5}}$
The ratio		1.688	1.462	1.212	1.223	1.154	1.106



In the new arrangement, the entry ratios generally manifest monotonicity, and the minor increase from s3/s2 to s4/s3 can be regarded as potential evidence that F4 is higher than F3 and entry barriers are high (recall that s4/s3 = v3/v4 * F4/F3). In addition, the sequence gradually converges to a value above 1.1, indicating that either the variable profit is still above competitive level, or entry cost is higher for new entrants. Both scenarios suggest that the U.S. aviation market is not highly competitive—profit margins still exist, or newcomers face entry hindrance from the incumbents. Furthermore, the sharp decrease from s1/s0 to s3/s2 is consistent with Bresnahan and Reiss's conclusion that post-entry competition intensifies at a rate that drops with the number of entrants; the small change from s4/s3 to s6/s5 implies that later entrants have relatively small influence on incumbent conducts.

Granted, some presumptions adopted by Bresnahan and Reiss might not be compatible with the airline market: the belief of homogeneous consumer makes it hard to model individual difference between budget-conscious and quality-sensitive travelers, and the hypothesis of zero-profit equilibrium is too strong: airlines might be earning positive profits in equilibrium.

- **6.2 Market Equilibria and Entry Thresholds Prediction**

In addition to inferring incumbent conducts, the available data enables likelihood estimation that predicts the equilibrium number of airlines (N) in a market with particular demand conditions. For a market to sustain N firms in equilibrium, the N+1th firm must earn negative profit upon entry. Suppose the profit function of an airline can be written as

$$\pi(N) = V_N \times S - F_N + \varepsilon_m = \overline{\pi(N)} + \varepsilon_m$$

, where S represents market size (approximated by the *pop* variable), V_N and F_N respectively denote per-capita variable profit and fixed cost of the Nth firm, and the unobservable error term ε_m is normally distributed, homoskedastic, and independent of explanatory variables.

The assumption of normally distributed errors justifies the use of probit functions to estimate the coefficients. The probability of observing no airlines operating in a market is $\Pr(\mathrm{N} = 0) = \Pr(\pi(1) < 0) = 1 - \Phi\left[\overline{\pi(1)}\right]$ ①, where $\Phi(\cdot)$ stands for the cumulative normal distribution function (normal CDF). I expect airline profits to decline with the entry of an additional operator (regardless of its type), and therefore the chance of observing N airlines in an equilibrium market is $\Pr(\mathrm{N}) = \Pr(\pi(N) > 0 \,\&\, \pi(N+1) < 0) = \Phi[\overline{\pi(N)}] - \Phi[\overline{\pi(N+1)}]$ ②. Expression ① and ② collectively define the probability density functions in the likelihood function (assuming markets are mutually independent).

The variable profits function for the N-th airline is assumed to be linear and contains additively separable components: $V_N = X\beta + \alpha_1 * d_1 - \sum_{i=2}^{N} \alpha_i * d_i = V_1 - \sum_{i=2}^{N} \alpha_i * d_i$ ③, where X matrix involves the demand shifters, coefficients α_i measure the change in per head variable profits when the i-th airline enters the market, d_i is a dummy variable which becomes 1 when the i-th airline enters, and V_1 equals the per head variable profits of a monopolist. The X matrix comprises *dist* and *ally*, and all α_i's are expected to be positive.

Similarly, the fixed cost function for the N-th firm is characterized as $F_N = Z\delta + \gamma_1 * d_1 + \sum_{i=2}^{N} \gamma_i * d_i = F_1 + \sum_{i=2}^{N} \gamma_i * d_i$ ④, where Z matrix holds the cost shifters, coefficients γ_i quantify the rise in fixed cost with the addition of another airline, and F_1 equals the fixed cost of a monopolist. The Z matrix consists of hub information, and all γ_i's are expected to be positive.

Additionally, the variable *pop* and *dist* are transformed using the following formula to produce a reasonable estimate for the coefficients:

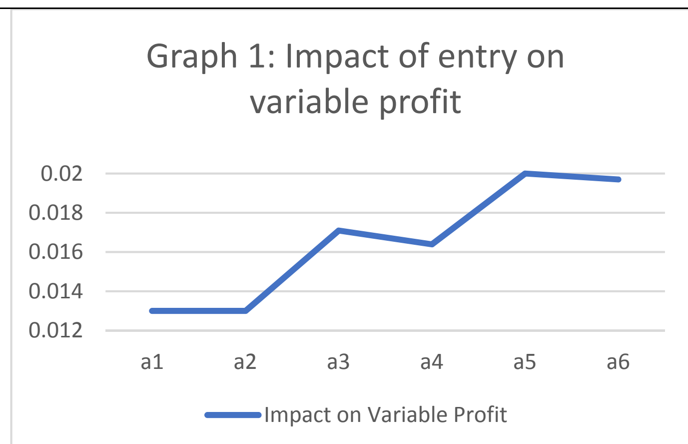
$$popst_m = \ln\left[\frac{pop_m}{mean(pop_m)}\right] \text{ ⑤};\ distst_m = \ln\left[\frac{dist_m}{mean(dist_m)}\right] \text{ ⑥}$$

Due to the transformation, variable values above the mean become positive and those below the mean become negative, and a value equal to the mean is converted to zero. This practice ensures that the estimated coefficients for market size (captured by *pop*) and distance are not too small.

Substituting in ③ and ④, the profit function becomes $\pi(N) = [X\beta + \alpha_1 d_1 - \sum_{i=2}^{N} \alpha_i d_i] \times S - [Z\delta + \gamma_1 d_1 + \sum_{i=2}^{N} \gamma_i d_i] + \varepsilon_m$, or equivalently:

$$\pi(N) = S \cdot X\beta + \alpha_1 d_1 \cdot S - S \cdot \sum_{i=2}^{N} \alpha_i d_i - Z\delta - \gamma_1 d_1 - \sum_{i=2}^{N} \gamma_i d_i + \varepsilon_m$$

Table A: Estimated Profit-function Parameters from MLE:

Coefficients		**Estimated Values**	
$S \cdot X\beta$: Interaction between S and demand shifters in X	popst*distst	0.0145	
	popst*ally	0.0087	
$S \cdot \alpha_i d_i$: interaction between S and all the di dummies (coefficients on popst*di)	α_1	0.0130	
	α_2	0.0130	
	α_3	0.0171	
	α_4	0.0164	
	α_5	0.0200	
	α_6	0.0197	
$-Z\delta$: impact of cost shifters (hubs) on fixed cost F	hub_b6	0.0103	
	hub_aa	0.0164	
	hub_dl	0.0159	
	hub_ua	0.0159	
	hub_sw	0.0515	
	hub_as	0.0156	
	hub_f9	0.0162	

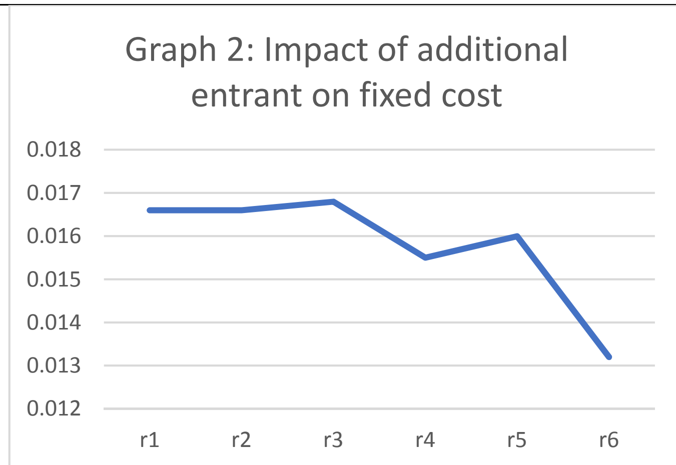
	hub_nk	0.0173	
Coefficients on di: impacts of additional entrant on fixed cost	γ_1	0.0166	
	γ_2	0.0166	
	γ_3	0.0168	
	γ_4	0.0155	
	γ_5	0.0160	
	γ_6	0.0132	
Constant (Intercept) _cons		8.0167	

Table A displays the parameter estimates of the profit function. Consistent with expectation, all the alpha and gamma estimates have positive signs, indicating that additional entrants bring down per capita variable profit and face higher fixed costs. However, contrary to Bresnahan and Reiss's finding, it is the later entrants, rather than the first few entrants, that have greater impact on variable profits (as visualized by graph 1). Since I have concluded that U.S. aviation market is not perfectly competitive, the incremental effects of new entrants on variable profit can be explained using cartel theory: when there are few competitors in a market, it is easy for the incumbents to collude and form a cartel to maximize profit. As new airlines enter the market, the cartel becomes increasingly hard to sustain, and when it breaks down, the drop in variable profit is striking.

Graph 2 shows that new entrants tend to face higher fixed costs, but the size of the increments is diminishing. I cannot distinguish from the data whether the rise in fixed cost is due to inefficient technologies or entry deterrence created by the incumbents. Nonetheless, the result echoes the findings in 6.1 that newcomers suffer from higher entry barriers.

The estimated parameters enable the prediction of profits under different market conditions. By imposing the two conditions $\pi(N) > 0$ and $\pi(N+1) < 0$, the equilibrium number of firms operating in a market (N) can be identified. The entry threshold, or the market size that can support a certain number of airlines, can then be estimated by $S(N)^{hat} = F(N)^{hat}/V(N)^{hat}$, where $F(N)^{hat}$ and $V(N)^{hat}$ are obtained using the estimated parameters and variable values in the dataset.

Note that the analysis here does not discern the intrinsic difference between low-cost carriers and full-service airlines: the dependent variable is just the total number of airlines flying nonstop in a market, and the differential effect of heterogeneous competitors on an airline's payoff cannot be distinguished. To study the effect of product differentiation on firm profits and market structure, a more comprehensive mathematical framework must be utilized, as discussed in the following section.

7. Adaptation of the Empirical Framework in Mazzeo (2002)

So far, the analysis has been presuming that airline entry is a two-stage sequential game, in which the FSAs have established themselves in stage 1, and LCCs make entry decisions in stage 2, given incumbents' behavior in the previous period. The assumption is made to avoid simultaneous movements in stage 2 (i.e. incumbent FSAs collude as a response to entry, while LCCs' entry decisions hinge on the anti-competitive behavior of the FSAs), because no pure strategy Nash Equilibrium can be found unless restrictive symmetries are imposed. As a consequence, only the code-sharing decisions and mergers right before an entry takes place were considered.[17] The sequential game presumption can be further clarified by permitting incumbent firms in stage 1 to predict the actions of later entrants when making their own profit maximizing decisions. Drawing upon Mazzeo (2002), I come up with an empirical framework to assess the effects of product differentiation and incumbent collusion on the profitability of different types of airlines.

Mazzeo argues that two decisions are endogenous to a firm: the decision of entry and what types of product to offer. Besides, his model is based on the following premises: a market will have N firms when $\pi(N) > 0$ and $\pi(N+1) < 0$; companies offering different kinds of product have separate payoff functions; profits are non-increasing with the number of competing firms, and firms can either make their entry and product offering decisions in the same period (the Stackelberg-style specification) or in two different periods (the two sub-stage specification).

- **7.1 Payoff Function**

In the airline entry context, suppose that FSAs and LCCs are providing two kinds of product: low-level service (L) by the low-cost carriers, and high-level service (H) by the full-service airlines. FSAs and LCCs have distinct payoff functions that rely on market demand features (represented by an X matrix), number of firms offering the same kind of product, number of firms offering heterogeneous type of product, and the anti-competitive behavior of

[17] The original statistics I used is on U.S. domestic travel information in 2016, and as a result only the code-sharing that were initiated before 2016 are treated as a valid entry for the "ally variable" in the constructed dataset for regression (e.g. AS and AA started to codeshare since 2004). In a similar vein, the AA and US merger was initiated in 2013 and completed by Oct. 2015, which is before the LCC entry game in 2016 that this paper has been focusing on. Other mergers (UA/CO in 2010; DL/NW in 2009) are too early to be regarded as collusive behavior against competition.

FSAs (the ally variable introduced in section 4, here denoted as d_m). Therefore, the payoff function is specified as $\pi_T^m = X_m\beta_T + f(\vec{N}, \gamma_T, d_m) + \epsilon_T^m$, where m is market, and T is the types of firm (H or L). The f function captures the effect on payoff by homogeneous and heterogeneous competitors: the 1*2 $\vec{N}$ vector stands for the number of FSA rivals (N_1) and LCC rivals (N_2); the gamma vector involves parameters that represent the incremental effect of a particular type of competitor (H or L) on payoffs: for example, γ_{LH3} is the effect of the 3rd H-type airlines on average L-type payoff; γ_{LL2} describes the effect of the 2nd L-type rival on average L-type payoff. Note that the specification does not permit heterogeneity within the same type (e.g. JetBlue's effect on FSA profitability cannot be distinguished from that of Spirit airlines), in order to keep the number of parameters to estimate at a reasonably low level. The dummy variable d_m captures whether FSA collusion (i.e. code-sharing, joining alliance, or merging) exists in market m, and the effects of collusion on payoffs are represented as α_L for LCCs and α_H for FSAs. Typically, FSAs collude to enhance their profits, and I expect α_H to be positive. Collusions by FSAs might hurt the profitability of LCCs, so I expect α_L to be negative. The unobservable in the payoff function, ϵ_T^m, is presumed to be independent of all the observables, additively separable, and different for each type of firms in a given market.

For instance, in the Denver to Los Angeles segment, 2 LCCs and 4 FSAs are operating nonstop in 2016Q1, so the observed market configuration is (L, H) = (2, 4). Since vector $\vec{N}$ captures only the rival carriers, $\vec{N} = (1,4)$ for each LCC and $\vec{N} = (2,3)$ for a FSA. We can parameterize the average payoff of a LCC as:

$$\pi_L^m = X_m\beta_L + \gamma_{LL1} + \gamma_{LH1} + \gamma_{LH2} + \gamma_{LH3} + \gamma_{LH4} + \alpha_L * d_m + \epsilon_L^m$$

, and the average payoff of a FSA as:

$$\pi_H^m = X_m\beta_H + \gamma_{HL1} + \gamma_{HL2} + \gamma_{HH1} + \gamma_{HH2} + \gamma_{HH3} + \alpha_H * d_m + \epsilon_H^m$$

❖ 7.2 Equilibrium Identification

An equilibrium market can be depicted by the following inequalities and conditions:

1) $\pi_L(X, L, H, d_m) > 0$ and $\pi_L(X, L+1, H, d_m) < 0$;
2) $\pi_H(X, L, H, d_m) > 0$ and $\pi_H(X, L, H+1, d_m) < 0$, and
3) $\frac{\partial \pi}{\partial L} \leq 0$ and $\frac{\partial \pi}{\partial H} \leq 0$

Define $h_T(X, L, H, d_m) = X_m\beta_T + f(\vec{N}, \gamma_T, d_m)$, and denote $\pi_T^m = h_T(X, L, H, d_m) + \epsilon_T^m$. Hence condition 1) and 2) can be represented by:

4) $h_L(X, L+1, H, d_m) < -\epsilon_L < h_L(X, L, H, d_m)$, and
5) $h_H(X, L, H+1, d_m) < -\epsilon_H < h_H(X, L, H, d_m)$

4) and 5) jointly define the region in which the (L, H) outcome in market m can be realized. In addition, if product differentiation does soften competition, the profit margin will change less drastically with the entry of a heterogeneous firm than with a homogeneous one. Mathematically: $\pi_L(X, L, H, d_m) - \pi_L(X, L+1, H, d_m) > \pi_L(X, L, H, d_m) - \pi_L(X, L, H+1, d)$ and $\pi_H(X, L, H, d_m) - \pi_H(X, L, H+1, d_m) > \pi_H(X, L, H, d_m) - \pi_H(X, L+1, H, d_m)$, which are equivalent to:

6) $\pi_L(X, L, H+1, d_m) > \pi_L(X, L+1, H, d_m)$, and
7) $\pi_H(X, L+1, H, d_m) > \pi_H(X, L, H+1, d_m)$

From which we can further derive $\pi_L(X, L, H+1, d_m) > \pi_H(X, L, H+1, d_m)$ and $\pi_H(X, L+1, H, d_m) > \pi_L(X, L+1, H, d_m)$. These two conditions, together with inequalities 4) and 5), characterize the equilibrium in each market.

❖ 7.3 Entry Game Specification

In contrast to the regression model in section 4 which treated incumbents' entry decisions and collusive behavior as exogenously given, here the decisions on entry and collusion by all firms (FSA and LCC) are endogenous. Since I observe the equilibrium market configuration (L, H) in each route, the rules of the entry game do not matter much for the likelihood estimation.

❖ 7.4 Estimation (MLE)

The parameters of the payoff functions can be estimated using maximum likelihood, which selects the parameter estimates that maximize the probability of the observed market structures[18] in the constructed dataset. For the 1703 observations, the likelihood function takes the form of

$$L = \prod_{m=1}^{1703} Pr(L, H, d)_m$$

Using conditions 1) to 7), any realization of (L, H) can be represented by a realization of (ϵ_L, ϵ_H). Assuming the error terms are bivariate normal, the joint PDF for (ϵ_L, ϵ_H) takes the form

$$f(\epsilon_L, \epsilon_H) = \frac{1}{2\pi * \sqrt{1-\rho^2} * \sigma_{\epsilon_L} * \sigma_{\epsilon_H}} * \exp(-\frac{z_1^2 + z_2^2 - 2*\rho*z_1*z_2}{2*(1-\rho^2)}), \text{ where } z_i = \frac{\epsilon_i - \mu_{\epsilon i}}{\sigma_{\epsilon_i}}.$$

In the dataset, L ranges from 0 to 2 and H from 0 to 5. A nested expression for the payoff of a type T airline operating in market m is

$$\pi_T^m = X_m\beta_T + \sum_{i=1}^{2} \gamma_{TLi} * d_{TLi} + \sum_{j=1}^{5} \gamma_{THj} * d_{THj} + \alpha_T * d_m + \epsilon_T^m$$

[18] A market is characterized by the numbers of LCCs (L) and FSAs (H), as well as whether the incumbent is "allying" (d) in a particular market.

, where d_{TLi} is a dummy which gets 1 if the i-th low-type competitor is present. Since a market will have N entrants when $\pi(N) > 0$ and $\pi(N+1) < 0$, the probability of observing 0 L-type firm is $1 - \Phi[h_L(X, 1, H, d_m)]$, of observing 1 L-type firm is $\Phi[h_L(X, 1, H, d_m)] - \Phi[h_L(X, 2, H, d_m)]$, and of observing 2 L-type firms is $\Phi[h_L(X, 2, H, d_m)]$.[19] Similarly, for H-type firms, the odds of observing 0 H-type firm is $1 - \Phi[h_H(X, L, 1, d_m)]$, of observing 1 H-type firm is $\Phi[h_H(X, L, 1, d_m)] - \Phi[h_H(X, L, 2, d_m)]$, and of observing 5 H-type firms is $\Phi[h_H(X, L, 5, d_m)]$. The likelihood function resembles:

$$L = \prod_{m=1}^{1703} \{\Phi[h_H(L, H)] - \Phi[h_H(L, H+1)]\} * \{\Phi[h_L(L, H)] - \Phi[h_L(L+1, H)]\}^{20}$$

Or in logarithm form:

$$L = \sum_{m=1}^{1703} ln\{\Phi[h_H(L, H)] - \Phi[h_H(L, H+1)]\} + ln\{\Phi[h_L(L, H)] - \Phi[h_L(L+1, H)]\}$$

In addition, *pop* and *dist* are transformed to *popst* and *distst*, using formula ⑤ and ⑥ in section 6.2.

❖ 7.5 MLE Results

Assuming the 1703 data points are independently drawn from the bivariate normal distribution, and the kernel of the log-likelihood function for a single observation is expressed as:

$$l = -\frac{1}{2} ln\sigma_1^2 - \frac{1}{2} ln\sigma_2^2 - \frac{1}{2}\ln(1-\rho^2) - \frac{1}{2(1-\rho^2)}[z_1^2 + z_2^2 - 2\rho z_1 z_2]$$

[19] As defined before, $h_T(X, L, H, d_m) = X_m\beta_T + f(\vec{N}, \gamma_T, d_m)$, and $\pi_T^m = h_T(X, L, H, d_m) + \epsilon_T^m$; $\Phi[\cdot]$ is the normal CDF;

[20] Matrix X and variable d_m are omitted in this expression to simplify notation.

Table B: Estimated Parameters from MLE:

Coefficient		Estimated values
Effect on L-type (LCC) payoffs:		
Constant (Intercept) _cons		0.5516
Market Demand Features (X)	popst	0.1536
	distst	-0.0576
ally		**0.0795**
Effect of another competitor (L or H type) on L-type payoff (d_ll1: impact of 1st L-rival on L payoff)	d_ll1	1.5927
	d_lh1	-0.6992
	d_lh2	0.2245
	d_lh3	0.0702
	d_lh4	0.0488
	d_lh5	0.0243
Effect on H-type (FSA) payoffs:		
Constant (Intercept) _cons		0.7888
Market Demand Features (X)	popst	0.2196
	distst	-0.0824
ally		**0.1137**
Effect of another competitor (L or H type) on H-type payoff (d_hl1: effect of 1st L-rival on H payoff)	d_hl1	-1.4301
	d_hl2	0.8477
	d_hh1	1.3210
	d_hh2	1.1004
	d_hh3	1.0698
	d_hh4	1.0348

Table B exhibits the coefficient estimates of the two payoff functions (π_L and π_H). Using the estimated parameters, one can predict the relative payoffs of operating as a LCC or a FSA under various market conditions (X and ally) and in different product-type configurations. For instance, the estimated intercepts indicate that, in markets with similar demand (X) conditions, monopolizing a market as a LCC is on average less profitable than as a FSA (_cons_LCC=0.5516 vs. _cons_FSA=0.7888).

Market demand condition (*popst*) has a positive and significant effect on payoffs of both LCCs and FSAs, and the difference in relative size of the coefficients ($\beta_{popst\_lcc} = 0.1536$ vs $\beta_{popst\_fsa} = 0.2196$) suggests that FSAs might favor markets with population above the mean (their payoffs grow faster with increasing market demands), while LCCs might prefer markets with below-mean population. In reality, the predilection for FSA remains when different values of *popst* are taken account of. For instance, holding *distst* at sample mean level and assuming *popst* is twice the sample mean in market *i*, a monopolizing FSA will in general acquire more profits (π_H=0.7888+0.2196*0.6931=0.9410) than a monopolizing LCC (π_L=0.5516+0.1536*0.6931=0.6581). If, holding others constant, *popst* is only 5% of the sample mean in market *j*, then a monopolizing FSA still earns more (π_H=0.7888+0.2196*(-2.9957) = 0.1309) than a monopolizing LCC (π_L=0.5516+0.1536*(-2.9957)=0.0915). The relationship will only be reversed when market size shrinks to approximately 2.7% of the sample mean and payoffs for both types (H and L) are negative, indicating that the market is too small to sustain either type of airlines. In sum, the baseline preference for offering high-level service is too high for demand condition (*popst*) to alter it.

Both estimations report positive estimates for the effect of collusion (*ally*) on airline payoff, suggesting that collusion among FSAs will increase the payoffs of both FSAs and LCCs. The first conclusion is compatible with traditional oligopoly theory, while the second is quite counter-intuitive. One plausible explanation is that demands are fairly inelastic in certain markets, and consequently, the colluding FSAs raise prices to maximize payoffs, making some customers switch from flying with them to with low-cost carriers.

However, the effects of rivals on airline payoffs are quite unanticipated. Instead of having negative signs, almost all estimates (except for d_lh1 and d_hl1) are positive in value, albeit the extent of impact is gradually declining (abs (d_lh_i) > abs (d_lh_i+1); abs (d_hh_i) > abs (d_hh_i+1)). The negative signs of d_lh1 and d_hl1 show that the first heterogeneous carrier significantly reduces the payoff of an airline, but later heterogeneous entrants and homogeneous firms improve the profitability of an airline, regardless of its type. The results are different from the conclusion in Mazzeo (2002), in which he argues that homogeneous competitors have stronger negative impact on payoffs than heterogeneous competitors, and firms are eager to differentiate.

Here I propose an alternative specification for the payoff functions that involves fewer parameters to estimate. Current MLE analysis requires estimation of 18 coefficients and 2 intercepts from π_L and π_H, and the relatively small sample size (N=1703) is insufficient to produce accurate estimates for the 20 parameters. Borrowing from Mazzeo (2002), I estimate the "average competition effect" of additional heterogeneous competitors[21] when there are more than one heterogeneous rivals. For example, in a market with 2 LCCs (L-type) and 4 FSAs (H-type), I estimate the impact of first H-type rival on L-type payoff, and the average impact of the second, third, and fourth H-type rivals on π_L. This parameterization of the payoff functions

[21] A LCC is a heterogeneous competitor for FSAs in a market, and a FSA is a heterogeneous competitor for LCCs.

helps avoid potential perfect collinearity issues among the dummy variables, and significantly reduces the number of coefficients to estimate. As a consequence, the payoff function for LCCs is transformed into

$$\pi_L^m = X_m\beta_L + \gamma_{LL1}D_{LL1} + \gamma_{L0H1}D_{L0H1} + \gamma_{L0HX}NUM_{L0HX} + \alpha_L * d_m + \epsilon_L^m \text{ ⑦}$$

, where D_{LL1} represents the presence of the first L-type competitor, D_{L0H1} stands for the presence of the first H-type competitor when there are no L-type rivals, and NUM_{L0HX} captures the number of additional H-type competitors when there are no other L-types. In a similar vein, FSA's payoff function can be parameterized by

$$\pi_H^m = X_m\beta_H + \delta_{HH1}D_{HH1} + \delta_{HH2}D_{HH2} + \delta_{H0L1}D_{H0L1} + \delta_{H0LX}NUM_{H0LX} + \alpha_H * d_m + \epsilon_H^m \text{ ⑧}$$

, where D_{HH1} denotes the presence of the first H-type competitor, D_{HH2} is the dummy for the second H-type competitor, D_{H0L1} represents the presence of the first L-type competitor when there are no H-type rivals, and NUM_{H0LX} includes the number of additional L-type competitors when there are no other H-types. In short, the only difference in the alternative specifications ⑦ and ⑧ lies in the f function in $\pi_T^m = X_m\beta_T + f(\vec{N}, \gamma_T, d_m) + \epsilon_T^m$, while the X matrix and error term stay the same.

Table C. MLE Result for the alternative payoff function specification

Coefficient		**Estimated values**
Effect on L-type (LCC) payoffs:		
Constant (Intercept) _cons		0.6429
Market Demand Features (X)	popst	0.2234
	distst	-0.0481
ally		**0.0350**

Effect of another competitor (L or H type) on L-type payoff	D_ll1	1.3470
	D_l0h1	**-0.5474**
	Num_l0hx	0.0652
<u>Effect on H-type (FSA) payoffs:</u>		
Constant (Intercept) _cons		0.8092
Market Demand Features (X)	popst	0.2007
	distst	-0.1033
ally		**0.1811**
Effect of another competitor (L or H type) on H-type payoff	D_hh1	1.0027
	D_hh2	1.3104
	D_h0l1	**-0.6012**
	Num_h0lx	0.2873

The estimation results are not much different from those in table B. A feasible interpretation of the outcome is that some necessary control variables might be missing in my model, and improvements can be made by finding and adding eligible controls that are exogenously predetermined or immutable.[22]

[22] By "immutable", I mean market characteristics that do not change over time, but are different across routes.

8. Conclusion and Potential Ways to Improve

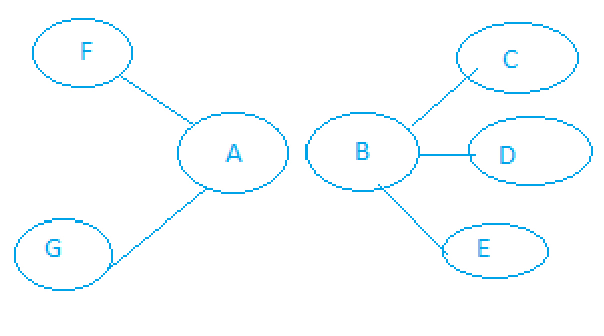
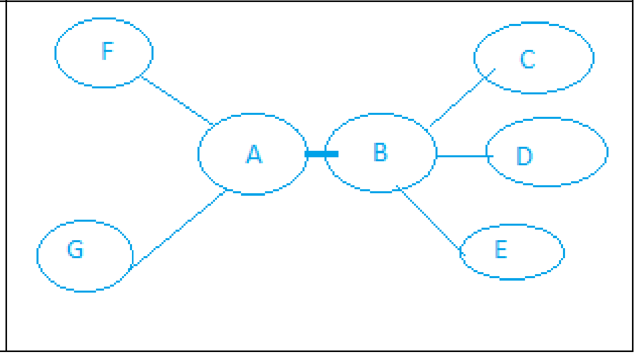
Evidence from the sequence of entry ratios shows that the U.S. aviation market is not perfectly competitive, and entrants face entry barriers. Empirical results from the regression further demonstrate that the anti-competitive behavior from incumbent airlines does affect the entry decisions of low-cost carriers. Although the results are significant, the empirical part of the paper does not take account of incumbent's reactions in stage 2 of the game. In addition, since entering into a new market does not happen frequently, the approach in section 4 and 5 treats the entry decision problem in a static manner by looking at cross-sections of markets and analyzing the impact of existing market structures[23] on entry decisions. The dynamic aspect of entry decisions, where LCCs choose to enter different routes across time, is not examined.

A feasible strategy to refine the current paper is to use a panel data method and choose a starting point (e.g. 2013Q1) as a baseline. Only the routes opened by a LCC after the cutoff time are regarded as valid observations of entry, and collusions that take place in one period only have anti-competitive impacts from the next period onwards. This method requires much greater effort in dataset construction, and since entries and exits are uncommon phenomena, the estimated parameters from regressions might be insignificant.

Additional factors that might affect entry decisions can be added to the regression equation. For instance, if opening up a new route leads to greater connectivity in an airline's network (i.e. by adding a new route into the airline's network, more one-stop routes are created), then the firm might be more willing to enter the market, and the number of new one-

[23] E.g. population, distance, hub presence, and collusions that happened before entries occur

stop itineraries that would be realized can be used as an explanatory variable that depicts the network-improving effect.

Prior to entering the A-B segment:	After entering the A-B segment: 5 new one-stop itineraries are enabled
	

However, simply adding regressors to the equation will dilute the effect of the target variable that this paper has been focusing on, and this new variable, capturing network-improving effects, might be highly correlated with the d dummies in the MLE, making it a bad control. As a result, the introduction of additional variables should be moderated.

Furthermore, incomplete information about the incumbents' willingness to fight entry via collusion can be added to probe the effect of deterrence by the existing airlines on the entry strategies of the LCCs. Kreps and Wilson study the equilibrium strategies of an incumbent monopolist and potential entrants in a sequential game, while assuming that the entrants cannot directly observe the monopolist's type ex ante. Their theoretical model that analyzes reputation's effect on forestalling entries in an asymmetric information context can be applied to the airline entry game. Since incumbent collusion and deterrence lower the probability of entry, and exclusionary practices like predation and code-sharing are costly, an incumbent airline might want to appear "tough" by signaling that it will fight entries whenever they occur. To make the threat of fighting entry credible, "fighting entry" should be the optimal strategy for the incumbent in every subgame (i.e. fighting is a subgame perfect Nash equilibrium strategy), or at least it should appear to the potential entrants that fighting is ex post optimal for the incumbent. As a consequence, possible strategies for the incumbent airlines involve actively expanding into new markets (according to Kreps and Wilson (1982), when there is a sequence of markets controlled by the incumbent that is enterable, the incumbent will always choose to fight in order to protect its long-term payoff), or building up capacities to cut down production cost. Such practices will raise the probability evaluated by the entrants that fighting is optimal for the incumbent in every subgame, and lower the chance of entry without the incumbent actually engaging in the costly fighting.

Appendix:

1. Terms and airline code:

Low-cost carriers that we're interested in	B6: JetBlue Airways	F9: Frontier Airlines	NK: Spirit Airlines
The 3 major full-service airlines in US	AA: American Airlines	DL: Delta Airlines	UA: United Airlines
The largest LCC in US	SW: Southwest Airlines		
Others	AS: Alaska airlines US: US Airways (merged with AA)		

2. Hub and focus cities for each airline (within the list of the 45 airports): 1 for yes, blank for no

Airport code	B6	F9	NK	AA	DL	UA	SW	AS	US
ATL		1			1		1		
BOS	1				1				
BWI							1		
CLE		1							
CVG		1			1				
DAL							1		
DCA				1					1
DEN		1				1	1		
DFW			1	1					
DTW			1		1				
EWR						1			
FLL	1		1						
IAD						1			
IAH						1			
JFK	1			1	1				
LAS			1				1		
LAX				1	1	1		1	
LGA				1	1				
MCO	1	1					1		
MDW							1		
MIA				1					
MSP					1				
OAK							1		
ORD		1	1	1		1			
PDX								1	
PHL		1		1					1
PHX				1			1		1
SAN							1	1	
SEA					1			1	
SFO						1		1	
SLC					1				
SJC								1	

Source of information: U.S. Department of Transportation, Bureau of Transportation Statistics, Data and Statistics: Carrier Snapshot, 2016 https://www.transtats.bts.gov/carriers.asp

3. Top 10 US domestic routes (city-pairs, not airport-pairs):
   1) Chicago to NYC
   2) Los Angeles to San Francisco
   3) Chicago to Los Angeles
   4) Los Angeles to NYC
   5) Atlanta to Chicago
   6) Atlanta to NYC
   7) Miami to NYC
   8) Chicago to San Francisco
   9) Chicago to Minneapolis
   10) Atlanta to Orlando

Source of information: https://www.transtats.bts.gov/

4. Airline merger that was completed right before 2016Q1 (period of interest): The merger between American Airlines and US Airways was completed in October 2015; AA took over US's fleet, and the former hub airports of US became AA's hubs.

5. Airline code-sharing: Alaska Airlines (AS) codeshares with American Airlines on several flight segments out of LAX and SEA. In addition, as AS and Virgin America (VX) is currently undergoing merger, the two airlines codeshare on all the flights operate by either airline.

6. How are the control variables constructed?

   The construction of route characteristic variables (Pop and Dist) are discussed in the main text. Entrants' and incumbents' features are represented by the hub dummies: if at least one of the endpoint airports in a flight segment is a hub airport for the entrant, then this hub dummy will get a value of 1. For example, when analyzing JetBlue's entry decision on the BOS-SFO segment, BOS is a hub for JetBlue, and thus hub_b6 is 1 for this observation. In addition, SFO is a hub for a rival airline (United), so the hub_ua variable will have a value of 1.

   The numrival variable describes the number of rival airline operating in a market. For instance, three airlines fly nonstop between Boston (BOS) and Denver (DEN)—United, Southwest, and JetBlue. When investigating JetBlue's entry decision on the route, the number of rivals is 2.

7. Regression outputs: table 1 to 3

1) Table 1

```
. probit entry_b6 pop dist ally numrival_b6 hub_b6 hub_aa hub_dl hub_ua hub_sw

Iteration 0:   log likelihood = -532.89527
Iteration 1:   log likelihood = -329.82156
Iteration 2:   log likelihood = -318.35735
Iteration 3:   log likelihood = -318.23007
Iteration 4:   log likelihood = -318.22997
Iteration 5:   log likelihood = -318.22997

Probit regression                               Number of obs     =      1,703
                                                LR chi2(9)        =     429.33
                                                Prob > chi2       =     0.0000
Log likelihood = -318.22997                     Pseudo R2         =     0.4028
```

entry_b6	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pop	.0000137	.0000107	1.28	0.199	-7.23e-06	.0000347
dist	.000083	.0000821	1.01	0.312	-.000078	.000244
ally	-.352239	.1492476	-2.36	0.018	-.644759	-.059719
numrival_b6	.0149166	.0364476	0.41	0.682	-.0565193	.0863525
hub_b6	1.928633	.1221556	15.79	0.000	1.689212	2.168054
hub_aa	.1421412	.1496304	0.95	0.342	-.1511289	.4354114
hub_dl	.1918699	.125928	1.52	0.128	-.0549445	.4386843
hub_ua	.2097655	.1436906	1.46	0.144	-.0718629	.4913939
hub_sw	-.2930843	.1271267	-2.31	0.021	-.542248	-.0439207
_cons	-2.416848	.178028	-13.58	0.000	-2.765777	-2.06792

.

2) Table 2

```
. probit entry_f9 pop dist ally numrival_F9 hub_f9 hub_aa hub_dl hub_ua hub_sw hub_nk

Iteration 0:   log likelihood = -486.28037
Iteration 1:   log likelihood = -329.05508
Iteration 2:   log likelihood = -313.72382
Iteration 3:   log likelihood = -313.45729
Iteration 4:   log likelihood = -313.45726

Probit regression                               Number of obs     =      1,703
                                                LR chi2(10)       =     345.65
                                                Prob > chi2       =     0.0000
Log likelihood = -313.45726                     Pseudo R2         =     0.3554
```

entry_f9	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pop	-.0000506	.0000145	-3.49	0.000	-.0000791	-.0000222
dist	-.000098	.0000985	-1.00	0.320	-.0002909	.000095
ally	.0856181	.1421589	0.60	0.547	-.1930081	.3642443
numrival_F9	-.1062696	.0647911	-1.64	0.101	-.2332578	.0207187
hub_f9	1.112606	.1305632	8.52	0.000	.8567072	1.368505
hub_aa	.2172534	.15429	1.41	0.159	-.0851494	.5196563
hub_dl	.1395069	.1359286	1.03	0.305	-.1269081	.405922
hub_ua	.9497996	.1451605	6.54	0.000	.6652903	1.234309
hub_sw	.983593	.1343397	7.32	0.000	.720292	1.246894
hub_nk	.2912198	.123152	2.36	0.018	.0498463	.5325933
_cons	-2.370408	.2122307	-11.17	0.000	-2.786372	-1.954443

Note: 1 failure and 0 successes completely determined.

3) Table 3

```
. probit entry_nk pop dist ally numrival_nk hub_nk hub_aa hub_dl hub_ua hub_sw hub_f9

Iteration 0:   log likelihood = -690.70486
Iteration 1:   log likelihood = -521.97223
Iteration 2:   log likelihood = -514.40913
Iteration 3:   log likelihood = -514.36982
Iteration 4:   log likelihood = -514.36982

Probit regression                               Number of obs     =      1,703
                                                LR chi2(10)       =     352.67
                                                Prob > chi2       =     0.0000
Log likelihood = -514.36982                     Pseudo R2         =     0.2553
```

entry_nk	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pop	-1.23e-06	9.15e-06	-0.13	0.893	-.0000192	.0000167
dist	.0000289	.0000735	0.39	0.694	-.0001152	.0001729
ally	-.2409768	.1169942	-2.06	0.039	-.4702812	-.0116723
numrival_nk	-.0070853	.0244097	-0.29	0.772	-.0549275	.0407569
hub_nk	1.323644	.0920636	14.38	0.000	1.143203	1.504085
hub_aa	.3773991	.1139233	3.31	0.001	.1541134	.6006847
hub_dl	.2086099	.0960687	2.17	0.030	.0203188	.396901
hub_ua	.4818536	.1087735	4.43	0.000	.2686615	.6950458
hub_sw	.4607827	.0969214	4.75	0.000	.2708202	.6507451
hub_f9	.1057235	.1028464	1.03	0.304	-.0958517	.3072987
_cons	-2.255027	.1568544	-14.38	0.000	-2.562456	-1.947598

.

8. Section 6: MLE Outcome

```
                                                Number of obs     =      1,703
                                                Wald chi2(7)      =          .
Log likelihood =  37553.534                     Prob > chi2       =          .

------------------------------------------------------------------------------
    num_firm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
popst_distst |   .0144892   .0017768     8.15   0.000     .0110066    .0179717
  popst_ally |    .008652          .        .       .            .           .
    popst_d1 |   .0129768          .        .       .            .           .
    popst_d2 |   .0129768          .        .       .            .           .
    popst_d3 |   .0171085    .038889     0.44   0.660    -.0591126    .0933296
    popst_d4 |   .0163818          .        .       .            .           .
    popst_d5 |   .0199562          .        .       .            .           .
    popst_d6 |   .0197128          .        .       .            .           .
      hub_b6 |   .0103277          .        .       .            .           .
      hub_aa |   .0163598   .0040751     4.01   0.000     .0083728    .0243468
      hub_dl |    .015935          .        .       .            .           .
      hub_ua |   .0158862   .0021356     7.44   0.000     .0117004    .0200719
      hub_sw |   .0514519          .        .       .            .           .
      hub_as |   .0155776    .002081     7.49   0.000     .0114988    .0196563
      hub_f9 |   .0161548   .0029668     5.45   0.000       .01034    .0219696
      hub_nk |    .017347          .        .       .            .           .
          d1 |   .0166119          .        .       .            .           .
          d2 |   .0166119          .        .       .            .           .
          d3 |    .016849    .005027     3.35   0.001     .0069963    .0267017
          d4 |   .0155088          .        .       .            .           .
          d5 |    .015953          .        .       .            .           .
          d6 |    .013194          .        .       .            .           .
       _cons |   8.016669   .0013692  5855.18   0.000     8.013985    8.019352
------------------------------------------------------------------------------
```

9. Section 7: Observed market configurations in each flight segment:

Market Structure (L, H) (L for LCC and H for FSA)	**Number of markets with this structure**	**Shares in total (%)**
(0,0)	306	17.97
(0,1)	544	31.94
(0,2)	280	16.44
(0,3)	99	5.81

(0,4)	15	0.88
(0,5)	7	0.41
(1,0)	51	2.99
(1,1)	122	7.16
(1,2)	133	7.81
(1,3)	41	2.41
(1,4)	10	0.59
(1,5)	6	0.35
(2,0)	26	1.53
(2,1)	25	1.47
(2,2)	26	1.53
(2,3)	10	0.59
(2,4)	2	0.12

10. Section 7 MLE Outcome 1: The original payoff function specification

	Number of obs	=	1,703
	Wald chi2(9)	=	1.10e+13
Log likelihood = 20710.442	Prob > chi2	=	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
eq1						
popst	.1535523	4.08e-07	3.8e+05	0.000	.1535515	.1535531
distst	-.0576242	3.51e-07	-1.6e+05	0.000	-.0576249	-.0576235
ally	.0794853	5.54e-07	1.4e+05	0.000	.0794842	.0794864
d_ll1	1.592746	8.72e-07	1.8e+06	0.000	1.592744	1.592748
d_lh1	-.6992479	2.73e-07	-2.6e+06	0.000	-.6992485	-.6992474
d_lh2	.2244821	5.32e-07	4.2e+05	0.000	.2244811	.2244831
d_lh3	.0702169	3.62e-07	1.9e+05	0.000	.0702161	.0702176
d_lh4	.0488181	7.75e-07	6.3e+04	0.000	.0488166	.0488196
d_lh5	.0242996	1.11e-06	2.2e+04	0.000	.0242974	.0243018
_cons	.5515689	3.42e-07	1.6e+06	0.000	.5515682	.5515695
eq2						
popst	.2195964	5.54e-07	4.0e+05	0.000	.2195953	.2195975
distst	-.0824088	5.06e-07	-1.6e+05	0.000	-.0824098	-.0824078
ally	.1136725	7.55e-07	1.5e+05	0.000	.1136711	.113674
d_hl1	-1.430108	3.80e-07	-3.8e+06	0.000	-1.430109	-1.430107
d_hl2	.8476907	1.03e-06	8.2e+05	0.000	.8476887	.8476927
d_hh1	1.321034	7.80e-07	1.7e+06	0.000	1.321032	1.321035
d_hh2	1.100418	.	.	.	.	.
d_hh3	1.069815	9.73e-07	1.1e+06	0.000	1.069813	1.069817
d_hh4	1.034751	1.62e-06	6.4e+05	0.000	1.034748	1.034754
_cons	.788803	3.73e-07	2.1e+06	0.000	.7888023	.7888037
eq3						
_cons	-2.024326	6.87e-07	-2.9e+06	0.000	-2.024327	-2.024325
eq4						
_cons	-1.308826	.	.	.	.	.
eq5						
_cons	12.10166	.	.	.	.	.

11. Section 7 MLE Outcome 2: The alternative payoff function specification

```
initial:       log likelihood =    -2636.5
alternative:   log likelihood = -2000.9978
rescale:       log likelihood = -2000.9978
rescale eq:    log likelihood = -818.28736
Iteration 0:   log likelihood = -818.28736  (not concave)
Iteration 1:   log likelihood =  784.33598  (not concave)
Iteration 2:   log likelihood =  1188.7028  (not concave)
Iteration 3:   log likelihood =  1379.1183
Iteration 4:   log likelihood =  1431.7701
Iteration 5:   log likelihood =  1457.5852
Iteration 6:   log likelihood =  1457.9218
Iteration 7:   log likelihood =  1457.9219

                                                Number of obs     =      1,703
                                                Wald chi2(6)      =    1961.21
Log likelihood =  1457.9219                     Prob > chi2       =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
eq1						
popst	.2233922	.0198435	11.26	0.000	.1844996	.2622848
distst	-.0481181	.0181407	-2.65	0.008	-.0836732	-.0125631
ally	.0349932	.0284222	1.23	0.218	-.0207133	.0906996
D_l11	1.346975	.0581297	23.17	0.000	1.233043	1.460907
D_l0h1	-.5473565	.0420054	-13.03	0.000	-.6296856	-.4650275
num_l0hx	.0651782	.0155124	4.20	0.000	.0347745	.0955819
_cons	.6428692	.0351415	18.29	0.000	.573993	.7117454
eq2						
popst	.2007234	.0195296	10.28	0.000	.1624461	.2390008
distst	-.1032962	.0174236	-5.93	0.000	-.1374458	-.0691466
ally	.1810845	.0272229	6.65	0.000	.1277286	.2344403
D_hh1	1.002749	.0296321	33.84	0.000	.944671	1.060827
D_hh2	1.310417	.0324606	40.37	0.000	1.246795	1.374039
D_h0l1	-.6011571	.0418088	-14.38	0.000	-.6831009	-.5192134
num_h0lx	.2873408	.0642783	4.47	0.000	.1613578	.4133239
_cons	.8091657	.0194906	41.52	0.000	.7709648	.8473667
eq3						
_cons	-1.608228	.0483096	-33.29	0.000	-1.702913	-1.513543
eq4						
_cons	-1.621652	.0413212	-39.25	0.000	-1.70264	-1.540664
eq5						
_cons	.7226988	.0461258	15.67	0.000	.6322939	.8131038

References:

1) Boguslaski, C., Ito,H., and Lee, D. "Entry Patterns in the Southwest Airlines Route System", *Review of Industrial Organization* 25 (2004), 317-350.
2) Bresnahan, T. and Reiss, P. "Entry and Competition in Concentrated Markets", *Journal of Political Economy* 99 (1991), 977-1009.
3) Goetz, C. and Shapiro, A. "Strategic Alliance as a Response to the Threat of Entry: Evidence from Airline Codesharing", *International Journal of Industrial Organization* 30 (2012), 735-747.
4) Mazzeo, M. "Product Choice and Oligopoly Market Structure", *RAND Journal of Economics* 30 (2002), 221-242.
5) Sinclair, A. "An Empirical Model of Entry and Exit in Airline Markets", *Review of Industrial Organization* 10 (1995), 541-557.
6) U.S. Department of Transportation, Bureau of Transportation Statistics, *Data and Statistics: Carrier Snapshot*, 2016 https://www.transtats.bts.gov/carriers.asp
7) Ibid., https://www.transtats.bts.gov/ Only the top 10 domestic routes are listed.
8) Ibid., Origin and Destination Survey: T-100 Domestic Market (U.S. Carriers) https://www.transtats.bts.gov/Fields.asp?Table_ID=258 or https://www.transtats.bts.gov/Tables.asp?DB_ID=110
9) Ibid., Origin and Destination Survey: DB1B Market https://www.transtats.bts.gov/Fields.asp?Table_ID=247
10) US Census Bureau, Population Estimates, *Metropolitan and micropolitan statistical area population and estimated components of change* (CBSA-EST2016-alldata). https://www.census.gov/data/datasets/2016/demo/popest/total-metro-and-micro-statistical-areas.html
11) Kreps, D. and Wilson, R. "Reputation and Imperfect Information", *Journal of Economic Theory* 27 (1982), 253-279.