Network Competition in the Airline Industry: An Empirical Framework*

Zhe Yuan† and Panle Jia Barwick‡

August 23, 2024

Abstract

The Hub-and-Spoke network is a defining feature of the airline industry. This paper is among the first in the literature to introduce an empirical framework for analyzing network competition among airlines. Airlines make market entry decisions and choose flight frequencies in the first stage, followed by price competition to attract passengers in the second stage. A key feature of this model is the linkage between direct and indirect flights, which is described by a technological relationship (and estimated using data) that proxies the Hub-and-Spoke network. The paper estimates the marginal costs of serving passengers and operating flights using first-order conditions, bounds the entry costs using inequalities derived from the reveal-preference argument, and employs a state-of-the-art econometric method to conduct inference for entry cost parameters. Ignoring network externality underestimates the benefits of operating an additional flight by 13.2%, and airlines would schedule 21.53% fewer one-stop flights had they made flight operation decisions independently for each market. To evaluate the impact of a hypothetical merger, the paper proposes a novel equilibrium concept that makes it feasible to compute the industry equilibria. Counterfactual analyses indicate that a hypothetical merger between Alaska

*We thank the editor and anonymous referees whose comments significantly improved this paper. We are grateful to Victor Aguirregabiria, Heski Bar-Isaac, Dwayne Benjamin, Loren Brandt, Yanyou Chen, Andrew Ching, Rahul Deb, Mara Lederman, Yao Luo, Ismael Mourifié, Rob Porter, James Roberts, Marc Rysman, Xianwen Shi, Xiaoxia Shi, Eduardo Souza-Rodrigues, Junichi Suzuki, Yuanyuan Wan, Miaojun Wang, Mo Xiao, Daniel Xu, and seminar participants at Albany, Renmin University, Ryerson, Saskatchewan, Toronto, Western, Zhejiang University, CEA Annual Conference 2016, IIOC Annual Conference 2016, SHUFE IO Conference, EARIE Annual Conference 2016, and China Meeting for Econometric Society 2017 for helpful comments and discussions. We thank Siyu Chen and Jianjuan Chen for their computational support. All errors are our own.

†School of Economics, Zhejiang University, Hangzhou, China 310000. Email: yyyuanzhe@gmail.com.

‡Department of Economics, University of Wisconsin-Madison, NBER, and CEPR. Email: pbarwick@wisc.edu.

and Virgin America would *increase* consumer surplus as the merged airline would offer direct flights in 10% more markets while the overall post-merger price effect would likely be muted.

Keywords: Network competition, Network externality, Airline industry, Entry model, Moment inequality

JEL: C51, L13, L14, L93

1 Introduction

The Hub-and-Spoke network is a defining characteristic of the airline industry. After the deregulation in the 1970s, all major U.S. carriers adopted this network model, allowing airlines to expand their direct flight operations to connecting services and serve more markets. It is easy to see the benefits of a hub-and-spoke network: connecting all C cities with direct flights requires $\frac{C \times C-1}{2}$ nonstop services, whereas using a hub that is connected to all cities would require only C nonstop services. This interconnection between direct and indirect flights through a hub creates what is known as network externality, a key feature of the airline industry that has been analyzed theoretically (Hendricks et al., 1997, 1999) and highlighted in DOJ's report (*U.S. v. Northwest and Continential*, 1998). As of the first quarter of 2014, the number of markets served by one-stop service was four times greater than those served by direct flights. Additionally, one-stop service accounted for 16.4% of the total passengers and contributed 18.8% of the industry's revenue. This network structure is not only prevalent among legacy carriers but also has been adopted by low-cost carriers such as JetBlue and Frontier.

Despite the importance of the hub-and-spoke networks, most empirical studies on the airline industry abstract away from the network externality, often treating flight frequency decisions in one market as independent of those in other markets. This paper addresses this gap by introducing an empirical framework that analyzes network competition among airlines. A key feature of this model is the linkage between direct and indirect flights – an airline can offer indirect flights only when it operates two direct flights sharing a common endpoint. This linkage, described by a technological relationship and estimated from data, serves as a proxy for the hub-and-spoke network. Airlines compete in two stages. In the first stage, i.e., the network formation stage, airlines choose the set of markets in which they will operate direct flights and determine flight frequencies for every nonstop service. These decisions dictate the number of connecting flights offered according to the technological relationship. In the second stage, airlines compete on prices in all markets, given their network structures and flight frequencies.

This paper aims to answer two research questions. First, how do network externalities shape the airline industry's network structure? Will network externalities allow airlines to serve more markets that would otherwise be unprofitable on their own? Second, in the context of an airline merger, how do network externalities affect the optimal network structure of the merged airline? Does the expansion of the network following a merger lead the newly formed airline to enter more markets, thereby alleviating the merger's pricing pressure and potentially delivering consumer welfare gains?

To answer these questions, this paper utilizes three datasets on the U.S. airline industry: the DB1B database on passenger prices and quantities, the OAG databases on flight schedules, and an airport gate database on airline gate allocations that we constructed. The empirical findings suggest that an additional nonstop flight generates an average variable profit of $\$6,729$ from nonstop services and $\$1,310$ from one-stop services. The total marginal variable profit from operating an additional flight is $7,753, after factoring in cannibalization effects. Notably, ignoring network externalities would underestimate the benefits of operating an additional flight by 13.2%. Had airlines chosen flight frequencies independently across markets and ignored the benefits generated by adding non-stop services in one market on other markets, they would have offered 3.17% fewer direct services and strikingly 21.53% fewer connecting services, resulting in a 2.78% reduction in consumer surplus.

A hypothetical merger between Alaska Air Group and Virgin America in the first quarter of 2014 would allow the newly merged airline to offer nonstop services in 10% more markets and increase the number of direct flights by 3.4%. The average local market concentration, measured by the Herfindahl-Hirschman Index (HHI) in revenue shares, would decrease slightly as the network expansion by Alaska and Virgin America more than offset the reduction in competition as a result of the merger. Aggregate consumer surplus would *increase* by 0.56%, driven by improved product quality from new or more frequent services in certain markets and overall muted price responses.

Computational and econometric challenges Conducting these empirical analyses presents a series of computational and econometric challenges. First, the number of potential network configurations increases exponentially with the number of markets and airlines.[1] Existing methods in the literature, such as order-of-move assumptions (Berry, 1992), bounding choice probabilities (Ciliberto and Tamer, 2009), or leveraging supermodularity properties Jia (2008)), are inadequate to address this curse of dimensionality. Instead of directly solving the industry equilibrium to

[1]With a total of C cities and K airlines, the number of possible network configurations is $2^{K\times C\times C/2}$.

estimate entry cost parameters, this paper follows Ho and Pakes (2014), Pakes et al. (2015) and Houde et al. (2023) and estimates entry cost parameters by exploiting the inequality restrictions implied by airlines' best responses in the entry game (the revealed-preference argument). If an airline operates in a market, its entry cost must be (weakly) lower than the variable profit from serving that market. This delivers an upper bound on the entry cost. Conversely, if an airline does not enter a market, the entry cost must be (weakly) higher than the counterfactual variable profit from serving the market. This generates a lower bound on the entry cost. The set of estimated entry cost parameters includes all cost vectors that satisfy these inequalities.

The second challenge arises because constructing the confidence region for parameters estimated through inequalities is non-standard. We draw on the recent moment inequality literature and construct the confidence region by collecting all parameter values for which we fail to reject the hypothesis that the moment inequalities described above are violated (see Molinari (2020) for an excellent review). Specifically, we adopt the method proposed by Cox and Shi (2023), a state-of-the-art approach that is computationally much faster than alternative solutions. It is based on a likelihood ratio test and uses the critical value from a conditional chi-squared distribution with a known degree of freedom, thereby avoiding the need to compute the critical value for each tested parameter vector.

Third, counterfactual analyses require solving for the industry equilibrium, which is computationally infeasible. We propose an equilibrium concept that is the same as the Subgame Perfect Nash Equilibrium (SPNE) except that we restrict airlines' action space to $k-$market deviations, which we term as the level-k "restricted" Subgame Perfect Nash Equilibrium ($k-$RNE). It is a weaker equilibrium concept (there could be many local RNE's for any given SPNE), but is computationally feasible because it drastically reduces airlines' action space. We explain how to compute $k-$RNEs in Section 6.

Related Literature This paper builds on and contributes to three strands of literature. First, it belongs to a growing literature on the estimation of entry games with network competition. Seim (2006) studies spatial competition in the video rental industry. Zhu and Singh (2009) employ a more flexible model of spatial competition and allow for more heterogeneity across firms. Ellickson et al. (2013) and Aguirregabiria et al. (2016) estimate network effects in retail chains and the banking industry, respectively. These papers focus on economies of density in a given market: operating more stores in a market delivers higher profits. In contrast, the network externality in the airline industry is about the economies of scope across markets: operating two direct flights in adjacent markets allows an airline to operate an indirect flight in a third market at zero cost. This type of network effect generated by the hub-and-spoke structure makes airline profit *across markets* interconnected.

The paper is also related to empirical studies of the airline industry. Many papers have examined the advantages of airline hubs, including cost efficiency (Berry, 1990, 1992; Brueckner and Spiller, 1994; Berry et al., 2006; Ciliberto and Tamer, 2009), demand benefits (Berry, 1990; Berry and Jia, 2010), and strategic entry deterrence (Hendricks et al., 1997, 1999; Aguirregabiria and Ho, 2012). However, these papers abstract away from the key feature of a hub-and-spoke network where airlines offer connecting services through combining two direct flights. Li et al. (2022) model the choice between nonstop and one-stop services but take one-stop service as exogenous. This paper endogenizes flight frequency decisions for all carrier-market combinations; in addition, an airline's entry decision in one market depends on its entry decisions in other markets. Bontemps et al. (2023) allow airlines to choose which set of markets to enter but they do not model flight frequency decisions. To the best of our knowledge, this is one of the first papers to propose and estimate a model of airline competition that endogenizes the hub-and-spoke structure and accounts for its network externality.[2]

Finally, this paper contributes to the literature on merger analysis. Nevo (2000) studies the price effect of a merger in the ready-to-eat cereal industry. Fan (2013) allows for changes in product characteristics in addition to prices after newspaper ownership consolidation. There is also an extensive literature on mergers in the airline industry. For instance, Richard (2003) finds that the welfare effect of mergers varies by market, Peters (2006) compares simulation predictions with actual post-merger prices for six airline mergers in the 1980s, and Ciliberto et al. (2019) use the graph theory's centrality measure to study the effect of airline consolidation. The closest research to the present paper is Benkard et al. (2010), which estimates dynamic changes in the airline industry after mergers. Their merger analysis is based on the simulation of policy functions (choice probabilities) and assumes that firm strategy functions do not change after the merger. Our merger simulations start from model primitives and endogenize the (entire) airline networks, in contrast to previous studies.

The remainder of the paper is organized as follows. Section 2 describes the data and construction of key variables, such as flight frequency measures. Section 3 introduces a model of airline

[2] A 2018 version of this paper can be found here https://papers.ssrn.com/abstract=3246222.

competition featuring the hub-and-spoke structure. Section 4 states the empirical strategies, Section 5 presents estimation results, and Section 6 discusses the counterfactual analysis. Section 7 summarizes and concludes the paper.

2 Data

The empirical analyses exploit three databases: the Data Bank 1B (DB1B), the Official Airline Guide (OAG) databases, and an author-constructed airport gate database. DB1B is part of TranStats, the Bureau of Transportation Statistics (BTS) online collection of databases, containing a 10% sample of all U.S. domestic tickets with information on prices and the number of travelers. The OAG database collects all domestic flight schedules, which we use to construct direct and indirect flight frequency measures. The airport gate database provides detailed information on all airport gate usage in 2014, which is compiled from daily domestic flight departure and arrival gate information collected from the flight statistics website http://www.flightstats.com/. We use this database to determine whether a gate is exclusively used by one airline or shared among multiple airlines. A gate is classified as belonging to a specific airline if 80% or more of the flights departing from it are operated by that airline; otherwise, it is a common-use gate. Gate allocation is an important factor that affects airline entry decisions: airlines rarely enter markets where they do not already serve both end cities (Berry, 1992).

We focus on the 100 busiest airports in the continental U.S. and aggregate them into 87 Metropolitan Statistical Areas (MSAs). For simplicity, we do not distinguish between different airports in the same MSA and use MSAs and cities interchangeably.[3] With 87 MSAs (or nodes), there are $M = \frac{C\times(C-1)}{2} = 3741$ markets in the network.

The sample spans from the first quarter of 2007 to the fourth quarter of 2014 for a total of 32 quarters. All prices are deflated and expressed in the constant 2007 dollars. There were 12 major airlines operating in the first quarter of 2007. Virgin America entered in 2008. After several mergers, nine major airlines remain at the end of the sample period. The unit of observation is airline-quarter-market, with a total of 329,448 observations in the sample. Appendix A summarizes various mergers and provides more details on data construction.

[3]Some MSAs have more than one airport. For instance, the Los Angeles MSA has four airports: Los Angeles International Airport (LAX), Long Beach Airport (LGB), Palm Springs International Airport (PSP), and John Wayne Airport in Orange County (SNA). The San Francisco MSA has two airports: San Francisco International Airport (SFO) and Oakland International Airport (OAK). We count the number of direct flights from Los Angeles to San Francisco but do not distinguish between the specific departure/arrival airports.

Table 1 provides a detailed comparison of the number of nonstop and one-stop markets served by each airline, along with the corresponding market share and the percentage of revenue generated from these services in Q1 2007 and Q1 2014. Legacy carriers (which are American Airlines, Continental Airlines, Delta Air Lines, Northwest Airlines, United Airlines, and U.S. Airways) operated direct flights in around 200 direct markets and offered one-stop services in thousands more in 2007. While the majority of passengers travel on non-stop flights, one-stop services still contribute a nontrivial portion of airline revenue. For instance, in Q1 2007, Delta Air Lines earned 24.7% of its domestic revenue from one-stop services. Southwest, known for its point-to-point business model, provided one-stop services to 9% of its passengers and brought in 13.1% of its revenue from these services. On average, airlines served 13.2% of their passengers through one-stop services, generating 15.6% of their total revenue. The statistics for 2014Q1 show similar patterns: one-stop services constitute an essential portion of airline operations and revenues and should not be ignored.

2.1 Hub Indices and Hub Cities

One-stop services, or connecting services in general, are predominately operated through the hub-and-spoke networks, a hallmark of the airline industry. We define the hub index of airline n at city i in quarter t, H_{nit}, as the number of nonstop markets connected to city i in quarter t. The hub index serves two roles in the analysis. First, the hub index provides a direct way to identify an airline's hub cities. Specifically, if airline n has a hub index above 20 in city i throughout the sample period, city i is classified as a hub for airline n. Over 90% of one-stop passengers connect through these hub cities in our sample. Second, it captures heterogeneity in consumer demand and operating costs across airlines for a given market. Airlines usually have higher demand (driven by convenience factors and loyalty programs, etc.) and lower costs in their hub cities (due to the economies of scale).

Table 2 presents the top two hubs for each airline at the beginning and the end of the sample period. The primary hubs of legacy carriers offered direct services to over 60 cities. Delta Air Lines, for instance, operated 77 and 81 nonstop flights from Atlanta, its largest hub, in Q1 2007 and Q1 2014, respectively. Some airlines expanded their hub operations over time. For example, Frontier Airlines increased its nonstop services from Seattle from 22 destinations in 2007 to 30 in 2014, while Southwest expanded its nonstop markets from Chicago from 45 cities to 58 during the same period.

2.2 Measuring Flight Frequencies

Direct Flight Frequency Direct flight frequency is defined as the average number of daily nonstop flights an airline offers between two cities in a quarter. There are $56,097$ unique airline-quarter-nonstop markets. Panel (a) of Figure 1 depicts the histograms of daily direct flight frequencies across markets. On average, airlines operate 7.7 daily direct flights per market, with a median of 5.7 flights. Southwest operates the busiest nonstop market from Southern California (Los Angeles-Long Beach-Anaheim) to the Bay Area (San Francisco-Oakland-Hayward) with over 108 daily flights. On average, nonstop markets accommodate 240 passengers daily, with a median of 131 passengers.

Indirect Flight Frequency As with direct flights, the frequency of indirect flights is a key quality measure of one-stop services. We define a (feasible) indirect flight service from city A to city B with a layover at city H if the following three conditions are satisfied: (a) airport H is a hub city, (b) the total distance of the indirect flight is less than 1.5 times the distance of the direct flights ($dist_{AH} + dist_{HB} < 1.5 \times dist_{AB}$), and (c) the scheduled departure time of the second flight is between 45 minutes and 4 hours after the scheduled arrival time of the first flight.[4] When multiple flights in market AH can connect with flights in market HB, the indirect flight frequency is the total number of feasible indirect connections. These discussions highlight an indirect benefit of adding nonstop services: expanding direct flight options can create more feasible indirect connections, thereby increasing total revenue.

There are 273,351 airline-quarter-markets with indirect services. Panel (b) of Figure 1 plots the distribution of indirect flight frequencies. The average indirect flight frequency is 10.5, with a median of eight. Airlines transport nine one-stop passengers daily between two cities, with a median of three.

3 Model

This section introduces a static model that endogenizes airlines' entry, flight frequency, and pricing decisions. We use matrices to represent these decisions, following the network literature (Jackson

[4]This connecting time threshold of 45 minutes to 4 hours follows Molnar (2013). The minimum time for domestic connections is usually 45 to 75 minutes, and the maximum time for domestic connections is typically 4 hours. Without these restrictions, the number of indirect flights will blow up and introduce two sources of bias. First, it overstates the number of indirect flights, especially at the large hubs, as few passengers would consider flights with layovers exceeding 4 hours. Second, not limiting travel distance would overestimate indirect flight options in the coastal regions. Both would bias coefficient estimates of flight frequency in passenger utility and marginal cost.

and Wolinsky, 1996). This paper is among the first to endogenize the airline's network structure and utilize matrix notation to describe airline networks, offering a streamlined method to record the network structures of airlines.

The model incorporates two major innovations. First, it includes a technological relationship that captures the formation of indirect flights from nonstop services, which is a critical source of network externalities in the airline industry's hub-and-spoke structure. Second, the model decomposes the marginal profit of adding a direct flight into four distinct channels, highlighting explicitly the impact on all relevant markets and the scope of network externalities.

The model has two stages. In stage one, carriers select the set of markets in which to offer non-stop services, which determines their network configuration. They also choose the number of nonstop flights in each market. These decisions are made simultaneously across carriers. In stage two, airlines compete on prices to attract passengers. We describe the model setup in Section 3.1, explain the creation of indirect services from direct services in Section 3.2, and present the timeline, information structure, and the equilibrium concept in Section 3.3. The model primitives include consumer preferences, airlines' variable costs of serving passengers and operating flights, and their fixed costs of managing gates and airport presence in each market. While our sample has a panel structure, we suppress the time subscript t throughout Sections 3 and 4 for notational simplicity.

3.1 Model Settings

The commercial aviation industry is characterized by N airlines (indexed by n) and C cities (denoted by i or j). A market is a non-directional city-pair in which airlines provide regular aviation services.[5] With C cities, there are a total of $M = \frac{C\times(C-1)}{2}$ markets, which are indexed by ij, with i and j representing the two endpoint cities.

Airlines offer two types of services (products) in a market: nonstop service, and one-stop service with a layover/transfer in a third city. We follow Aguirregabiria and Ho (2012) and ignore services with more than one stop, which comprise less than 3% of air travel. For notational simplicity, let subscript g represent a product, which is defined by the following triplet: (1) airline n, (2) market ij, and (3) nonstop indicator x ($x = 1$ for nonstop services).

[5] A market is a non-directional city pair where airlines transport passengers from one city to the other. Technically, "direct" means that passengers do not change planes between origin and destination, whereas "nonstop" means that the flight does not stop between origin and destination. We use these two terms interchangeably. The definition of markets as city pairs follows Berry (1992); Berry et al. (2006); Aguirregabiria and Ho (2012) and ignores competition between airports within a city. Borenstein (1989) and Ciliberto and Tamer (2009) define markets as airport pairs. We assume markets are non-directional for simplicity and do not distinguish flights from i to j separately from flights from j to i, though the model can be extended to accommodate directional markets.

Demand The demand model follows the classic discrete-choice literature.[6] Passengers value frequent services because they provide more flexible departure times and more connecting possibilities. Let f_g denote flight frequency, $b(f_g, \xi_g; \alpha)$ denote travelers' willingness to pay for product g (which depends on quality f_g, demand shock ξ_g and preference parameters α), and p_g denote price. The indirect utility of traveler ι purchasing product g is:

$$U(f_g, p_g, \xi_g, v_{\iota g}; \alpha) = b(f_g, \xi_g; \alpha) - p_g + v_{\iota g}, \tag{1}$$

where $v_{\iota g}$ is the traveler-product specific random shock. We allow $v_{\iota g}$ to have an arbitrary variance and normalize the price coefficient to one. This specification is isomorphic to one that estimates a price coefficient but normalizes the variance of $v_{\iota g}$, as the scale of the utility function in discrete-choice models is undefined. The utility of the outside good (not traveling or taking an alternative form of transportation) is normalized to zero ($U_{\iota 0} = 0$).

Let MS_{ij} denote the total number of potential travelers (market size) in market ij. Travelers observe flight frequencies ($\mathbf{f}$), prices ($\mathbf{p}$), and demand shocks (ξ) of all products in market ij and choose one product (or the outside alternative) to maximizes their utility. We integrate over the traveler-idiosyncratic component $v_{\iota g}$ to obtain the aggregate demand for product g:

$$\underbrace{q_g(\mathbf{f}_{ij}, \mathbf{p}_{ij}, \xi_{ij}; \alpha)}_{\text{Demand for Product } g} = \underbrace{MS_{ij}}_{\text{Market Size}} \times \int \mathbf{1}\left[\underbrace{U(f_g, p_g, \xi_g, v_{\iota g}; \alpha)}_{\text{Utility from Product } g} \geq \underbrace{U(f_{g'}, p_{g'}, \xi_{g'}, v_{\iota g'}; \alpha)}_{\text{Utility from Product } g'}, \forall g' \neq g\right] d\mathbf{v}_\iota,$$

where $\mathbf{f}_{ij}$ and $\mathbf{p}_{ij}$ are vectors of characteristics and prices for all products in market ij, and $\mathbf{v}_\iota$ denotes a vector that contains all product-specific random tastes of individual ι, respectively.

Airline Choices Airline n's choices include its entry ($\mathbf{A}_n$), flight frequencies ($\mathbf{F}_n$), and pricing decisions ($\mathbf{P}_n$) in all markets. Entry decisions are described by a $C \times C$ symmetric matrix $\mathbf{A}_n$, with a_{nij} as its (i, j)-th element. Let $a_{nij} = 1$ if airline n enters market ij, and $a_{nij} = 0$ otherwise. Airline entry decisions determine its network structure, i.e., the set of direct and indirect services it offers across all markets. An airline can only operate direct flights in markets it has entered. However, it can schedule indirect flights when it operates direct flights in two markets that share a common end-point. Mathematically, airline n can provide one-stop service between city-pair ij

[6] Existing literature examining airline demand includes: Berry (1990), Berry et al. (2006), Berry and Jia (2010), Aguirregabiria and Ho (2012), Ciliberto et al. (2021) and Li et al. (2022). Our demand-side analysis follows closely Aguirregabiria and Ho (2012).

with a connection in k if it operates direct flights in both market ik and market kj (i.e., $a_{nik} = 1$ and $a_{nkj} = 1$).

Differences in flight frequencies introduce important quality differentiation across airline-markets.[7] There are two flight frequency measures, non-stop ($\mathbf{F}_n^{NS}$) and one-stop ($\mathbf{F}_n^{OS}$) frequencies, both of which are $C \times C$ symmetric matrices, where the (i,j)-th elements, f_{nij}^{NS} and f_{nij}^{OS}, are the number of direct and indirect flights airline n operates in market ij. Note that $a_{nij} = 0$ iff $f_{nij}^{NS} = 0$. Once airlines decide on direct flight frequencies, indirect flight frequencies are determined according to a technological relationship that we describe in Section 3.2.

The pricing decisions $\mathbf{P}_n$ are denoted by two $C \times C$ symmetric matrices $\mathbf{P}_n^{NS}$ and $\mathbf{P}_n^{OS}$, with P_{nij}^{NS} and P_{nij}^{OS} as the (i,j)-th elements, respectively. Given that there are N airlines competing, we use arrays $\mathbf{A}$, $\mathbf{F}$, $\mathbf{P}$ to describe the network structure, flight frequencies, and prices for the entire industry (airline n and its competitors): $\mathbf{A} = \{\mathbf{A}_n, \mathbf{A}_{-n}\}$, $\mathbf{F} = \{\mathbf{F}_n, \mathbf{F}_{-n}\}$, and $\mathbf{P} = \{\mathbf{P}_n, \mathbf{P}_{-n}\}$.

Profit and Costs Airline n's total profit consists of the variable profit of serving passengers minus the costs of operating flights and the costs of maintaining gates and airport presence in all markets:

$$\underbrace{\Pi_n(\mathbf{A}, \mathbf{F}, \mathbf{P}, \xi, \omega, \varepsilon, \kappa; \alpha, \delta, \gamma, \eta)}_{\text{Total Profit}} = \underbrace{\pi_n(\mathbf{A}, \mathbf{F}, \mathbf{P}, \xi, \omega; \alpha, \delta)}_{\text{Total Variable Profit}} - \underbrace{\Gamma_n(\mathbf{A}_n, \mathbf{F}_n^{NS}, \varepsilon; \gamma)}_{\text{Flight Frequency Cost}} - \underbrace{FC_n(\mathbf{A}_n, \kappa; \eta)}_{\text{Total Fixed Cost}}.$$

We explain each item in detail below.

Total Variable Profit The total variable profit, π_n, is the sum of variable profits from both non-stop and one-stop services across all markets:

$$\underbrace{\pi_n(\mathbf{A}, \mathbf{F}, \mathbf{P}, \xi, \omega; \alpha, \delta)}_{\text{Total Variable Profit}} = \frac{1}{2} \sum_i \sum_{j \neq i} \sum_{x \in \{NS, OS\}} \underbrace{\pi_{nij}^x(\mathbf{f}_{ij}, \mathbf{p}_{ij}, \xi_{ij}, \omega_{ij}; \alpha, \delta) \times a_{nij}^x}_{\text{Variable Profit of Service } x},$$

where a_{nij}^x is a dummy variable indicating whether airline n offers service x (nonstop or one-stop) in market ij, and $\mathbf{f}_{ij}$ and $\mathbf{p}_{ij}$ are vectors of flight frequencies and prices for all nonstop and one-stop

[7] We ignore heterogeneity in aircraft type and flight departure (arrival) times. Williams (2008) models available seats as a capacity choice instead of a flight frequency decision and allows airlines to invest in capacity to lower variable costs.

services in market ij, respectively. The variable profit from service x in market ij is:

$$\underbrace{\pi^x_{nij}(\mathbf{f}^x_{ij}, \mathbf{p}^x_{ij}, \xi_{ij}, \omega_{ij}; \alpha, \delta)}_{\text{Variable Profit of Service } x} = p^x_{nij} \times q^x_{nij}(\mathbf{f}_{ij}, \mathbf{p}_{ij}, \xi_{ij}; \alpha) - VC^x_{nij}(f^x_{nij}, q^x_{nij}, \omega_{ij}; \delta), \tag{2}$$

where VC^x_{nij} is airline n's variable cost of serving passengers on the plane for either nonstop or one-stop services in market ij. It depends on flight frequency (f), the number of passengers served (q), variable cost shock (ω), and variable cost parameters (δ).

Airlines' variable profits from serving passengers depend on flight frequencies through several channels. As described above, demand for both direct and indirect services increases with flight frequency. In addition, the costs of serving passengers also depend on flight frequencies (in addition to the costs of operating flights). When flight frequency is high, many seats are likely to empty, and the marginal cost of serving passengers is low. Conversely, when flight frequency is low, the opportunity cost of serving an additional passenger is high. In the extreme case of a fully booked plane, the airline's marginal cost of serving an additional passenger could be the cost of adding another flight because there is no empty seat for the passenger.

Flight Frequency Costs The total flight operating costs consist of the expenses incurred in operating all *nonstop* flights across the markets served by airline n. Note that one-stop flight services do not entail additional operating costs. When an airline operates two nonstop flights with a common endpoint, it can combine these two flights (segments) to offer a one-stop service at no extra costs. This is the source of the network externality in this industry. The total flight operating costs are defined as:

$$\underbrace{\Gamma_n(\mathbf{A}_n, \mathbf{F}^{NS}_n, \varepsilon_n; \gamma)}_{\text{Flight Frequency Costs}} = \frac{1}{2}\sum_i \sum_{j \neq i} \underbrace{\Gamma_{nij}(f^{NS}_{nij}, \varepsilon_{nij}; \gamma) \times a_{nij}}_{\text{Flight Frequency Cost in Market } ij} .$$

where Γ_{nij} denotes airline n's cost of operating nonstop flights in market ij. It depends on the flight frequency of the nonstop service f^{NS}_{ij}, an operating cost shock ε_{nij}, and cost parameters γ.

Fixed Cost In addition to the costs of serving passengers and operating flights, airlines pay fixed costs to enter into a market (gates fees and maintenance costs for airport presence, etc.). Airline

n's total entry cost sums up its fixed costs in all markets:

$$\underbrace{FC_n(\mathbf{A}_n, \kappa_n; \eta)}_{\text{Total Fixed costs}} = \frac{1}{2} \sum_i \sum_{j \neq i} \underbrace{FC_{nij}(\kappa_{nij}; \eta) \times a_{nij}}_{\text{Fixed Cost in Market } ij}.$$

where FC_{nij} denotes airline n's entry cost into market ij, which depends on entry cost shock (κ_{nij}) and fixed cost parameters (η).

3.2 Technological Relationship Defining Indirect Flight Frequencies

A key source of network externality in the airline industry is carriers' ability to offer connecting services via non-stop flights. This section discusses how the flight frequencies of one-stop services are determined. Specifically, the frequency of airline n's indirect flight between cities i and j with a connection in k ($f_{nij}^{OS(k)}$) depends on the direct flight frequencies on its two segments that comprise the one-stop route: f_{nik}^{NS} and f_{nkj}^{NS}:

$$\underbrace{f_{nij}^{OS(k)} = \Lambda_{nij}^{(k)} \left(f_{nik}^{NS}, f_{nkj}^{NS} \right),}_{\text{Technological Relationship in a Market}}$$

where $\Lambda_{nij}^{(k)}$ represents the technological relationship that links direct and indirect flight frequencies between cities i and j with a connection in k.[8] This function is specific for each airline-market-connection city combination because airlines may employ different schedules or connection technologies across different markets and connection cities. The total indirect flight frequency for airline n's indirect flight service in market ij is the sum of the indirect flight frequencies between cities i and j across all feasible connecting cities:

$$f_{nij}^{OS} = \sum_k f_{nij}^{OS(k)}.$$

Airline n's indirect flight frequencies across all markets are denoted by a $C \times C$ symmetric matrix, $\mathbf{F}_n^{OS}$, with its (i, j)-th element being f_{nij}^{OS}:

$$\underbrace{\mathbf{F}_n^{OS} = \Lambda_n(\mathbf{F}_n^{NS})}_{\text{Technological Relationship in a Network}},$$

[8]Section 2.2 describes how we measure the flight frequencies for indirect services. We specify the functional form of $\Lambda_{nij}^{(k)}$ in Section 5.1.

where function $\Lambda_n = \{\sum_k \Lambda_{nij}^{(k)},\ \forall i, j\}$ summarizes the technological relationship between airline n's indirect and indirect flight frequencies across its entire operation network.

A key characteristic of this technological relationship is that, given a total of C cities, a change in flight frequency in market ij could potentially affect up to $2 \times C - 4$ one-stop markets that include ij as part of their services. Specifically, there are $C - 2$ one-stop routes from city i to other cities with a connection at city j, and an additional $C - 2$ one-stop routes from city j to other cities with a stop at city i. This is the network externality that our study focuses on in this paper.

3.3 Timeline, Information Structure, and Equilibrium Concept

Timeline and Information Structure The model is a two-stage game. In **Stage one**, airlines observe profit shifters, the fixed costs associated with entering specific markets, and variable costs related to flight operations. These include the entry cost shock κ_n and flight operating cost shock ε_n for all airlines, which are unobservable to the econometrician. Airlines form expectations over the distribution of demand shocks ξ and variable cost shocks of serving passengers ω. All carriers simultaneously decide on the set of markets where they offer direct services ($\mathbf{A}$) and the corresponding flight frequencies in each market entered ($\mathbf{F}$). These joint decisions of all airlines collectively shape the network structure of the airline industry. In **Stage two**, airlines observe the actual consumer demand and variable costs of serving passengers. They also observe the realizations of ξ, ω for all products, which are unobservable to the econometrician. Given the industry's network structure, all airlines compete on prices for nonstop and one-stop services across all markets ($\mathbf{P}$) to attract passengers.

We assume airlines know the distribution of demand and marginal cost shocks (ξ, ω) in stage one but do not observe the realizations of these shocks until after they have entered a market and offered flight services in stage two. This timing assumption follows Aguirregabiria and Ho (2012), Sweeting (2013), and Eizenberg (2014) where firms do not know their products' demand shocks or marginal cost shocks before entering a market.[9]

Equilibrium Concept The equilibrium of this two-stage game is a Subgame Perfect Nash Equilibrium (SPNE). We now describe the best response functions and equilibrium outcomes:

(i) In the first stage, i.e., the network formation stage, given the optimal network structure of

[9] Ciliberto et al. (2021), and Li et al. (2022) estimate demand and entry decisions simultaneously and allow correlated demand and entry cost shocks in a local market. For computational reasons, joint estimation of the demand and market entry in a network model is beyond the scope of this paper.

competitors ($\mathbf{A}^*_{-n}$,$\mathbf{F}^*_{-n}$), cost shocks κ_n and ε_n, and airlines' optimal pricing strategies in the second stage ($\mathbf{P}^*(.)$), airline n makes the entry ($\mathbf{A}_n$) and flight frequency decisions ($\mathbf{F}_n$) to maximize its expected profit:

$$\{\mathbf{A}^*_n, \mathbf{F}^*_n\} = \underset{\mathbf{A}_n,\mathbf{F}_n}{argmax} \quad \underbrace{E_{\xi,\omega}[\pi_n(\mathbf{A}_n, \mathbf{F}_n, \mathbf{P}^*_n(\xi,\omega); \alpha, \delta | \mathbf{A}^*_{-n}, \mathbf{F}^*_{-n}, \mathbf{P}^*_{-n}(\xi,\omega))]}_{\text{Total Variable Profit}} \qquad (3)$$
$$-\underbrace{\Gamma_n(\mathbf{A}_n, \mathbf{F}^{NS}_n, \varepsilon; \gamma)}_{\text{Flight Frequency Cost}} - \underbrace{FC_n(\mathbf{A}_n, \kappa; \eta)}_{\text{Total Fixed Cost}}.$$

where we omit the arguments $\mathbf{A}$ and $\mathbf{F}$ in the pricing strategies $\mathbf{P}^*(.)$ to simplify notations. At the industry equilibrium, Equation (3) holds for all carriers $n \in N$.

(ii) In the second (pricing) stage, airlines compete on prices in every local market. In market ij, given airlines' flight frequencies ($\mathbf{f}_{ij}$) and competitors' optimal pricing strategy ($\mathbf{p}^*_{-nij}(\xi_{ij}, \omega_{ij})$), airline n's best price response function for both nonstop and one-stop services is:

$$\mathbf{p}^*_{nij}(\xi_{ij}, \omega_{ij}) = \underset{\mathbf{p}_{nij}(\xi_{ij},\omega_{ij})}{argmax} \quad \sum_{x\in\{NS,OS\}} \underbrace{\pi^x_{nij}(\mathbf{f}_{ij}, \mathbf{p}_{nij}(\xi_{ij}, \omega_{ij}); \alpha, \delta | \mathbf{p}^*_{-nij}(\xi_{ij}, \omega_{ij})) \times a^x_{nij}}_{\text{Variable Profit of Service } x}$$

where $\mathbf{p}_{nij}$ is a two-by-one column vector of airline n's nonstop and one-stop prices in market ij and $\mathbf{p}_{-nij}$ includes prices of all other competing products in market ij. Similarly, this optimal pricing equation holds for all airlines $n \in N$ in equilibrium.

We assume the existence of a pure-strategy SPNE $\mathbf{A}^*, \mathbf{F}^*, \mathbf{P}^*$ for the two-stage game specified, but we do not assume the uniqueness and allow for multiple equilibria. The estimation strategy (detailed in Section 4) does not require solving the industry equilibria, thereby circumventing the issue of multiple equilibria. We revisit the equilibrium concept in counterfactual analyses in Section 6 below.

4 Estimation Strategies and Identification

We now outline the parametric assumptions and empirical strategies employed for model estimation. Results are presented in Section 5.1. The exposition of this section is structured in the following order: Section 4.1 explains the technological relationship that determines the flight frequencies for indirect services. Section 4.2 details the estimation of passengers' travel preferences and marginal costs of serving passengers. Section 4.3 leverages the optimality conditions for flight frequencies to recover the flight frequency cost parameters. Section 4.4 uses the reveal-preference argument and moment inequalities to estimate entry costs.

4.1 Technological Relationship for Indirect Flight Frequency

We assume airline n's indirect flight frequency between city i and j with a stop at hub city k is a symmetric Cobb-Douglas function with respect to its direct flight frequencies in market ik and market kj:

$$\underbrace{\ln f_{nij}^{OS(k)}}_{\text{Indirect Flight Frequency}} = h + \lambda \times \underbrace{\left(\ln f_{nik}^{NS} + \ln f_{nkj}^{NS}\right)}_{\text{Direct Flight Frequency}} + \epsilon_{nij}^{k}, \tag{4}$$

where ϵ_{nij}^{k} is assumed to be i.i.d with mean zero. The exact production technology for indirect flight frequencies is complicated. We use a Cobb-Douglas function for two reasons. First, it fits data remarkably well, as shown in Section 5.1. Second, it is straightforward to examine whether there are increasing or decreasing returns to scale with a Cobb-Douglas function.

4.2 Demand and Marginal Costs of Serving Passengers

The empirical specification for the demand model in Equation (1) is a nested logit following Aguirregabiria and Ho (2012). Passengers first decide which airline to travel with and then choose between direct and indirect flights. Passengers' willingness to pay for flight service g is specified as $b_g = \mathbf{W}_g^d \alpha + \xi_g$, where $\mathbf{W}_g^d$ is a vector of regressors that includes a nonstop product dummy, flight frequencies for both nonstop and one-stop services, hub indices for the two endpoint cities, the distance between the two endpoint cities, airline fixed effects, city fixed effects, and quarter fixed effects. The term ξ_g is a product-specific demand shock. The individual error term $v_{\iota g}$ consists of two terms: $v_{\iota g} = \sigma_1 v_{\iota nijt}^{(1)} + \sigma_2 v_g^{(2)}$, where $v_g^{(2)}$ is i.i.d extreme value distributed and $v_{\iota nijt}^{(1)}$ has a distribution such that $v_{\iota g}$ is also extreme value distributed, following Berry (1994) and Cardell (1997). The parameter σ_1 governs the extent of between-group substitutions, while σ_2 is a scaling parameter that can be interpreted as the inverse of price sensitivity since we normalize the price coefficient to 1. Putting everything together, the demand model can be expressed as a system of equations that is linear in parameters:

$$ln(s_g) - ln(s_0) = \frac{b_g}{\sigma_1} - \frac{p_g}{\sigma_1} + (1 - \frac{\sigma_2}{\sigma_1})ln(s_g^*) = \mathbf{W}_g^d \frac{\alpha}{\sigma_1} - \frac{p_g}{\sigma_1} + (1 - \frac{\sigma_2}{\sigma_1})ln(s_g^*) + \frac{\xi_g}{\sigma_1}, \forall g \tag{5}$$

where s_g denote the market share of product g in market ij, i.e., $s_g = q_g/MS_{ij}$ and s_0 denote the market share of the outside option. Market size MS_{ij} is measured by the total population of the origin and destination cities. The second last term on the right-hand-side s_g^* is the within-group market share of product g, where a group consists of the two products offered by airline n in market ij: direct and connecting services.

Both price p_g and (log) within-group market share $ln(s_g^*)$ are endogenous because products with larger demand shocks (ξ_g) are more likely to have higher prices as well as higher within-group market shares. Note that the timing assumption in Section 3.3 implies that demand shock ξ is independent of flight frequency (an element of $\mathbf{W}^d$) since flight frequency is chosen before the realization of ξ. We estimate demand preferences via the Generalized Method of Moments (GMM) and use the characteristics of other products in the same market as instruments for prices and within-group shares, following Berry (1992) and Berry et al. (1995). Specifically, there are seven instrument variables in total: the average flight frequency for both nonstop and one-stop services of the competitors, hub indices for both the origin and destination cities of the competitors, the fractions of competing products providing direct services, and two dummy variables denoting the absence of nonstop and one-stop competing products, respectively. Characteristics of rival products are correlated with airline n's own prices through firms' pricing competition but are uncorrelated with airline n's product-specific demand shocks. Consequently, they are valid instruments for both prices p_g and within-group shares $ln(s_g^*)$.

The total variable cost of serving a passenger on flight g is defined as:[10]

$$VC_g(f_g, q_g, \omega_g; \delta) = c_g(f_g, \omega_g; \delta) \times q_g.$$

where c_g is the marginal cost of serving passengers. It is specified as $c_g(\omega_g; \delta) = \mathbf{W}_g^c \delta + \omega_g$, where $\mathbf{W}_g^c$ is a vector of regressors that includes a nonstop product dummy, flight frequencies, hub indices for the two endpoint cities, the distance between the two endpoint cities, airline $\times$ city fixed effects, and quarter fixed effects. The term ω_g is a product-specific marginal cost shock. We use airline $\times$ city fixed effects (instead of airline fixed effects plus city fixed effects) to control for differences in cost structures across airlines and cities.

Using the demand estimates and under the Nash-Bertrand pricing assumption, we recover the marginal cost of serving passengers as $c_g = p_g - \sigma_1 (1 - \bar{s}_g)^{-1}$, where σ_1 will be estimated in the

[10] By construction, $VC_g \equiv VC_{nij}^x$. We use VC_{nij}^x in Equation (2) in Section 3.1 when summing over markets ij, and VC_g here for notation simplicity.

demand analysis, and $\bar{s}_g = (\sum_{g' \in \mathbf{G}_{nij}} e_{g'})^{\frac{\sigma_2}{\sigma_1}}[1 + \sum_{n'=1}^{N}(\sum_{g' \in G_{n'ij}} e_{g'})^{\frac{\sigma_2}{\sigma_1}}]^{-1}$, with $e_g = exp\{(b - p_g)/\sigma_1\}$ and G_{nij} representing the set of products offered by airline n in market ij. Assuming no further endogeneity issues, we use OLS to estimate marginal cost parameter δ.

4.3 Flight Frequency Costs

The costs of operating flights in a market is a linear function of flight frequency:

$$\Gamma_{nij}(f_{nij}^{NS}, \varepsilon_{nij}; \gamma) = c^f(\varepsilon_{nij}; \gamma) \times f_{nij}^{NS}.$$

where $c^f(\varepsilon_{nij}; \gamma)$ is the marginal cost airline n incurs for scheduling one more direct flight in market ij. It is specified as $c^f(\varepsilon_{nij}; \gamma) = \mathbf{W}_{nij}^f \gamma + \varepsilon_{nij}$, where $\mathbf{W}_{nij}^f$ is a vector of regressors that includes hub indices for the two endpoint cities, the distance between the two endpoint cities, airline $\times$ city fixed effects, and quarter fixed effects. The term ε_{nij} is an airline-market-specific cost shock to operating an additional flight.

We estimate cost parameters γ exploiting the fact that flight frequencies are optimally chosen and hence equate the expected marginal variable profit from an additional direct flight with the marginal cost c^f. The optimality condition for flight frequency is (treating it as differentiable for simplicity):

$$\frac{\partial E_{\xi,\omega}[\pi_n(\mathbf{A}_n^*, \mathbf{F}_n, \mathbf{P}_n^*(\xi,\omega); \alpha, \delta | \mathbf{A}_{-n}^*, \mathbf{F}_{-n}^*, \mathbf{P}_{-n}^*(\xi,\omega))]}{\partial f_{nij}^{NS}} = \underbrace{\frac{\partial \Gamma_{nij}(f_{nij}^{NS}, \varepsilon_{nij}; \gamma)}{\partial f_{nij}^{NS}}}_{=c^f(\varepsilon_{nij};\gamma)}.$$

A change in direct flight frequency in one market creates a chain of reactions and affects an airline's total variable profit in many different markets:

$$\begin{aligned} \frac{\partial E_{\xi,\omega}[\pi_n(.)]}{\partial f_{nij}^{NS}} = & \underbrace{\frac{\partial E_{\xi,\omega}[\pi_{nij}^{NS}(.)]}{\partial f_{nij}^{NS}}}_{\text{(a) Additional Nonstop Service}} + \underbrace{\frac{\partial E_{\xi,\omega}[\pi_{nij}^{OS}(.)]}{\partial f_{nij}^{NS}}}_{\text{(b) Cannibalization on Existing One-stop Services}} \\ & + \underbrace{\sum_{i' \neq i} \frac{\partial E_{\xi,\omega}[\pi_{ni'j}^{OS}(.)]}{\partial f_{nij}^{NS}} + \sum_{j' \neq j} \frac{\partial E_{\xi,\omega}[\pi_{nij'}^{OS}(.)]}{\partial f_{nij}^{NS}}}_{\text{(c) Additional One-stop Services}} + \underbrace{\sum_{i' \neq i} \frac{\partial E_{\xi,\omega}[\pi_{ni'j}^{NS}(.)]}{\partial f_{nij}^{NS}} + \sum_{j' \neq j} \frac{\partial E_{\xi,\omega}[\pi_{nij'}^{NS}(.)]}{\partial f_{nij}^{NS}}}_{\text{(d) Cannibalization on Nonstop Services in markets in (c)}}. \end{aligned} \quad (6)$$

where we suppress the arguments in π for notation simplicity. One additional direct flight in market ij affects airline n's expected variable profit through four channels: (a) it earns more variable profit from nonstop services in market ij; (b) but loses profit due to cannibalization effects on existing one-stop services in market ij; (c) gains variable profit from one-stop services in other markets that now offer more indirect flights as a result of the increase in f_{nij}^{NS}; (d) and loses profit when these additional indirect flights in other markets cannibalize profit from existing nonstop services in the corresponding markets. Channels (b)-(d) potentially affect a large number of markets and capture the extensiveness of network externality in the airline industry. Equation (6) illustrates a distinguishing feature of our network model from models without network effects where channels (c) and (d) are absent because decisions in one local market would not affect other markets.

All components from (a) to (d) can be calculated using demand estimates and the marginal cost estimates of serving passengers in Section 4.2, though two major computational challenges arise. First, adding a direct flight in one market may affect the expected variable profit in as many as $2 \times C - 3$ markets, which is computationally difficult to calculate.[11] Second, the calculation of expected profit is very involved.[12] We utilize parallel computing to expedite this process and use GMM to estimate parameter γ.

4.4 Entry Costs and Moment Inequalities

We assume the costs of entering market ij is linear in airline and market characteristics $\mathbf{Z}_{nij}$ and entry cost shock κ_{nij}:

$$FC_{nij}(\kappa_{nij}; \eta) = \mathbf{Z}_{nij}\eta + \kappa_{nij}$$

The number of airline network structures grows exponentially with the number of markets (2^M). In addition, for a specific network configuration (which carriers operate in which markets), figuring out the optimal flight frequency choices and pricing decisions across the entire network for all carriers is also a daunting task. Consequently, it's impractical to directly solve for the industry equilibria, which presents a mixture of combinatorial problems with a vast number of potential choices and continuous decisions. Therefore, instead of exploring the optimality entry condition for profit maximization in Equation (3), we estimate the entry cost parameters η using a revealed-preference

[11] Channel (a) affects the nonstop service in one market, channel (b) affects the non-stop service in that market, channel (c) affects one-stop service in up to $(C-2) \times 2$ other markets, and channel (d) affects nonstop service in the corresponding markets (up to $(C-2) \times 2$ markets).

[12] To do so, we randomly draw ξ's and ω's for all products $\times$ carriers $\times$ markets 50 times from their empirical distribution. For each set of simulation draws, we solve the optimal prices and calculate the variable profits. Then we take the average over simulation draws to obtain the expected variable profits for all carriers $\times$ markets.

argument and moment inequalities. By the revealed-preference argument, airlines' observed network structure should deliver higher profits than alternative network configurations on average.

Construction of Inequalities Let $\Pi_n(\mathbf{A}_n^*, \kappa; \eta|\mathbf{A}_{-n})$ denote airline n's expected total profit from its observed network structure $\mathbf{A}_n^*$ (where ξ and ω are integrated out and the other arguments except for $\mathbf{A}$ and κ are omitted for notation simplicity). Let $\Pi_n(\mathbf{A}_n^a, \kappa; \eta|\mathbf{A}_{-n})$ denote the expected total profit from an alternative network configuration. By revealed-preference argument:

$$\Pi_n(\mathbf{A}_n^a, \kappa; \eta) - \Pi_n(\mathbf{A}_n^*, \kappa; \eta) = \Delta\Pi_n(\mathbf{A}_n^a, \mathbf{A}_n^*; \eta) + \tau_n^a(\kappa) \leq 0. \tag{7}$$

where $\Delta\Pi_n(\mathbf{A}_n^a, \mathbf{A}_n^*; \eta) = \Delta\Pi_n^1 + \Delta\Pi_n^2$ captures differences in total profit that are unrelated to entry cost shocks. The first part of $\Delta\Pi_n$ is differences in variable profit and flight operating costs (where arguments other than $\mathbf{A}$ and $\mathbf{F}$ are omitted):

$$\Delta\Pi_n^1 = [E_{\xi,\omega}[\pi_n\left(\mathbf{A}_n^a, \mathbf{F}_n^a\right)] - E_{\xi,\omega}[\pi_n\left(\mathbf{A}_n^*, \mathbf{F}_n^*\right)]] - [\Gamma_n\left(\mathbf{A}_n^a, \mathbf{F}_n^a\right) - \Gamma_n\left(\mathbf{A}_n^*, \mathbf{F}_n^*\right)] \tag{8}$$

As explained above, the calculation of these profit deviations is quite involved, as any change in one market creates a chain of adjustments and affects profits in many other markets through network externalities. The second part of $\Delta\Pi_n$ is differences in entry costs:

$$\Delta\Pi_n^2 = \frac{1}{2}\sum_i \sum_{j\neq i} \mathbf{Z}_{nij}\eta \times (a_{nij}^a - a_{nij}^*)$$

We interpret the residual profit difference τ_n^a in Equation (7) as measurement errors following Houde et al. (2023), and "cancel" them out using a vector of non-negative instruments H that are correlated with changes in the observable profit components but uncorrelated with τ_n^a:

$$E[H \cdot \Delta\Pi_n(\mathbf{A}_n^a, \mathbf{A}_n^*; \eta)] + \underbrace{E[H \cdot \tau_n^a]}_{=0} \leq 0.$$

Then we follow Pakes et al. (2015) to construct sample moment inequalities and estimate η using the sample moment inequalities:

$$\frac{1}{N_{\text{alt}}} \sum_{n,i,j\neq i, a_{nij}=a} H_h \cdot [\Delta\Pi_n(\mathbf{A}_n^a, \mathbf{A}_n^*; \eta)] = \bar{m}_h(\eta) \leq 0. \qquad \forall h = 1, \ldots, |H|. \tag{9}$$

where $|H|$ is the number of moment inequalities and N_{alt} is the number of alternative network configurations.

Alternative Network Structures We consider alternative network structures that consist of "single market deviations". Specifically, if an airline is active in a market, we consider an alternative scenario where it exits that market. Conversely, for markets where an airline is not currently active, we examine a scenario where it enters. We use 2014 for this exercise, a period in which we have comprehensive gate operation data. There are a total of 6,934 airline-market observations where airlines entered a market; we use them to estimate the upper bound of entry costs. Similarly, there are 14,626 airline-market combinations where airlines did not serve these markets; we use them to estimate the lower bound of entry costs.[13] Specifically, for markets where airlines are currently operating, we have:

$$\frac{1}{6,934} \cdot \quad \frac{1}{2} \cdot \sum_{n,i,j \neq i, \text{ with } a^*_{nij}=1} H_h \cdot (\Delta \Pi^1_{nij} + \mathbf{Z}_{nij}\eta) \leq 0. \tag{10}$$

where $\Delta \Pi^1_{nij}$ is defined in Equation (8) and captures differences in variable profit and flight operating costs. These inequalities generate an upper bound of the entry cost parameters because $\mathbf{Z}$ is nonnegative. Similarly, for markets where airline n does not operate direct flights, its highest profit upon entry must be non-positive:

$$\frac{1}{14,626} \cdot \quad \frac{1}{2} \cdot \sum_{n,i,j \neq i, \text{ with } a^*_{nij}=0} H_h \cdot (\Delta \Pi^1_{nij} - \mathbf{Z}_{nij}\eta) \leq 0. \tag{11}$$

which generates a lower bound of the entry cost.

Inference Our inference procedure follows Cox and Shi (2023). Specifically, we construct a likelihood ratio test statistic:

$$T_n(\eta) = min_{\mu, \mu \leq 0} \quad n(\bar{m}_n(\eta) - \mu)' \hat{\Sigma}_n(\eta)^{-1} (\bar{m}_n(\eta) - \mu), \tag{12}$$

[13] An airline is considered a potential entrant in a market if it is active at both endpoints. This restriction is based on the observations that airlines rarely enter markets where they do not already serve both end cities (Berry, 1992). Consequently, the number of total observations used for the inequalities (6,934+14,626) is fewer than the product of the number of markets and the number of airlines.

where $\bar{m}_n(\eta)$ is an $|H| \times 1$ vector of moment inequalities defined in Equation (9), $\hat{\Sigma}_n(\eta)$ denotes an estimator of $Var(\sqrt{n}\bar{m}_n(\eta))$, the variance covariance matrix of the moments. The critical value of this test with a significance level of α is denoted by $\chi^2_{\hat{r},1-\alpha}$, the $100(1-\alpha)$ quantile of the Chi-squared distribution with $\hat{r}$ degree of freedom, $\chi^2_{\hat{r}}$. Note that $\hat{r}$ is the number of inequalities that are binding at a given value of η and needs to be evaluated at every η vector that is tested. The estimated confidence region is:

$$\phi_n^{CC}(\eta, \alpha) = \{\eta \in R^{|\eta|} : T_n(\eta) \leq \chi^2_{\hat{r},1-\alpha}\}. \tag{13}$$

where the superscript CC denotes "conditional chi-squared" following the notation in Cox and Shi (2023) and $|\eta|$ is the dimension of η. The confidence region is a collection of parameter values for which we fail to reject the test that $\bar{m}_n(\eta) \leq 0$. We standardize the explanatory variables $\mathbf{Z}_{nij}$ and carry out a grid search of the η vector when estimating the confidence region. Appendix B.2 provides more details.

5 Empirical Results

This section discusses the empirical results, in the order of: technologies that define the flight frequency for indirect services (Section 5.1), demand preferences (Section 5.2), marginal costs of serving passengers and operating flights (Section 5.3), and entry costs estimates (Section 5.4). Given our panel data structure, we reintroduce the period subscript t (quarter) when necessary.

5.1 Technological Relationship

Table 3 summarizes the estimates for the Cobb-Douglas technological relationship between the indirect flight frequency and direct flight frequency (Equation (4) in Section 4.1). A simple Cobb-Douglas function without fixed effects fits the data surprisingly well, with R^2 close to 0.8 for a sample of nearly 2.5 million observations. The estimates indicate that there exists an increasing return to scale: when direct flight frequency doubles in both markets ik and jk, indirect flight frequency in market ij increases by 149.2 percent.[14] Results are robust when controlling for airline and market fixed effects (Columns 2 and 3). Direct flights generate about 1.1 indirect flights per

[14] The coefficient in Column 1 of Table 3 suggests that when direct flight frequency doubles in either leg, indirect flight frequency increases by 74.6%. When direct flight frequency doubles in both legs, indirect flight frequency increases by $74.6\% \times 2 = 149.2\%$.

connecting city.[15]

5.2 Demand Estimates

Table 4 reports demand estimates.[16] The odd columns report OLS estimates, and the even columns report IV estimates, where IVs include the average flight frequency for both nonstop and one-stop services, hub indices for both the origin and destination cities, and nonstop product dummies of competing products. Columns (1) and (2) control for all regressors, Columns (3) and (4) exclude flight frequencies as explanatory variables, and Columns (5) and (6) exclude hub indices. While the parameter estimates are similar across columns, the Pseudo R-squared is noticeably lower in Columns (3)-(6), especially when flight frequencies are excluded. The F-values are 287.72 and 805.74 for the first-stage regressions of price and within-group shares on IVs, respectively, indicating strong BLP instruments (Appendix Table B.1). In addition, there is no evidence of serial correlations among the demand shocks. Column (2) is our preferred specification.

On average, consumers are willing to pay \$200 more for nonstop services compared to one-stop services ($\hat{\alpha}_1/\hat{\sigma}_1 = 1.972/0.985 * 100 = 200$), similar to the estimate in Aguirregabiria and Ho (2012), who found a willingness-to-pay of \$152 more for a non-stop flight over a connecting flight. The aggregate price elasticity, i.e., the percentage change in total demand when all product prices increase by one percent, is estimated at 1.64, corroborating the findings in Berry and Jia (2010), where aggregate price elasticity was found to be 1.55 in 1999 and 1.67 in 2006. Passengers are willing to pay extra for more frequent flight services, valuing one additional nonstop service at \$13.87 and one additional one-stop service at \$6.44.[17] A ten percent increase in hub size increase willingness-to-pay by \$3.9 to \$4.5.[18] Air travel demand increases with travel distance initially, though it dampens when distances become too far.

[15] Airlines operate on average 7.7 direct flights per market. According to column (1) of Table 3, the expected number of indirect flights per $ik - kj$ route is $exp(-2.949) \times 7.7^{0.746} * 7.7^{0.746} = 1.1$. The flight frequency of indirect services between two cities sums up the indirect flight frequencies across all routes/hub cities.

[16] Passengers' willingness to pay for product g, defined in Equation (5) in Section 4.2, is specified as:

$$b_{gt} = \underbrace{\alpha_1 \mathbf{1}[x = NS]}_{\text{Nonstop Dummy}} + \underbrace{\alpha_2 \ln(f_{gt}^{NS}) \times \mathbf{1}[x = NS] + \alpha_3 \ln(f_{gt}^{OS}) \times \mathbf{1}[x = OS]}_{\text{Flight Frequencies}}$$
$$+ \underbrace{\alpha_4 \ln(H_{nit}) + \alpha_5 \ln(H_{njt})}_{\text{Hub Indices}} + \underbrace{\alpha_6 d_{ij} + \alpha_7 d_{ij}^2}_{\text{Distance}} + \underbrace{\alpha_n}_{\text{Airline FE}} + \underbrace{\alpha_i + \alpha_j}_{\text{City FE}} + \underbrace{\alpha_t}_{\text{Quarter FE}} + \underbrace{\xi_g}_{\text{Shock}}.$$

[17] There are on average 7.7 daily direct flights and 10.5 daily one-stop flights. WTP is $\frac{\hat{\alpha}_2}{\text{Avg Direct Flights}} \times \frac{1}{\hat{\sigma}_1} \times 100 = \frac{1.052}{7.7} \times \frac{1}{0.985} \times 100 = 13.87$ for one additional direct flight and $\frac{\hat{\alpha}_3}{\text{Avg Indirect Flights}} \times \frac{1}{\hat{\sigma}_1} \times 100 = \frac{0.666}{10.5} \times \frac{1}{0.985} \times 100 = 6.44$ for one additional connecting flight.

[18] $\hat{\alpha}_4/\hat{\sigma}_1 \times 10\% \times 100 = 0.385/0.985 \times 10\% \times 100 = 3.9$ and $\hat{\alpha}_5/\hat{\sigma}_1 \times 10\% \times 100 = 0.441/0.985 \times 10\% \times 100 = 4.5$.

5.3 Variable Cost of Serving Passengers

Once we obtain demand estimates, we recover the marginal cost of serving passengers using $\hat{c}_{gt} = p_{gt} - \hat{\sigma}_1(1-\bar{s}_{gt})^{-1}$, where $\bar{s}_{gt} = (\sum_{g' \in \mathbf{G}_{nijt}} e_{g'})^{\frac{\sigma_2}{\sigma_1}}[1 + \sum_{n'=1}^{N}(\sum_{g' \in G_{n'ijt}} e_{g'})^{\frac{\sigma_2}{\sigma_1}}]^{-1}$, as explained in Section 4.2. The average marginal cost of serving an additional passenger is \$95. Given that the average travel distance is 1,187 miles, the per-mile cost is \$95/1187=\$0.08/mile, comparable to Berry and Jia (2010)'s estimate of \$0.06/ mile. We regress the marginal cost of serving passengers c_g on product, airline, and market attributes and estimate cost parameters δ's with OLS.[19]

Table 5 reports the estimated marginal cost of serving consumers. Column (1) includes all controls, Column (2) excludes flight frequencies, and Column (3) excludes city hub indices. Although travelers value nonstop services more than one-stop services, the marginal cost of serving a nonstop passenger is \$30.3 lower than serving a one-stop passenger. It is more costly to serve one-stop passengers because they occupy seats on two different flights and travel more miles by construction. As expected, marginal costs reduce with more frequent services: the marginal cost of serving nonstop passengers decreases by \$1.38 with one additional direct flight, and that of serving one-stop passengers drops by \$0.76 with one extra connecting flight.[20] The marginal cost is convex in distance; at the median travel distance, a 10% increase in miles traveled is associated with a \$3.7 increase in marginal cost. The economies of scale associated with hubs exist, though economically small, potentially due to the extensive set of airline $\times$ city fixed effects included in the estimation.

5.4 Flight Frequency Cost and Entry Cost

Figure 2 decomposes the marginal variable profit from an additional direct flight into four components according to Equation (6) and plots the box charts for each element. These components are 1) the marginal variable profit from this additional direct flight, 2) the cannibalization effect on existing one-stop services in the same market, 3) the marginal variable profit from one-stop services in markets where flight frequencies for indirect services increase as a result of this additional direct flight, and 4) the additional cannibalization effect from these indirect flights on existing nonstop

[19] The empirical specification for the marginal cost of serving passengers c_g is:

$$\begin{aligned} c_{gt} &= \underbrace{\delta_1 \mathbf{1}[x=NS]}_{\text{Nonstop Dummy}} + \underbrace{\delta_2 \ln\left(f_{gt}^{NS}\right) \times \mathbf{1}[x=NS] + \delta_3 \ln\left(f_{gt}^{OS}\right) \times \mathbf{1}[x=OS]}_{\text{Flight Frequency}} \\ &\quad \underbrace{+\delta_4 \ln\left(H_{nit}\right) + \delta_5 \ln\left(H_{njt}\right)}_{\text{Hub Indices}} + \underbrace{\delta_6 d_{ij} + \delta_7 d_{ij}^2}_{\text{Distance}} + \underbrace{\delta_{ni} + \delta_{nj}}_{\text{Airline} \times \text{City FE}} + \underbrace{\delta_t}_{\text{Quarter FE}} + \underbrace{\omega_g}_{\text{Shock}}. \end{aligned}$$

[20] $\frac{\hat{\delta}_2}{\text{Avg Direct Flights}} \times 100 = \frac{-0.106}{7.7} \times 100 = -1.38$, and $\frac{\hat{\delta}_3}{\text{Avg Indirect Flights}} \times 100 = \frac{-0.08}{10.5} \times 100 = -0.76$.

services in corresponding markets. The last column reports the total marginal variable profit, which is the sum of the first four columns.

On average, if an airline schedules an additional non-stop flight, it generates a daily marginal variable profit of $\$6,730$ from nonstop services (Table 6). This number is substantially higher than the median profit depicted in the first column of Figure 2 because the profit distribution is very right-skewed: a substantial fraction of airlines' (marginal) profits are accounted for by the top quartile of the most profitable markets. Similarly, an additional non-stop flight generates a daily marginal profit of $1,310 from one-stop services. Consequently, ignoring the airline network externality underestimates the costs and benefits of operating an additional flight by 13.2%. The cannibalization effects from the 2nd and 4th channels are much smaller relative to the profit gains and average at around $300.

The optimality condition for flight frequency implies that the marginal cost of operating an additional flight should equal the marginal variable profit, which is estimated at $\$7,753$ on average. Assuming the (log) marginal cost of operating flights depends linearly on flight and market characteristics, we estimate the operation cost parameters via OLS.[21]

Table 7 reports these cost estimates. The marginal cost of operating an additional flight increases with distance, as expected.[22] The marginal operating cost in a market is higher when an endpoint of this market has a higher hub index, likely due to congestion or capacity constraints.

5.5 Fixed Cost Estimation

We now direct our attention to the parameters governing the entry cost. Airline n's entry cost into market ij in quarter t (FC_{nijt}) depends upon several factors, such as the fraction of gates it operates at the two endpoint cities, market size, and whether it is a legacy carrier.[23]

[21]The empirical specification of the (log) marginal cost of operating flights is

$$\ln c^f_{nijt} = \underbrace{\gamma_1 \ln(H_{nit}) + \gamma_2 \ln(H_{njt})}_{\text{Hub Indices}} + \underbrace{\gamma_3 d_{ij} + \gamma_4 d^2_{ij}}_{\text{Distance}} + \underbrace{\gamma_{ni} + \gamma_{nj}}_{\text{Airline} \times \text{City FE}} + \underbrace{\gamma_t}_{\text{Quarter FE}} + \underbrace{\varepsilon_{nijt}}_{\text{Shock}}.$$

The log specification $\ln c^f_{nijt}$ fits data better than a level specification c^f_{nijt} due to the right-screwed empirical distribution of the operating costs.

[22]While the coefficient for the quadratic distance term is negative, only 5% of the sample with the longest travel distance has its marginal operating cost decreasing in distance.

[23]We specify FC_{nijt} as follows:

$$FC_{nijt} = \eta_1 + \eta_2 \times \underbrace{G_{nijt}}_{\text{Gate Share}} + \eta_3 \times \underbrace{ln(MS_{ij})}_{\text{Market Size}} + \eta_4 \times \underbrace{L_n}_{\text{Legacy Air Dummy}} + \underbrace{\kappa_{nijt}}_{\text{Entry Cost Shock}},$$

where $G_{nijt} = G_{nit} + G_{njt}$ is the sum of the share of gates leased to airline n in city i and j in quarter t. Ciliberto and Williams (2010) find that entry cost decreases with the share of gates an airline operates at an airport. $MS_{ij} =$

Table 8 presents the entry cost estimates, where we project the confidence region into confidence intervals for each parameter. The last row of Table 8 displays the number of inequalities used. Depending on the specifications, we select varying numbers of inequalities to achieve more accurate estimates. The first specification includes only the constant term. There are a total of 16 inequalities, with eight inequalities that are based on evenly-spaced population quantiles and eight inequalities based on evenly-spaced gate ownership quantiles.

Results from the first specification indicate that the confidence interval for the average daily entry cost is $[\$5,100, \$5,600]$, which equates to \$0.46 to \$0.51 million every quarter. The second specification adds the gate share and market size, while the third specification further controls for the legacy carrier fixed effect. The gate coefficient is negative, implying that airlines with more gates at the airport have lower entry costs. A one standard deviation increase in gate share (equivalent to a 21 percentage point increase) results in a quarterly entry cost savings of \$0.027 to \$0.060 million (i.e., 21% $\times$ 90 days $\times$ [\$-3,200, -\$1,400]). This significant difference demonstrates a substantial entry cost advantage for airlines that have a dominant position at an airport. Finally, legacy carriers face higher entry costs than other airlines.

6 Counterfactual Simulation

In 2016, the U.S. Department of Justice approved the merger between Alaska Air Group (AS) and Virgin America (VX).[24] It was not obvious ex-ante how the merger would affect the airline network structure. After estimating the demand and cost parameters, we now turn to the counterfactual analysis. We first investigate the importance of network externalities and then examine the implications of a hypothetical merger between Alaska and Virgin America in 2014. Section 6.1 describes the simulation strategy and introduces a weaker equilibrium concept than SPN. Section 6.2 discusses results. Appendix C provides more details.

6.1 A Weaker Equilibrium Concept

As elaborated in Section 4.4, solving an industry equilibrium for the entire network is computationally infeasible. We propose an equilibrium concept that is the same as the SPNE except that we restrict airlines' action space to $k-$market deviations, which we term as the level-k "restricted"

$Pop_i + Pop_j$ measures the size of market ij and L_n is the legacy carrier fixed effect.

[24] Alaska Air Group won a bidding war to acquire Virgin America in April 2016. We do not examine previous mergers, such as that between United Airlines and Continental Airlines, because they took place before 2014, and relevant cost-side data, particularly gate usage, is unavailable.

Subgame Perfect Nash Equilibrium ($k-$RNE). For example, a $1-$RNE is an equilibrium where no airline has any profitable one-market deviation. When $k = M$ (with M denoting the total number of markets), the $M-$RNE is the same as SPNE.

The "restricted" Subgame Perfect Nash Equilibrium (RNE) is a weaker equilibrium concept than SPNE. Indeed, there could be many local RNE's for any given SPNE. However, it offers two key advantages. First, it is computationally feasible because it drastically reduces airlines' action space. For each airline, the number of possible one-market deviations is M, and the number of feasible two-market deviations is $C_2^M = \frac{M(M-1)}{2}$. These figures are significantly smaller than the 2^M possible network configurations for a given airline in the standard SPNE, making the computation burden manageable. Second, there is a natural constructive approach to computing the $k-$RNE, where we begin with the observed network structure in the data and examine all possible $k-$market deviations for every airline. If there are no profitable deviations, then we have found a $k-$RNE. Otherwise, we modify the airline network to incorporate the profitable deviations and repeat this process until there are no profitable deviations. Despite these simplifications, solving the $k-$RNE's remains a difficult task. We focus on the $2-$RNE's in the counterfactual analyses discussed below. It takes five hours to compute the $2-$RNE once on a 200-core server. Appendix C explains computation tricks and additional assumptions that facilitate the solution of a $k-$RNE.

6.2 Counterfactual Results

We begin our analysis by establishing a pre-merger benchmark. In this step, we draw a set of model primitives and simulate the 2-RNE based on observed data. The resulting network configuration closely aligns with the actual network observed in the data, with minor differences due to model fit and random shocks. To investigate the role of network externalities, we shut down the network effects and re-solve the 2-RNE. Specifically, we exclude channels (c) and (d) in Equation (6) when deriving the marginal benefit of operating a non-stop flight in market ij and only keep the direct profit change from more non-stop services in that market.[25] In other words, airlines do not internalize the benefits generated by adding one non-stop service in market ij on other markets. By comparing outcomes with and without network externalities, we quantify the extent to which these externalities shape airline network structures. Next, we simulate the post-merger network, where Alaska Airlines (AS) and Virgin America (VX) operate as a single entity and optimize their

[25]The technological relationship between direct flight and indirect flight frequencies as described in Equation (4) does not change.

network jointly while other carriers adjust in response. Flight frequencies and prices of all products change accordingly. Comparing the market outcomes before and after the merger illustrates the effect of a merger. For each counterfactual scenario, we draw the entry cost parameters 100 times from their confidence region, solve the 2-RNE equilibrium for each parameter vector, and average the equilibrium outcomes across these simulations.

Network Externality Table 9 compares the airline network structure with (Column 1) and without network externality (Column 2). There are two countervailing forces when we shut down the network externality. On the one hand, airlines may schedule fewer direct flights as they no longer internalize additional revenues from one-stop services in other markets as a result of the network externality (channel (c) in Equation (6)). On the other hand, airlines may schedule more direct flights as they do not need to worry about the cannibalization effect in other markets (channel (d) in Equation (6)).

The simulation results indicate that, in the absence of network externalities, airlines would schedule 3.17% fewer direct flights. Hence, the first channel dominates. The reduction in nonstop frequency is more pronounced in the top 25% largest markets, which are typically connected to airline hubs. In addition, airlines would offer 21.53% fewer indirect flights. Note that the frequency of indirect flights diminishes more markedly than that of direct flights. Several factors explain these findings. First, without network externalities, airlines have less incentive to coordinate their direct flights and offer one-stop services. As a result, airlines reduce the number of direct flights linked to hub cities, particularly in large markets where profits from one-stop services are important under network externalities. Second, a change in direct flight frequency in one market can impact the frequency of indirect flights in up to $2 \times C - 3$ markets (Section 3.2), with a greater number of markets affected in denser hub-and-spoke networks. Consequently, the cumulative changes in flight frequency may be much more pronounced for one-stop services compared to nonstop services. As airlines reduce their flight frequency, their variable profit declines by 8.51% on average. Consumer surplus would decrease by 2.78%, with larger markets experiencing more significant reductions due to the greater drop in flight frequencies in these markets.

Post-merger Airline Network Table 10 presents a comparison of Alaska and Virgin America's network configurations before and after their merger. The first two columns summarize changes in nonstop services, while the last two columns focus on one-stop services. Prior to the merger, Alaska

Airlines and Virgin America jointly operated in 504 markets, with a total of 5,133 daily flights. After the merger, the newly formed airline introduced nonstop services in 52 markets (or 10% of the markets) where direct flights had not previously been offered by either carrier. Of these new markets, seven were previously served by indirect flights, and the remainder were markets that had been unprofitable for either airline. The merger enabled the new airline to capitalize on a more extensive network and operate 3.4% more daily flights, bringing the total daily flights to 5,311. The expanded network also facilitated more one-stop services, with the number of daily one-stop flights increasing from 456 pre-merger to 464 post-merger. The increase in connecting services was modest, partly because Vergin America did not have a hub, and Alaska only had one hub in Denver.

The post-merger airline also restructured its existing network across markets. We categorize all markets into four mutually exclusive and collectively exhaustive categories based on the pre-merger (benchmark) network: A) markets where both Alaska Airlines and Virgin America operated direct flights, B) markets where only one of the two operated direct flights, C) markets where they only provided indirect flights, and D) markets that were not served by either airline before the merger. Following the merger, the new airline maintained direct flights in all 87 markets in group A, though it significantly reduced the number of direct flights in these markets, from 1,482 pre-merger to 922 post-merger. In contrast, the airline expanded its operations in group B markets, where only one of the airlines had previously served nonstop flights, and especially groups C and D that lacked non-stop services before the merger. The total number of nonstop flights in groups C and D surged from 0 pre-merger to 487 post-merger.

At the industry level, the total number of markets served and total daily flights operated by all carriers increased slightly, indicating that the network expansion by Alaska and Virgin American was not offset by adjustments from other carriers.[26]

Table 11 analyzes the merger's impact on market structure and consumer surplus. The average effect on concentration is ambiguous ex-ante and varies by market. On the one hand, concentration may increase due to reduced competition following the merger. On the other hand, the newly merged airline might expand its operations to additional markets, potentially diminishing the market power of dominant airlines in those markets. Results from counterfactual simulations suggest that the average Herfindahl-Hirschman Index (HHI) across markets dropped slightly, from 3,063 before the merger to 3,051 after the merger. Meanwhile, the average price remains stable.

[26] We take the hub cities for each airline as given and do not endogenize them in this exercise. Endogenizing the hub status is a straightforward extension.

As a result, consumer surplus *increased* post-merger, driven by improved product quality resulting from new or more frequent services in a subset of markets and muted price responses overall. The combined profit of Alaska and Virgin America increased noticeably, while the industry's aggregate profit saw a modest rise of 0.5%. These results underscore the importance of accounting for network changes when analyzing airline mergers. Otherwise, one might erroneously conclude that the horizontal merger between Alaska and Virgin America would likely increase prices and harm consumers, while the merger between the two small carriers could actually help offset market power in some locally concentrated markets, thereby improving consumer welfare.

7 Conclusion

This paper introduces an empirical framework to analyze network competition among airlines, featuring the hub-and-spoke network structure. Ignoring network externality underestimates the benefits of operating an additional flight by 13.2%, and airlines would schedule 21.53% fewer one-stop flights had they made flight operation decisions independently for each market. To circumvent the curse of dimensionality of solving the industry equilibrium network configurations, we propose a novel equilibrium concept, the level $k-$ restricted SPNE, or $k-$RNE. A hypothetical horizon merger between Alaska and Virgin America in the first quarter of 2014 could benefit consumers, as the merged airline would expand non-stop services in more markets, thereby improving product quality and offsetting market power by dominant carriers in certain markets. These findings highlight the importance of endogenizing the network structure in merger analyses concerning the airline industry.

References

Aguirregabiria, Victor and Chun-Yu Ho, "A dynamic oligopoly game of the US airline industry: estimation and policy experiments," *Journal of Econometrics*, 2012, *168* (1), 156–173.

_ **, Robert Clark, and Hui Wang**, "Diversification of geographic risk in retail bank networks: evidence from bank expansion after the Riegle-Neal Act," *The RAND Journal of Economics*, 2016, *47* (3), 529–572.

Benkard, C Lanier, Aaron Bodoh-Creed, and John Lazarev, "Simulating the dynamic effects of horizontal mergers: US airlines," *Manuscript, Yale University*, 2010.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 1995, pp. 841–890.

_ **, Michael Carnall, and Pablo T Spiller**, "Airline hubs: costs, markups and the implications of customer heterogeneity," *In: Lee, D. (Ed.), Advances in Airline Economics. Vol. 1: Competition Policy and Antitrust. Elsevier Press, Amsterdam*, 2006, pp. 183–214.

Berry, Steven T, "Airport presence as product differentiation," *The American Economic Review*, 1990, pp. 394–399.

_ , "Estimation of a model of entry in the airline industry," *Econometrica: Journal of the Econometric Society*, 1992, pp. 889–917.

_ , "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

_ **and Panle Jia**, "Tracing the woes: an empirical analysis of the airline industry," *American economic journal. Microeconomics*, 2010, *2* (3), 1–43.

Blevins, Jason R, "Structural estimation of sequential games of complete information," *Economic Inquiry*, 2015, *53* (2), 791–811.

Bontemps, Christian, Cristina Gualdani, and Kevin Remmy, "Price competition and endogenous product choice in networks: Evidence from the us airline industry," Technical Report, Working Paper 2023.

Borenstein, Severin, "Hubs and high fares: dominance and market power in the US airline industry," *The RAND Journal of Economics*, 1989, pp. 344–365.

Brueckner, Jan K and Pablo T Spiller, "Economies of traffic density in the deregulated airline industry," *Journal of Law and Economics*, 1994, pp. 379–415.

Cardell, N Scott, "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory*, 1997, *13* (2), 185–213.

Ciliberto, Federico and Elie Tamer, "Market structure and multiple equilibria in airline markets," *Econometrica*, 2009, *77* (6), 1791–1828.

_ **and Jonathan W Williams**, "Limited access to airport facilities and market power in the airline industry," *Journal of Law and Economics*, 2010, *53* (3), 467–495.

_ **, Charles Murry, and Elie Tamer**, "Market structure and competition in airline markets," *Journal of Political Economy*, 2021, *129* (11), 000–000.

_ **, Emily E Cook, and Jonathan W Williams**, "Network Structure and Consolidation in the US Airline Industry, 1990–2015," *Review of Industrial Organization*, 2019, *54* (1), 3–36.

Cox, Gregory and Xiaoxia Shi, "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *The Review of Economic Studies*, 03 2023, *90* (1), 201–228.

Dou, Liyu, Jakub Kastl, and John Lazarev, "Quantifying Delay Propagation in Airline Networks," Technical Report, working paper 2020.

Eizenberg, Alon, "Upstream innovation and product variety in the US home pc market," *The Review of Economic Studies*, 2014, *81* (3), 1003–1045.

Ellickson, Paul B., Stephanie Houghton, and Christopher Timmins, "Estimating network economies in retail chains: a revealed preference approach," *The RAND Journal of Economics*, 2013, *44* (2), 169–193.

Fan, Ying, "Ownership consolidation and product characteristics: a study of the US daily newspaper market," *The American Economic Review*, 2013, *103* (5), 1598–1628.

Hendricks, Ken, Michele Piccione, and Guofu Tan, "Entry and exit in hub-spoke networks," *The RAND Journal of Economics*, 1997, pp. 291–303.

_ , _ , **and** _ , "Equilibria in networks," *Econometrica*, 1999, *67* (6), 1407–1434.

Ho, Kate and Ariel Pakes, "Hospital choices, hospital prices, and financial incentives to physicians," *American Economic Review*, 2014, *104* (12), 3841–84.

Houde, Jean-François, Peter Newberry, and Katja Seim, "Nexus Tax Laws and Economies of Density in E-Commerce: A Study of Amazon's Fulfillment Center Network," *Econometrica*, 2023, *91* (1), 147–190.

Jackson, Matthew O and Asher Wolinsky, "A strategic model of social and economic networks," *Journal of economic theory*, 1996, *71* (1), 44–74.

Jia, Panle, "What happens when wal-mart comes to town: an empirical analysis of the discount retailing industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Li, Sophia, Joe Mazur, Yongjoon Park, James Roberts, Andrew Sweeting, and Jun Zhang, "Repositioning and market power after airline mergers," *The RAND Journal of Economics*, 2022, *53* (1), 166–199.

Mele, Angelo, "A structural model of dense network formation," *Econometrica*, 2017, *85* (3), 825–850.

Molinari, Francesca, "Microeconometrics with Partial Identification," *Handbook of Econometrics*, 2020, *7*, 355–486.

Molnar, Alejandro, "Congesting the commons: A test for strategic congestion externalities in the airline industry," *Working Paper*, 2013.

Nevo, Aviv, "Mergers with differentiated products: The case of the ready-to-eat cereal industry," *The RAND Journal of Economics*, 2000, pp. 395–421.

Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii, "Moment inequalities and their application," *Econometrica*, 2015, *83* (1), 315–334.

Peters, C, "Evaluating the performance of merger simulation: Evidence from the US airline industry," *The Journal of law and economics*, 2006, *49(2)*, 627–649.

Richard, Oliver, "Flight frequency and mergers in airline markets," *International Journal of Industrial Organization*, 2003, *21* (6), 907–922.

Seim, Katja, "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 2006, *37* (3), 619–640.

Sweeting, Andrew, "Dynamic product positioning in differentiated product markets: the effect of fees for musical performance rights on the commercial radio industry," *Econometrica*, 2013, *81* (5), 1763–1803.
U.S. v. Northwest and Continential

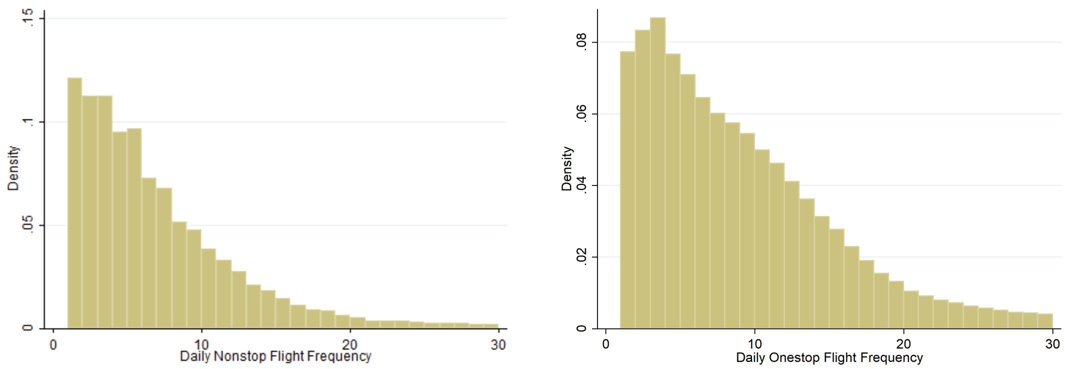
U.S. v. Northwest and Continential, *Department of Justice, 1998.*

Williams, Jonathan W, *"Capacity investments, exclusionary behavior, and welfare: a dynamic model of competition in the airline industry," Exclusionary Behavior, and Welfare: A Dynamic Model of Competition in the Airline Industry (December 1, 2008), 2008.*

Zhu, Ting and Vishal Singh, *"Spatial competition with endogenous location choices: an application to discount retailing," Quantitative Marketing and Economics, 2009, 7 (1), 1–35.*

Figure 1: Flight Frequency Distribution

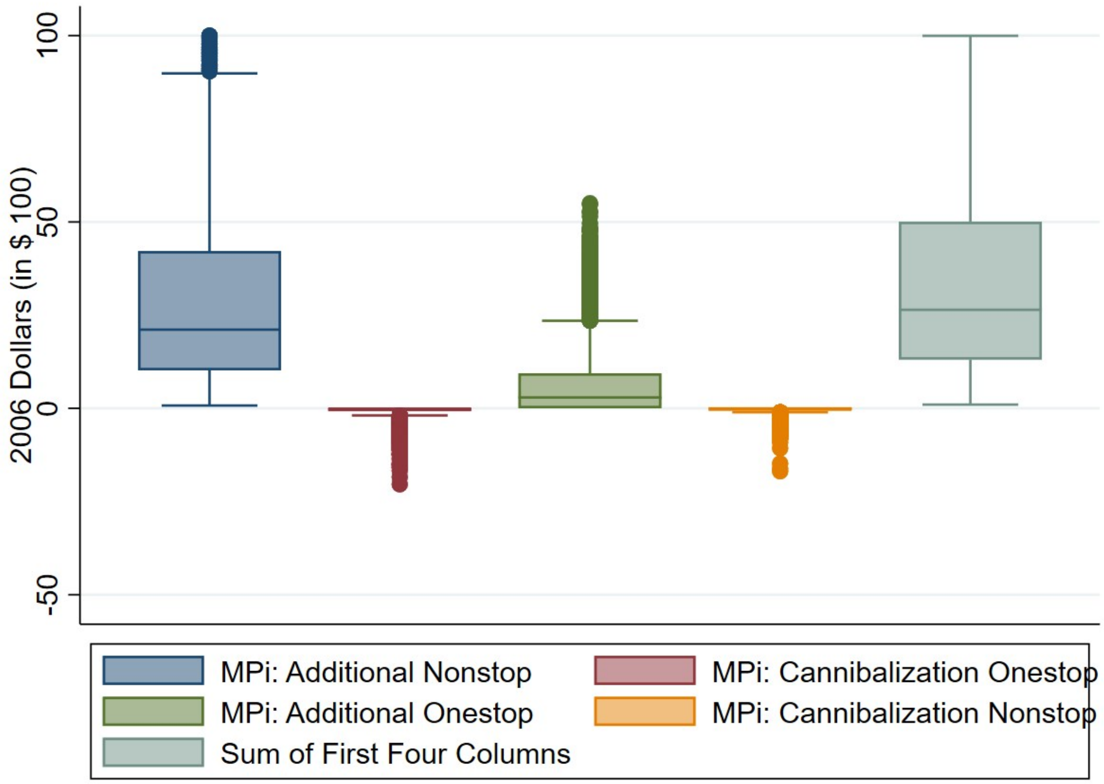
(a) Distribution of Direct Flight Frequency (b) Distribution of Indirect Flight Frequency



Note: This figure reports the distribution of flight frequencies. Panel (a) depicts the histograms of daily direct flight frequencies across markets. On average, airlines operate 7.7 daily direct flights per market, with a median of 5.7 flights. Panel (b) plots the distribution of indirect flight frequencies. The average indirect flight frequency is 10.5, with a median of eight.

(c) gains variable profit from one-stop services in other markets that now offer more indirect flights as a result of the increase in f_{nij}^{NS}

Figure 2: Decomposition of Marginal Variable Profit



Note: This figure decomposes the marginal variable profit from an additional non-stop flight into four components and summarizes their empirical distribution across airline-markets. 'MPi' is the abbreviation of marginal variable profit. The first column reports direct profit gains in nonstop services from one additional direct flight, with an average of $\$6,730$ and a median of \$3292 across airline-markets. The second column reports profit losses from the cannibalization effect on existing one-stop services in the same market, with an average of \$182 and a median of \$19. The third column reports profit gains from one-stop services in other markets that now offer more indirect flights upon the introduction of the additional non-stop flight. The average and median of such profit gains is $\$1,310$ and \$37, respectively. The fourth column reports profit losses from the cannibalization effect on nonstop services in corresponding markets, with an average of \$117 and a median of \$7. The last column reports the total marginal variable profit associated with an additional non-stop flight and is the sum of the first four columns, with an average of $\$7,753$ and a median of $\$3,938$.

Table 1: Summary Statistics of Nonstop and One-stop Services

	2007Q1					
	Nonstop Services			**One-stop Services**		
Airline Code (Name)	# Markets	% of Pass	% of Rev	# Markets	% of Pass	% of Rev
WN(Southwest Airlines)	359	91.0%	86.9%	1,115	9.0%	13.1%
DL(Delta Air Lines)	262	74.5%	75.3%	2,619	25.5%	24.7%
AA(American Airlines)	243	89.5%	87.8%	561	10.5%	12.2%
US(U.S. Airways)	240	84.4%	82.2%	1,410	15.6%	17.8%
UA(United Airlines)	158	84.0%	81.6%	1,214	16.0%	18.4%
NW(Northwest Airlines)	151	79.3%	79.8%	1,451	20.7%	20.2%
CO(Continental Airlines)	94	87.9%	86.7%	698	12.1%	13.3%
FL(AirTran Airways)	90	80.7%	78.9%	489	19.3%	21.1%
B6(JetBlue Airways)	67	98.6%	98.5%	78	1.4%	1.5%
AS(Alaska Airlines)	45	98.2%	97.7%	86	1.8%	2.3%
F9(Frontier Airlines)	45	82.4%	77.9%	460	17.6%	22.1%
NK(Spirit Airlines)	21	100.0%	100.0%	0	0.0%	0.0%
Total	1,775	86.8%	84.4%	10,181	13.2%	15.6%
	2014Q1					
	Nonstop Services			**One-stop Services**		
Airline Code (Name)	# Markets	% of Pass	% of Rev	# Markets	% of Pass	% of Rev
WN(Southwest Airlines)	488	86.5%	82.6%	1,420	13.5%	17.4%
DL(Delta Air Lines)	355	73.0%	73.3%	2,631	27.0%	26.7%
AA(American Airlines)	406	77.9%	77.0%	1,883	22.1%	23.0%
UA(United Airlines)	174	88.5%	88.5%	567	11.5%	11.5%
B6(JetBlue Airways)	85	99.4%	99.4%	75	0.6%	0.6%
AS(Alaska Airlines)	55	99.3%	98.9%	90	0.7%	1.1%
F9(Frontier Airlines)	30	78.8%	76.4%	125	21.2%	23.6%
NK(Spirit Airlines)	52	100.0%	100.0%	0	0.0%	0.0%
VX(Virgin America)	22	100.0%	100.0%	0	0.0%	0.0%
Total	1,667	83.6%	81.2%	6,791	16.4%	18.8%

Notes: This table summarizes the nonstop and one-stop services for all major airlines in Q1 2007 and Q1 2014. The airlines are ranked by the number of markets with direct services. 'Pass' stands for the number of passengers and 'Rev' denotes revenue. The table provides a detailed comparison of the number of nonstop and one-stop markets served by each airline, along with the corresponding market share and the percentage of revenue generated from these services in Q1 2007 and Q1 2014. While the majority of passengers travel on non-stop flights, one-stop services still contribute a nontrivial portion of airline revenue. Virgin America (VX) entered in 2008. Northwest merged with Delta, Continental with United, and U.S. Airways with American Airlines; all three mergers occurred between 2008 and 2013. AirTran exited in 2010. More details can be found in Appendix A.

Table 2: Summary Statistics of Hub Cities

	2007Q1			
Airline Code (Name)	Top Hub	Hub index	Second Hub	Hub index
WN (Southwest Airlines)	Las Vegas	47	Chicago	45
DL (Delta Air Lines)	Atlanta	77	Cincinnati	67
US (U.S. Airways)	Dallas	75	Chicago	71
AA (American Airlines)	Charlotte	59	Philadelphia	48
UA (United Airlines)	Chicago	54	Denver	44
NW (Northwest Airlines)	Detroit	55	Minneapolis	53
CO (Continental Airlines)	Houston	46	New York	37
FL (AirTran Airways)	Atlanta	37	Orlando	18
B6 (JetBlue Airways)	New York	36	Boston	19
F9 (Frontier Airlines)	Seattle	22	Portland	14
AS (Alaska Airlines)	Denver	43	San Francisco	3
NK (Spirit Airlines)	Miami	9	Detroit	8
	2014Q1			
Airline Code (Name)	Top Hub	Hub index	Second Hub	Hub index
WN (Southwest Airlines)	Chicago	58	Denver	49
DL (Delta Air Lines)	Atlanta	81	Detroit	65
AA (American Airlines)	Dallas	71	Charlotte	61
UA (United Airlines)	Chicago	39	Denver	34
B6 (JetBlue Airways)	Boston	29	New York	29
F9 (Frontier Airlines)	Seattle	30	Portland	18
AS (Alaska Airlines)	Denver	30	Kansas City	1
NK (Spirit Airlines)	Dallas	13	Chicago	12
VX (Virgin America)	San Francisco	13	Los Angeles	9

Note: this table summarizes the top two largest domestic hub cities for each major airline in the first quarter of 2007 and 2014, respectively. The airlines are listed in the same order as in Table 1. 'Hub index' measures the number of nonstop markets that an airline serves out of the hub city. Legacy carriers' hubs are connected to almost all major cities with a hub index of over 50. Southwest's largest hub also connects to over 40 airports.

Table 3: Estimation of Technological Relationship btw One-stop and Nonstop Flight Frequencies

ln(# of indirect flights)	(1)	(2)	(3)
	OLS	OLS	OLS
ln(# of flights in first leg)	.746***	.729***	.746***
+ ln(# of flights in second leg)	(.002)	(.002)	(.002)
Airline FE		Yes	Yes
City-pair FE			Yes
Constant	–2.949***	–2.827***	–2.813***
	(.021)	(.030)	(.027)
Pseudo. R^2	.764	.774	.796
Observations	2,487,262	2,487,262	2,487,262

Notes: This table reports the estimates that describe the technological relationship between one-stop and nonstop flight frequencies. An observation is an Origin-ConnectionCity-Destination combination in a quarter for a given airline. The dependent variable is the log of the number of indirect flights. The empirical specification is:

$$\underbrace{\ln f_{nij}^{OS(k)}}_{\text{Indirect Flight Frequency}} = h + \lambda \times \underbrace{\left(\ln f_{nik}^{NS} + \ln f_{nkj}^{NS}\right)}_{\text{Direct Flight Frequency}} + \epsilon_{nij}^{k}.$$

Standard errors are clustered at the airline-market level and displayed in parentheses. *p<0.10; **p<0.05; ***p<0.01.

Table 4: Demand Estimation

$ln(s_g) - ln(s_0)$	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Nonstop Dummy	2.281***	1.972***	2.116***	1.632***	2.776***	2.722***
	(.018)	(.043)	(.010)	(.050)	(.017)	(.018)
ln(Direct Flight Frequency)	1.116***	1.052***			1.293***	1.262***
	(.008)	(.015)			(.008)	(.009)
ln(Indirect Flight Frequency)	.713***	.666***			.848***	.888***
	(.005)	(.008)			(.005)	(.007)
ln(Origin Hub Index)	.255***	.385***	.447***	.648***		
	(.004)	(.019)	(.004)	(.019)		
ln(Destination Hub Index)	.317***	.441***	.533***	.718***		
	(.004)	(.018)	(.005)	(.018)		
Distance (in 1,000 miles)	2.558***	2.699***	1.828***	2.157***	2.686***	2.758***
	(.021)	(.027)	(.022)	(.031)	(.021)	(.028)
Distance-squared	–.566***	–.538***	–.471***	–.441***	–.575***	–.563***
	(.008)	(.008)	(.008)	(.009)	(.008)	(.008)
FARE in \$100 (Coefficient $-\frac{1}{\sigma_1}$)	–.382***	–.985***	–.399***	–1.404***	–.344***	–.696***
	(.005)	(.063)	(.005)	(.068)	(.005)	(.060)
$ln(s_g^*)$ (Coefficient $1 - \frac{\sigma_2}{\sigma_1}$)	.420***	.506***	.485***	.625***	.325***	.495***
	(.002)	(.019)	(.002)	(.020)	(.002)	(.022)
Pseudo. R^2	.632	.601	.590	.505	.621	.600
Observations	329,448	329,448	329,448	329,448	329,448	329,448
Test of residual serial correlation						
p-value	.979	.884	.766	.775	.885	.833

Notes: This table reports the demand estimation results. The dependent variable is the difference between a product's market share and the market share of the outside option. The empirical specification is $ln(s_g) - ln(s_0) = \mathbf{W}_g^d \frac{\alpha}{\sigma_1} - \frac{p_g}{\sigma_1} + (1 - \frac{\sigma_2}{\sigma_1}) ln(s_g^*) + \frac{\xi_g}{\sigma_1}$. 'Hub index' measures the number of nonstop markets that an airline serves out of a city. All specifications include airline fixed effects, market fixed effects, and quarter fixed effects. We conducted a test for residual serial correlation and found no evidence of serial correlation across different periods. Standard errors are clustered at the airline-market-quarter level and displayed in parentheses. ***1%, **5%, *10% significance level.

Table 5: Marginal Cost of Serving Passengers

Dept var: c_{gt}, marginal cost in \$100s	(1)	(2)	(3)
	OLS	OLS	OLS
Nonstop Dummy	–.303***	–.267***	–.304***
	(.008)	(.003)	(.008)
ln(Direct Flight Frequency)	–.106***		–.106***
	(.004)		(.004)
ln(Indirect Flight Frequency)	–.080***		–.080***
	(.003)		(.003)
ln(Origin Hub Index)	–.008*	–.007	
	(.005)	(.005)	
ln(Destination Hub Index)	–.002	–.002	
	(.004)	(.004)	
Distance (in 1,000 miles)	.225***	.291***	.226***
	(.008)	(.008)	(.008)
Distance-squared	.061***	.053***	.061***
	(.003)	(.003)	(.003)
Pseudo. R^2	.407	.405	.407
Observations	329,448	329,448	329,448

Notes: This table reports the estimation of the marginal cost of serving passengers. The empirical specification is:

$$\begin{aligned} c_{gt} \quad = \quad & \underbrace{\delta_1 \mathbf{1}[x = NS]}_{\text{Nonstop Dummy}} + \underbrace{\delta_2 \ln\left(f_{gt}^{NS}\right) \times \mathbf{1}[x = NS] + \delta_3 \ln\left(f_{gt}^{OS}\right) \times \mathbf{1}[x = OS]}_{\text{Flight Frequency}} \\ & + \underbrace{\delta_4 \ln\left(H_{nit}\right) + \delta_5 \ln\left(H_{njt}\right)}_{\text{Hub Indices}} + \underbrace{\delta_6 d_{ij} + \delta_7 d_{ij}^2}_{\text{Distance}} + \underbrace{\delta_{ni} + \delta_{nj}}_{\text{Airline} \times \text{City FE}} + \underbrace{\delta_t}_{\text{Quarter FE}} + \underbrace{\omega_g}_{\text{Shock}}. \end{aligned}$$

Hub index H_{nit} measures the number of nonstop markets that airline n serves out of city i. All specifications include airline $\times$ city fixed effects and quarter fixed effects. Standard errors are clustered at the airline-market-quarter level and displayed in parentheses. ***1%, **5%, and *10% significance level.

Table 6: Decomposition of Marginal Variable Profit from an Additional Direct Flight in \$100s

Channels	**Mean**	**Std. Dev.**
$\Delta\pi_{nij}$ from an additional Direct Flight	67.297	91.745
Cannibalization on Indirect Flight	-1.827	4.583
$\Delta\pi_{ni'j} + \Delta\pi_{nij'}$ from additional Indirect Flights in Other Markets	13.104	25.910
Cannibalization on Direct Flights in Other Markets	-1.176	3.444
Total $\Delta\pi_n$	77.532	103.365

Notes: This table decomposes the marginal variable profit from an additional direct flight into four channels for each of the 6,934 airline-market combinations. The first and second columns report the average and standard deviations of the profit gains, respectively. The first row shows the gains in variable profit from nonstop services in the same market, the second row reports the profit losses due to the cannibalization effect on existing one-stop services in the same market, the third row reports profit gains from one-stop services in markets where flight frequencies for indirect services increase as a result of this additional direct flight, and the fourth row reports the additional cannibalization effect from these indirect flights on existing nonstop services in the corresponding markets. The last row shows the total marginal variable profit associated with an additional direct flight, the sum of the first four rows.

Table 7: Marginal Cost of Operating Flights

	(1)
c^f_{nijt}	Coefficient / Standard Error
ln(Origin Hub Index)	.295***
	(.024)
ln(Destination Hub Index)	.223***
	(.018)
Distance (in 1,000 miles)	1.910***
	(.020)
Distance-squared	−.460***
	(.008)
Pseudo. R^2	.974
Observations	6,934

Notes: This table reports the estimated coefficients for the marginal cost of adding flight frequency (in $100). The dependent variable is the marginal cost of operating flights (c^f). The empirical specification is:

$$\ln c^f_{nijt} = \underbrace{\gamma_1 \ln(H_{nit}) + \gamma_2 \ln(H_{njt})}_{\text{Hub Indices}} + \underbrace{\gamma_3 d_{ij} + \gamma_4 d^2_{ij}}_{\text{Distance}} + \underbrace{\gamma_{ni} + \gamma_{nj}}_{\text{Airline} \times \text{City FE}} + \underbrace{\gamma_t}_{\text{Quarter FE}} + \underbrace{\varepsilon_{nijt}}_{\text{Shock}}.$$

Hub index H_{nit} measures the number of nonstop markets that airline n serves out of city i. All specifications include airline × city fixed effects and quarter fixed effects. Standard errors are clustered at the airline-market-quarter level and displayed in parentheses. ***1%, **5%, and *10% significance level.

Table 8: Entry Cost Estimates

	Specification 1		Specification 2		Specification 3	
	LB	UB	LB	UB	LB	UB
Constant	51	56	47	61	55	70
Gate Shares			-21	-4	-32	-14
log (Market Size)			-5	12	-11	0
Legacy Carrier Dummy					10	40
# Instruments	8		32		36	

Notes: This table reports the estimated confidence intervals of entry cost parameters (in $100 per day) via Cox and Shi (2023). The empirical specification for the entry cost is:

$$FC_{nijt} = \eta_1 + \eta_2 \times \underbrace{G_{nijt}}_{\text{Gate Share}} + \eta_3 \times \underbrace{ln(MS_{ij})}_{\text{Market Size}} + \eta_4 \times \underbrace{L_n}_{\text{Legacy Air Dummy}} + \underbrace{\kappa_{nijt}}_{\text{Entry Cost Shock}},$$

where $G_{nijt} = G_{nit} + G_{njt}$ is the sum of the share of gates leased to airline n in city i and j in quarter t, $MS_{ij} = Pop_i + Pop_j$ measures the size of market ij, and L_n is the legacy carrier dummy. The bounds of entry cost parameters are defined by Inequalities (10) and (11). The first specification only includes a constant, which is estimated based on four evenly-spaced cells of market size (the largest quartile of markets, the second largest quartile, etc.) and four evenly-spaced cells of gate shares. The second specification includes gate shares and market size and are estimated from 16 evenly-spaced cells of market size and gate shares each. The third specification also control for the legacy carrier dummy and is based on 36 cells. See Appendix B.3 for more details.

Table 9: Effect of Network Externality on Flight Frequency, Variable Profit, and CS

	(1) With Externality	(2) Without Externality	(3) % Change
Flight Frequency for Nonstop Services	138,192	133,807	-3.17%
In largest quartile of markets	41,137	38,123	-7.33%
In 2nd largest quartile of markets	31,694	30,500	-3.77%
In 3rd largest quartile of markets	34,143	33,987	-0.45%
In smallest quartile of markets	31,219	31,196	-0.07%
Flight Frequency for Onestop Services	72,793	57,119	-21.53%
In largest quartile of markets	38,076	30,337	-20.33%
In 2nd largest quartile of markets	18,825	14,404	-23.49%
In 3rd largest quartile of markets	10,532	8,213	-22.02%
In smallest quartile of markets	5,360	4,166	-22.29%
Variable Profit (in $mill per quarter)	56,283	51,491	-8.51%
In largest quartile of markets	32,317	29,335	-9.23%
In 2nd largest quartile of markets	13,528	12,395	-8.37%
In 3rd largest quartile of markets	8,081	7,550	-6.57%
In smallest quartile of markets	2,358	2,211	-6.24%
Customer Surplus (in $mill per quarter)	27,696	26,927	-2.78%
In largest quartile of markets	15,902	15,299	-3.79%
In 2nd largest quartile of markets	6,181	6,024	-2.54%
In 3rd largest quartile of markets	3,611	3,603	-0.23%
In smallest quartile of markets	2,002	2,001	-0.08%

Notes: This table compares the network structure, variable profit, and customer surplus with and without the network externality. Market quartiles are defined based on market size ($Pop_i + Pop_j$).

Table 10: Network Configurations Pre-/Post-Merger for AS and VX

	(1)	(2)	(3)	(4)
	Nonstop Services		Onestop Services	
	Pre-merger	Post-merger	Pre-merger	Post-merger
Total Number of Markets Served	504	555	102	104
Market groups defined based on the pre-merger network:				
A: Both AS and VX served non-stop	87	87	21	22
B: Either AS/VX (not both) served non-stop	417	416	61	63
C: AS/VX only served connecting services	0	7	19	19
D: Markets not served by AS/VX pre-merger	0	45	0	1
Total Daily Flight Frequency	5,133	5,311	456	465
Frequency in the following markets:				
A: Both AS and VX served non-stop	1,482	922	98	105
B: Either AS/VX (not both) served non-stop	3,651	3,903	265	267
C: AS/VX only served connecting services	0	69	92	91
D: Markets not served by AS/VX pre-merger	0	418	0	2

Notes: This table reports the number of markets served and daily flights operated by Alaska (AS) and Virgin America (VX) pre- and post-merger. The first two columns report statistics on nonstop services, and the last two columns report those for one-stop services. We categorize all markets into four mutually exclusive and collectively exhaustive groups, based on the pre-merger (benchmark) network: A) markets where both AS and VX operated direct flights, B) markets where either AS or VX (but not both) operated direct flights, C) markets where AS and VX only operated indirect flights (no direct flights), and D) markets that were not served by AS and VX before the merger.

Table 11: Consumer Surplus and Industry Profit Pre- and Post-Merger

	(1)	(2)	(3)
	Pre-merger	Post-merger	% Change
HHI	3,063	3,051	-0.40%
Average price of nonstop services	1.401	1.400	-0.04%
Average price of one-stop services	1.493	1.493	0.00%
Customer Surplus in mill $ per quarter	27,541	27,696	0.56%
Variable Profit (AS & VX) in mill $ per quarter	768	1,046	36.19%
Variable Profit in mill $ per quarter	56,003	56,283	0.50%

Notes: This table reports market concentration, prices, consumer surplus, and variable profit pre- and post-merger. HHI is the Herfindahl-Hirschman Index (HHI) measured by revenue shares. Prices are measured in $ 100. Customer surplus and variable profit are measured in million USD per quarter.

Online Appendix for "Network Competition in the Airline Industry: An Empirical Framework" – Not for Publication

A Data and sample selection

Our analysis focuses on airlines with at least 1 million passengers during the sample period. We focus on the major airlines for three reasons: First, small airlines tend to concentrate their services in small markets and have a negligible presence in the sample dataset. Second, most small airlines employ a point-to-point business model and usually carry an insignificant proportion of indirect passengers. Low-cost carriers often operate point-to-point business models, which rely more on direct flights and do not have major hubs. Third, eliminating these airlines can save substantial computational time, proportional to the number of airlines in the dataset.

Some airline bankruptcies and mergers take place during the sample period. We consider four major mergers: Delta and Northwest announced their merger on Apr. 14th, 2008, and completed the transaction on Dec. 31st, 2009; United Airlines and Continental merged on May. 3rd, 2010, with a closing day of Oct. 1st, 2010; Southwest controlled AirTran's assets after AirTran's bankruptcy on Sep. 27th, 2010. AMR Corporation, the former parent company of American Airlines completed its merger with U.S. Airways Group on Dec. 9th, 2013. For the analysis, two merging airlines are treated as separate airlines until the closing day and then as one (post-merger) airline. Given this, Northwest flights are considered Delta in our dataset in 2010 Q1; Continental flights are considered United Airlines flights after 2010 Q4; the few AirTran tickets in 2008 Q2 are counted as Southwest tickets; U.S. Airways tickets and operations become a part of American Airlines after 2014 Q1.

Table A.1 provides the summary statistics for the data used in our model estimation and is divided into three panels. The first panel shows statistics for nonstop service, containing 56,097 observations. The second panel presents data for one-stop service, with 273,351 observations. In both of these sections, information is provided on prices, flight frequencies, the hub index (representing the average number of routes connected to the endpoint city), population (the number of residents in the endpoint city), and distance (measured between the two endpoint cities). The final panel provides information on gate data and is sourced from a balanced sample of airline-market pairs from 2014, with 272,484 observations. The average fares for nonstop and one-stop services stand at \$ 176 and \$ 203 respectively. The average nonstop service caters to a larger population and covers a shorter distance in comparison to the one-stop service.

Table A.1: Descriptive Statistics

	mean	sd	min	median	max
	Panel II: Nonstop Service				
	mean	sd	min	median	max
Fare (in \$100)	1.766	0.648	0.234	1.655	9.592
Flight Frequency	7.651	7.316	1	5.666	108.285
Hub Index	25.368	10.595	2	25	76
Population (in 1000's)	9,101	5,937	490	7,344	33,080
dist (in 1000 miles)	.924	.593	.063	0.795	2.719
	Panel II: One-stop Service				
	mean	sd	min	median	max
Fare (in \$100)	2.038	0.698	0.225	1.935	16.624
Flight Frequency	10.485	9.751	1	8.047	183.428
Hub Index	8.602	9.151	0	5	76
Population (in 1000's)	5,658	4,656	129	4,182	33,080
dist (in 1000 miles)	1.240	.627	.093	1.116	2.723
	Panel III: Gate				
	mean	sd	min	median	max
Gate share	.074	.101	0	0.029	.723

Notes: This table provides the summary statistics for the data used in our model estimation and is divided into three panels. The first panel shows statistics for nonstop service, containing 56,097 observations. The second panel presents data for one-stop service, with 273,351 observations. In both of these sections, information is provided on prices, flight frequencies, the hub index (representing the average number of routes connected to the endpoint city), population (the number of residents in the endpoint city), and distance (measured between the two endpoint cities). The final panel provides information on gate data and is sourced from a balanced sample of airline-market pairs from 2014, with 272,484 observations.

B Estimation

B.1 Demand Estimation: First Stage

Flight frequency is exogenous in the demand estimation because demand shocks arrive after airlines schedule their flights. Prices and within-group market shares are endogenous. Table B.1 reports the first-stage estimation results. Column (1) and Column (2) reports the first stage estimation for prices and within-group market shares, respectively. Most of the main coefficients (nonstop fixed effect, direct flight frequencies) are significant in the main specification. F-values for the two endogenous variables are 287.72 and 805.74, respectively. Both values are greater than 10, indicating strong BLP instruments.

Table B.1: Demand Estimation First Stage

	(1) $ln(s_g^*)$ [Parameter $1-\frac{\sigma_2}{\sigma_1}$]	(2) FARE (in $100) [Parameter $-\frac{1}{\sigma_1}$]
Competitors:		
ln(Avg Direct Flight Frequency)	−.166***	.001
	(.004)	(.002)
ln(Avg Indirect Flight Frequency)	−.044***	.000
	(.004)	(.003)
ln(Avg Origin Hub Index)	.270***	.034***
	(.013)	(.009)
ln(Avg Destination Hub Index)	.236***	−.006
	(.009)	(.007)
Fraction of Nonstop Competitors	−1.038***	−.322***
	(.022)	(.010)
Dummy (Absence of Nonstop Competitors)	−.069***	.014***
	(.010)	(.004)
Dummy (Absence of One-stop Competitors)	.077***	.174***
	(.008)	(.007)
Focal Airline:		
Nonstop Dummy	1.045***	−.358***
	(.014)	(.007)
ln(Direct Flight Frequency)	.652***	−.023***
	(.006)	(.003)
ln(Indirect Flight Frequency)	.144***	−.055***
	(.005)	(.002)
ln(Origin Hub Index)	−.703***	.116***
	(.004)	(.002)
ln(Destination Hub Index)	−.647***	.114***
	(.004)	(.002)
Market:		
Distance	.143***	.258***
	(.016)	(.010)
Distance squared	.052***	.051***
	(.006)	(.003)
Pseudo. R^2	.410	.356
Observations	329,448	329,448

Notes: This table reports the demand estimation first stage results. We report the regression results for the two endogenous variables: $ln(s_g^*)$ (Column 1) and FARE (in $100) (Column 2). The first seven rows reports the instrument variables. there are seven instrument variables in total: the average flight frequency for both nonstop and one-stop services for the competitors, average hub indices for both the origin and destination cities for the competitors, and the fractions of competing products providing direct services and two dummy variables denoting the absence of nonstop and one-stop competitors. OLS estimation results. All specifications include airline fixed effects, market fixed effects and quarter fixed effects. Standard errors are clustered at the airline-market-quarter level and displayed in parentheses. ***1% significance level. **5% significance level. *10% significance level.

B.2 Estimation Details of the Moment Inequality Approach

Following Cox and Shi (2023)'s approach, for a given significance level $\alpha \in (0,1)$, we can construct a test $\phi_n(\eta;\alpha)$ for $H_0 : \eta = \eta_0$, where $\phi_n(\eta;\alpha) = 1$ indicates rejection and $\phi_n(\eta;\alpha) = 0$ indicates a failure of rejection. We can obtain a confidence set for η by calculating

$$CS_n(1-\alpha) = \{\eta \in \Gamma : \phi_n(\eta;\alpha) = 0\}.$$

Specifically, we construct a likelihood ratio statistic,

$$T_n(\eta) = min_{\mu,\mu\leq 0} \quad n(\bar{m}_n(\eta)-\mu)'\hat{\Sigma}_n(\eta)^{-1}(\bar{m}_n(\eta)-\mu), \tag{B.1}$$

where $\bar{m}_n(\eta)$ is an $|H|\times 1$ vector of moment inequalities defined in Equation (9), $\hat{\Sigma}_n(\eta)$ denotes an estimator of $Var(\sqrt{n}\bar{m}_n(\eta))$, the variance covariance matrix of the moments. The critical value of this test with a significance level of α is denoted by $\chi^2_{\hat{r},1-\alpha}$, the $100(1-\alpha)$ quantile of the Chi-squared distribution with $\hat{r}$ degree of freedom, $\chi^2_{\hat{r}}$. Note that $\hat{r}$ is the number of inequalities that are binding at a given value of η and needs to be evaluated at every η vector that is tested. The estimated confidence region is:

$$\phi_n^{CC}(\eta,\alpha) = 1[T_n(\eta) \leq \chi^2_{\hat{r},1-\alpha}]. \tag{B.2}$$

where the superscript CC denotes "conditional chi-squared" following the notation in Cox and Shi (2023) and $|\eta|$ is the dimension of η. The confidence region is a collection of parameter values for which we fail to reject the test that $\bar{m}_n(\eta) \leq 0$. We standardize the explanatory variables $\mathbf{Z}_{nij}$ and carry out a grid search of the η vector when estimating the confidence region.

Let us illustrate the estimation of the constant term (as shown in Column 1 in Table 8). For each value of η between -100 and 100 with a step of one (201 values), we compute the likelihood ratio statistics $T_n(\eta) =$, as described in Equation B.1. We then compare this statistic with a χ^2 distribution with a degree of freedom of 16, which is the number of inequalities used in the estimation. We obtain a set of parameters whose likelihood ratio statistic ($T_n(\eta)$) is smaller than the critical value $\chi^2_{\hat{r},1-\alpha}$. We then report the minimum and maximum of the set of estimated parameters.

Let's delve into the estimation of the constant term, which is represented in Column 1 of Table 8. We calculate the likelihood ratio statistics $T_n(\eta)$ for each value of η ranging from -100 to 100,

Table B.2: Entry Cost Estimates with Different Number of Instrument

	LB	UB	LB	UB	LB	UB
Constant	48	78	51	74	55	70
Gate Share	-55	-7	-38	-10	-32	-14
log Market Size	–20	12	-18	10	-11	0
Legacy Dummy	-10	60	0	50	10	40
# Instruments	56		64		72	

Note: In Specification 1, we have 56 instruments for both population and gate share. Specification 2 comprises 32 instruments each for population and gate share. Meanwhile, Specification 3 incorporates 36 instruments each for population and gate share.

with a step interval of one (yielding 201 values). This calculation is performed as per Equation B.1. Subsequently, we compare this statistic with a χ^2 distribution that has a degree of freedom of 16, which corresponds to the number of inequalities used in the estimation. This comparison yields a set of parameters with a likelihood ratio statistic ($T_n(\eta)$) that is less than the critical value $\chi^2_{\hat{r},1-\alpha}$. The minimum and maximum values from this set of estimated parameters are then reported.

B.3 Selection of Instruments

The estimation process involves varying the quantity of instruments utilized. Fewer instruments often lead to a broader estimate set, whereas increasing the number of instruments results in a more precise set. However, overloading with too many instruments can result in an empty estimate set. In our empirical study, we increment the number of instruments until an empty set is reached. We then report the most precise non-empty estimate.

Table B.2 shows the entry cost estimates as we alter the number of instruments. The first specification includes only the constant term. There are a total of 16 inequalities, with eight inequalities that are based on evenly-spaced population quantiles and eight inequalities based on gate ownership. Specifically, $H_h, h = 1, \cdot, 8$ in Equation (9) is a cell indicating whether market ij's population is in the hth bucket and $H_h, h = 9, \cdot, 16$ is a cell indicating whether airline n's airport presence is in the $(h-8)$th bucket. $\frac{1}{N_{\text{alt}}} \sum_{n,i,j \neq i, a_{nij}=a} H_h \cdot [\Delta\Pi_n(\mathbf{A}^a_n, \mathbf{A}^*_n; \eta)] = \bar{m}_h(\eta) \leq 0. \qquad \forall h = 1, \ldots, |H|$.

As we raise the quantity from 56 to 64 and ultimately to 72, the estimate set becomes progressively tighter. In fact, adding more instruments beyond this point (72 instruments) yields an empty estimate set.

C Counterfactual Experiment

C.1 Description of the Counterfactual Analysis

Every time we compute an equilibrium, we draw a set of model primitives and simulate two types of airline networks. The first is a pre-merger (benchmark) network, whereas the second is a post-merger (counterfactual) network. A comparison of these two market outcomes illustrates the effect of a merger.[27] We further compare these networks with simulated market outcomes without the network externality to explore the effect of the network effect. The following paragraphs discuss the setup of the simulations.

C.2 Equilibrium Concept

The counterfactual experiments suffer from similar computational difficulties as the estimation section. For instance, to validate whether $\mathbf{A}_n^*$ is airline n's optimal network decision, we have to compare the profit from $\mathbf{A}_n^*$ with profits from $2^M - 1$ other network structures. It is computationally infeasible to conduct this comparison.

Instead, we propose a "restricted" NE concept, where airlines can deviate in one market, two-market pairs and upto K-market set (denoted by a set $\Omega(.)$). In this "restricted" NE, an airline cannot profitably deviate from its present network structure ($\mathbf{A}_n^*$) to any other network structures in $\Omega(\mathbf{A}_n^*)$.

Empirically, we restrict airline choice sets to all possible one-market and selected two-market deviations. In a network with M markets, the set $\Omega(\mathbf{A}_n^*)$ contains at most $M + C_2^M - 1 = \frac{M(M+1)}{2} - 1$ different network structures, which is substantially lower than the number of possible networks ($2^M - 1$). This restriction reduces the computation burden of computing the equilibrium.

The "restricted" NE also has two stages. Although its second (pricing) stage is the same as the model's second stage in Section 3.3, the first stage is different:

(i) In the first (network formation) stage, we impose restrictions on the set of network structures an airline can choose. Given the optimal network structure of the competitors ($\mathbf{A}_{-n}^*$,$\mathbf{F}_{-n}^*$), airline's flight frequency cost shock (ε) and fixed cost shock (κ) and airlines' optimal pricing strategies in the second stage ($\mathbf{P}^*(.)$), airline n makes its entry ($\mathbf{A}_n$) and flight frequency decisions ($\mathbf{F}_n$) to

[27] We do not compare the observed network structure with a counterfactual network because the observed network may be very different from the simulated network structure. The counterfactual experiments are helpful when studying the change in the number of direct and indirect flights after the merger for the entire network, instead of isolated market-specific airline choices.

maximize its expected profit:

$$\begin{aligned}\{\mathbf{A}_n^*, \mathbf{F}_n^*\} = \underset{\mathbf{A}_n, \mathbf{F}_n}{argmax} \quad & \underbrace{E_{\xi,\omega}[\pi_n(\mathbf{A}_n, \mathbf{A}_{-n}^*, \mathbf{F}_n, \mathbf{F}_{-n}^*, \mathbf{P}_n^*(\xi,\omega), \mathbf{P}_{-n}^*(\xi,\omega); \alpha, \delta)]}_{\text{Total Variable Profit}} \\ & -\underbrace{\Gamma_n(\mathbf{A}_n, \mathbf{F}_n^{NS}, \varepsilon; \gamma)}_{\text{Flight Frequency Cost}} - \underbrace{FC_n(\mathbf{A}_n, \kappa; \eta)}_{\text{Total Fixed Cost}}. \\ & \text{subject to: } \mathbf{A}_n \in \underbrace{\Omega(\mathbf{A}_n^*)}_{\text{Choice Set}},\end{aligned}$$

where $\Omega(\mathbf{A}_n^*)$ is a set of network structures including $\mathbf{A}_n^*$ and one-market, two-market pairs and upto K-market deviations from $\mathbf{A}_n^*$.

The present equilibrium concept differs from the equilibrium concept in the theoretical model, where an airline's choice set includes all 2^M different network configurations. However, this equilibrium concept is consistent with our empirical strategy of estimating entry costs based on profit inequalities where no airline has profitable one-market deviations. This equilibrium concept is also similar to Pairwise Stability in the strategic network formation literature, seen in Jackson and Wolinsky (1996), where no players have profitable deviations by adding or removing a link.

C.3 Iterated Best Responses

This subsection proposes an iterated single-deviation responses algorithm as a method for selecting equilibrium. This convergence of optimal responses within the network forms what we term a "restricted" Nash Equilibrium. We first define a sequence of airline moves and airline best responses.[28] Airlines move sequentially according to this sequence and respond sequentially according to their best responses. Then, we use a two-step procedure to simulate an equilibrium. In Step One, we evaluate the airline's best responses, market by market, according to this sequence until convergence. Suppose Step One converges to a network where no airlines have the incentive to deviate in any single market. Step Two checks whether there are profitable two-market and upto K market deviations.[29] When this sequence of best responses converges, and there are no profitable deviations in the K-market set, the algorithm reaches a "restricted" NE. This sequence of moves is a method for selecting a specific "restricted" NE. Mele (2017) also considers a similar simulation algorithm in social networks.[30]

[28] Blevins (2015) studies the estimation of a sequential-move game with complete information.

[29] Empirically, we consider upto two-market deviation.

[30] In Mele's model, player payoffs depend on both direct links and link externalities, and players meet sequentially at random, myopically updating their links.

Sequence of Moves Airlines make decisions according to a sequence of airline-market pairs. Let airlines first move in larger markets and then move in smaller markets. Within each market, airlines move sequentially by profitability.[31] The largest market is New York - Los Angeles (NY-LA). The most profitable airline is JetBlue in the NY-LA market, and the second most profitable airline is American Airlines. The proposed sequence would assign JetBlue and American Airlines the first and second movers in the NY-LA market, followed by other airlines in descending order of market profitability. After all the airlines in the NY-LA market have moved, the sequence proceeds to airlines in the second-largest market, New York - Chicago. Again, airlines move sequentially by profitability. In this way, we obtain a sequence of all airline-market pairs.

Airline Best Response Airlines make decisions according to their best response rules in each airline-market pair. When an airline makes its optimal entry and flight frequency decisions in a market, it treats entry and direct flight frequencies in other markets as given. There is no closed-form solution to optimal entry and flight frequency decisions. We numerically compute airline profits for different flight frequencies and select airline decisions that maximize airline profits. We first compute airline counterfactual network structures for different possible flight frequencies in a market. Then, we compute the Bertrand-Nash equilibrium prices in all markets and calculate the total profit of the airline. An airline's optimal entry and flight frequency decisions maximize its total profit. To reduce the computation burden, we restrict airline direct flight choices to six frequency levels. An airline can stay out (zero daily flight) or enter with two, four, six, eight, or ten daily flights.

Step One: One-market Best Response Simulation We propose an iterated single-deviation responses method and compute airline best responses market by market until airlines have no profitable deviations in any market. Starting with the network structures observed in the data, we evaluate the airline's best response, airline-market pair by airline-market pair, according to the sequence defined above. Specifically, starting with the first airline-market pair in the sequence, we predict the optimal entry and flight frequency decisions of the airline in this market. Then, we proceed to the second airline-market pair in the sequence. Given the decisions in the first airline-market, we evaluate the best response of the second airline and update its network structure. We update airline network structures every time an airline enters, exits, or changes flight frequencies

[31] Berry (1992) estimates a model of sequential airline market entry. He orders airlines according to profitability and incumbency status.

in a market. After every airline-market pair has been visited, we re-visit the first airline-market pair and re-evaluate the best responses of the entire airline-market sequence.[32] If this sequence converges, we obtain a potential equilibrium in Step One, a network structure with no profitable one-market deviation.

Step Two: Two-market Deviation Refinement With the potential equilibrium in Step One, we check whether there are profitable K-market deviations. Take a two-market pair (AB, AC) for example. We evaluate whether any airline has positive deviations by changing its entry and flight frequency decisions in AB and AC markets simultaneously.[33] Specifically, in this two-market pair, an airline has 36 possible actions: six different choices in market AB (i.e., stay out of market AB or enter the market with two, four, six, eight, or ten daily flights) cross with six different choices in market AC (stay out of market AC or enter the market with two, four, six, eight, or ten daily flights). We compute airline profits for all 36 actions. When there are profitable deviations, the potential equilibrium in Step One fails to pass the refinement. It does not qualify as a "restricted" NE, and another iteration of Step One would be needed from the more profitable network configuration. When there are no profitable two-market deviations, the potential equilibrium in Step One passes the two-market deviation refinement and is a "restricted" NE. [34] The total number of two-market pairs is $C_2^M = \frac{M(M-1)}{2}$=6,995,670. It is computationally demanding to compute total profit with all possible two-market deviations. Empirically, we randomly draw $8,000$ two-market pairs and check whether airlines have profitable deviations in these two-market pairs.[35]

C.4 Model Settings

Pre-Merger Network To simulate pre-merger network structures, we assume the values of the model primitives are the same as their estimates.[36] Specifically, airlines maintain their gate allocations at each airport. The demand, variable cost of serving passengers, and variable cost of operating flights are the same as the estimated value.[37] For the entry cost, all parameter

[32]In the robustness check, we consider other sequences of airline-market where the order of all airline-market pairs is randomized.

[33]Given that this network structure is obtained from the one-market best response, there are no profitable one-market deviations.

[34]Once the one-market equilibria do not pass two-market deviation refinement, we stop and re-start with a new trial when an algorithm does not converge in ten iterations.

[35]There were nine airlines in 2014. Each airline chooses from 36 different actions in a two-market pair. Therefore, we compute for $8,000$ two-market pairs $\times$ nine airlines $\times$ 36 actions = 2.6 million possible profit values.

[36]The primitives of the model include consumer utility functions, the variable cost of serving passengers and of operating flights, and entry cost.

[37]We set the value of all structural errors to zero.

estimates are intervals. In any simulation trial, we separately draw each parameters from a uniform distribution where the upper and lower bounds are the two endpoints of the estimated intervals. Then, we compute airline entry cost with these parameter draws from equation (14).

Post-Merger Network To simulate the post-merger network structures, we assume that non-merging airlines maintain their gate allocations, unobserved product quality in consumer demand, cost of serving passengers, and operating flights. The two merging airlines (Alaska Airlines and Virgin America) were eliminated and replaced with a new airline. The new airline owns all gates from the two merging airlines.[38] Post-merger unobserved product quality in demand, variable cost, and fixed cost are calculated as the weighted average of the two airlines. For markets in which Alaska Airlines and Virgin America were both active pre-merger, weights are based on the two airline's relative flight frequencies. For markets in which neither Alaska Airlines nor Virgin America was active pre-merger, weights are based on their gate shares at the two endpoints.[39]

The Network Externality To study how the network externality shapes network structures, we consider two scenarios in the simulation: a model with the network externality (Scenario I) and a model without the network externality (Scenario II). In Scenario I, an airline considers profit externality in other markets when deciding its entry and flight frequency decisions in a market. In this scenario, airlines make entry and flight frequency decisions to maximize their overall network profits. An airline may enter a market where operating direct flights is unprofitable. Still, the increase in its one-stop variable profit in other markets compensates for its loss in this market. In Scenario II, an airline makes entry/exit decisions in each city-pair independently without considering the externalities to other markets. In this scenario, airlines make entry and flight frequency decisions to maximize their local market profits. A comparison of the network structures in Scenarios I and II illustrates how the network externality shapes airline network structures.[40]

[38] In reality, there may be some gate or slot re-allocation post-merger. We assume there is no change in gate allocations for simplicity. This model can be extended to study further how gate re-allocation affects airline network structure.

[39] This merger setting is consistent with the Average Case Scenario in Ciliberto et al. (2021).

[40] Dou et al. (2020) develop a framework for quantifying delay propagation in airline networks. The current paper focuses on entry and flight frequency changes.

C.5 Simulation Analysis

Consumer surplus is computed according to the following formula:

$$W = \sigma_1 ln[1 + \sum_G (\sum_g exp(\frac{V_{g,G}}{1 - \frac{\sigma_2}{\sigma_1}}))^{(1-\frac{\sigma_2}{\sigma_1})}], \tag{C.1}$$

where $V_{(.)}$ is the deterministic component of the indirect utility function.