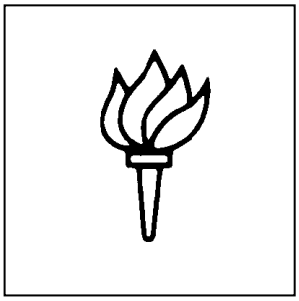
NEW YORK UNIVERSITY SCHOOL OF LAW

NYU Center for Law, Economics and Organization



Regulation and the Nature of the Firm: The Case of U.S. Regional Airlines

Michael E. Levine

July 2011

LAW & ECONOMICS RESEARCH PAPER SERIES
WORKING PAPER NO. 11-24

Regulation and the Nature of the Firm: The Case of U.S. Regional Airlines

Michael E. Levine*

ABSTRACT

In his pathbreaking article, Ronald H. Coase postulated that the structure of production was determined by the comparative advantage between contracting and hierarchy in securing and coordinating complementary resources (human or material) when the efforts of more than one person were required for production. While Coase considered the possibility that the regulatory context in which activity takes place might also influence the way it is structured, he considered the problem only very generally. This paper suggests that airline deregulation has profoundly affected the structure of firms that operate airline networks by affecting the contractual conditions under which airlines purchase labor inputs, by removing constraints on the extent and nature of the firm's route network and by changing the competitive environment in which airline firms operate. Where a stable uniform firm structure existed under regulation, airline networks now are organized with a variety of firm structures and individual networks have changed as particular conditions have changed over time. No single structure dominates, although some are more common than others.

***

The organization of airline networks, and particularly of the interaction between the less dense parts of an airline network with the more dense parts, is a particularly good example of the operation of two of Coase's main points in "The Nature of the Firm"[1] and subsequent articles: first, that the choice of institutions chosen to organize production is a function of economic circumstances, including regulation[2] and technology[3] and second, that there is

---

* Distinguished Research Scholar and Senior Lecturer, New York University School of Law. This paper was prepared for a conference organized by the Journal of Law and Economics to celebrate the 100th birthday of Professor Ronald H. Coase. The author studied with Professor Coase as Law and Economics Fellow at the University of Chicago Law School in 1967-68, was greatly influenced in his subsequent thinking by Professor Coase and wishes to express his profound gratitutde for the opportunity that was provided to him. The author wishes to thank the Journal for its support, an anonymous referee for his or her suggestions and L. Rush Atkinson V for valuable research assistance.

[1] R. H. Coase, "The Nature of the Firm, *Economica, New Series*, Vol. 4, No. 16 (November 1937), pp. 386-405

[2] *The American Economic Review*, Vol. 54, No. 3, Papers and Proceedings of the Seventy-sixth Annual Meeting of the American Economic Association (May, 1964), pp. 192-197.

[3] *Supra n. 1* at p.397

no general outcome that economic theory predicts, but rather that the result always depends on the particular circumstances and choices available and that it will change as circumstances change. [4]

In his pathbreaking article, Ronald H. Coase postulated that the structure of production was determined by the comparative advantage between contracting and hierarchy in securing and coordinating complementary resources (human or material) when the efforts of more than one person were required for production. He noted that cooperation could be achieved by market purchases of materials or labor required for production[5], or that production could be organized according to directives issued within hierarchical firms. Coase postulated that firms exist when the costs of contracting, including information costs, bargaining costs, the costs of specification about behavior under uncertainty and enforcement costs, exceed the "administrative" costs of hierarchy, including hiring, employment contracting, observation of work and outputs and enforcement of work rules. Coase did not explicitly mention the presence of unions and collective bargaining, but to the extent that they raise or lower contracting or administrative costs in organizing production labor, his theory easily accommodates the notion that they would affect the choice of organization.

Williamson elaborated Coase's deep insight, specifying in much more detail the conditions under which transaction costs can arise. In the process of doing so, he focused on incentive problems that inhibited cooperation, distorted contractual relationships and promoted structural invention. Williamson focused with some particularity on the various institutions that parties create to minimize transaction costs and harmonize incentives.[6]

While Coase considered the possibility that the regulatory context in which activity takes place might also influence the way it is structured, he considered the problem only very generally, noting taxation, quotas and "methods of price control which imply that there is rationing, and which do not apply to firms

---

4 R. H. Coase, "The Nature of the Firm: Meaning", *Journal of Law, Economics, & Organization*, Vol. 4, No. 1 (Spring, 1988), pp. 19-32 , 27

5 Presumably including subunits of production, or processes, not specifically mentioned.

6 Williamson, The Economic Institutions of Capitalism, 1985. For a focused discussion of his views on the relationship between contracting and organizations, along with a literature review, *see* Oliver E. Williamson, "The Theory of the Firm as Governance Structure: From Choice to Contract". 16 *Journal of Economic Perspectives* (3): 171–195(2002).

producing such products for themselves…".[7] He went on to observe that "…it is difficult to believe that it is measures such as have been mentioned in this paragraph that have brought firms into existence."[8] Later assertions that can be used to expand that insight noted that in determining the efficacy of regulation, one has to compare the real-world operation of markets with the real-world operation of regulation[9] and that to effectively theorize about economics, one had to look at real-world institutions.[10] Similarly, while Williamson has considered economic regulation for the limited purpose of assessing its implications for structuring and monitoring monopoly franchises, particularly in cable networks[11], he did not focus more broadly on the direct impact of regulation on the organization of firms in otherwise competitive markets.

This Article examines in more detail the impact of regulation, the contracts that govern labor inputs (particularly pilot labor) and coordination technology (including contracts between firms) on the organization of a particular activity, the coordination of U.S. domestic airline service in lower-density markets with the rest of the airline network. It pursues further both Coase's suggestion that regulation may have an impact on the coordination of production and his injunction to examine real-world institutions in assessing and theorizing both regulation and the operation of markets. It concludes that, at least in this case, the impact of regulation was more profound and pervasive than the language of Coase's seminal paper would suggest and that the tension between labor contracts and coordination technology governed the organizational choices ultimately made. This is by no means to suggest that Coase's or analytical framework is incorrect, but rather to expand on it in combining real-world facts with economic theory to predict and recommend outcomes, an approach that he has followed himself in several important papers[12]. In doing so, I will add Williamson's analysis of opportunism, holdups and asset specificity to Coase's insight to develop my argument.

The basic unit of airline service from the consumer standpoint (demand) is a trip from an origin to a destination, involving a flight from one airport to another plus coordinated ground transportation. Passengers value both time

---

[7] P 393
[8] *ibid.*
[9] R.H. Coase The Regulated Industries, Discussion, 54 *American Economic Review,* Papers and Proceedings, May 1964
[10] R. H. Coase "The Institutional Structure of Production", 82 *The American Economic Review*, 713-719 (1992)
[11] O. Williamson, *The Economic Institutions of Capitalism* (1985), Ch. 13
[12] R.H. Coase, "Payola in Radio and Television Broadcasting", 22 Journal of Law and Economics 269-328 (1979); "The Federal Communications Commission", 2 *Journal of Law and Economics* 1-40 (1959)

and convenience, preferring nonstop service to and from a nearby airport at a departure time convenient to them to one-stop or connecting service or service to less-convenient airports. They make substitutions between convenience attributes, such as infrequent nonstop service or service from distant airport *vs.* more convenient connecting service, and they trade convenience for cost at substitution rates that vary with the purpose of the trip and the value placed on time by the entity paying for it.

The economics of production of air transportation favors aggregating individual passengers for joint production of trips in aircraft whose unit costs decline with size over much of the relevant production range. Relatively few nonstop airport-to-airport markets (aggregations of passenger demand) allow the use of aircraft of optimally efficient size on journeys arranged to maximize convenience. The production technology used to reconcile demand and supply is to trade aircraft size for schedule convenience and to combine trips by either collecting passengers over time for a nonstop trip that is optimally timed for relatively few or none of them or to assemble them at a more convenient time into a group with multiple destinations and fly them to an intermediate airport (a "hub") where they can be recombined into new groups of passengers with different origins all destined for the same airport. These connecting trips are also flown in aircraft whose size often trades cost for comfort and convenience.

Since trip costs per passenger rise fairly sharply when using aircraft smaller than about 100 seats, the larger a network of connections is, the smaller the airport - pair market it can serve at acceptable cost by using connecting passengers to fill flights. This produces an economy of network size and scope in the presence of an indivisibility constraint. But network size and scope come with a price: the problem of coordinating flights becomes exponentially more difficult as the hub becomes larger and the spokes (nonstop destinations from the hub) more varied, especially since flight times are affected by stochastic events like congestion and weather. In addition, capital and gate staffing cost per unit of output goes up as the hub gets larger, because more gates need to be active simultaneously and then inactive waiting for the next hub "bank" of flights. And the percentage of aircraft time available for actually transporting passengers goes down as aircraft wait on the ground for other aircraft to arrive carrying connecting passengers and spend time queuing to depart.

Indivisibilities create a second large problem: A typical airport-to-airport journey requires combining many inputs with common costs such as flight costs (one cannot practically transport only a few passengers at a time), fleet

maintenance (involving expenditures and investments that support many flights for many aircraft), marketing costs (usually efforts involving substantial economies of scale or public goods) and coordination technology and labor costs to produce individual passenger journeys. These costs must somehow recovered by allocating them to each customer and each producer. This produces numerous complexities of cost allocation both within each firm flying each flight and between firms flying connecting flights combined to produce one journey.

In addition, there are numerous possibilities for revenue inefficiency within firms: for example, displacing a passenger flying from Dothan, Alabama to Tokyo via Atlanta with one flying only from Dothan to Atlanta presents a conflict because the local fare that can be charged for a business passenger between Dothan and Atlanta is disproportionately large on a unit revenue basis when compared to the leisure fare that can be charged between Dothan and Tokyo via Atlanta but would result in lower total combined revenue for the two segments than the leisure fare if the Atlanta-Tokyo seat would otherwise remain vacant. A single firm considers its marginal total revenue and compares it with its total joint cost. If more than one firm is involved, possibilities for externality and opportunities for holdups become significant. The Dothan-Atlanta carrier will only take into account its own revenue and costs and may either displace a Tokyo passenger with a local one or threaten to sell a seat to her only if it is allowed to extract compensation (including rents where possible) from the Atlanta-Tokyo carrier.

This is quintessentially the sort of problem that Coase and Williamson address: how should one organize production and marketing of aircraft flights that combine multiple passenger journeys? One could imagine internalizing all the inefficiency and holdup problems by organizing all the activity within one firm. Alternatively, one can imagine a contract between two or more firms designed to produce jointly the aircraft flight pattern (the "schedule") that accommodates the individual passenger trips. In the same way, choices need to be made about how the trips shall be sold and whether they will be sold by one firm or many. If the trip is produced or the tickets are sold by more than one firm, this presents the question of how the revenue collected for the trip should be divided. And as we have seen, the fact that many of the inputs for coordination themselves exhibit economies of scope and scale complicates the problem.

I will argue that regulation influenced greatly the extent and shape of airline networks, and affected in particular the "make or buy" decisions of large

airlines with respect to serving passengers flying in markets requiring the use of relatively small aircraft, either because of the size of the cities in which passengers originated or for which they were destined or because competitive considerations dictated frequent service without providing many passengers per flight. A major factor was the impact of regulation on the provision of labor. In addition to specifying the ultimate extent of each airline's network, and hence which labor force could provide inputs and ultimately on what competitive terms, safety regulation indirectly influenced these decisions by influencing the specificity of labor assets, thus limiting the ability of airlines to substitute one pilot for another. These factors in turn, ultimately influenced the labor organizations with which airlines contracted for pilot services and the internal politics of pilot unions influenced the terms of the contracts that they could ratify. Ultimately, these contracts in turn influenced service on lower-density routes by affecting, in many cases specifying, what kinds of aircraft would be kept inside which firm and how they would be deployed.

Regulation also mandated the contract terms on which connecting passengers could be exchanged between flights offered by different airlines. It limited the number of alternative providers of connecting service. It also specified or greatly influenced the degree of vertical integration in marketing the services produced and the terms on which ticket were marketed and sold.

A final and critically important influence on the structure of airline networks and the nature of the firms that operate them is and has been the interdependence and asset-specificity of all the elements -- aircraft, ground equipment, maintenance programs, maintenance equipment, information technology, passenger-handling procedures, and many more -- that make up the network and the complexity of the coordination task given the many elements, some stochastic, that can influence its operation. This makes any contracts between firms necessarily "relational" in Williamsonian terms and creates serious potential for opportunism and holdups, as well as allowing the reliability and standards of performance of one part of the network to have a very significant operational and financial impact on the rest.

Implicit in Coase's analysis is the notion that when the conditions under which contracting or administration occurs change, the optimum organization of production can change. If one accepts that the regulatory context in which production takes place can be important, then it also follows that a change in regulatory regime may well produce a change in organization. Service by U.S. airlines in domestic markets was regulated in detail between 1938 and 1978 and has been mostly deregulated since. Accordingly, changes in this service

following changes in regulation can serve as a "natural experiment" allowing us to observe the operation of regulation on the structure of the firm in the context of Coase's pathbreaking theory.

As noted, airline networks were greatly influenced by regulation and since deregulation have been greatly influenced by the tension between the incentives created by pilot labor contracts on the one hand and the need to manage relational impacts on the other. In addition, deregulation and its consequent changes also influenced capital structure and this in turn has also influenced firm structure. No single form has emerged, because labor contracts are different, the coordination technology changes, and different contractual structures to manage relationships have different strengths and weaknesses, meaning that whether one is superior to the other depends on specific technology, firm cultures[13], characteristics and relationships. As Coase predicted, the "optimal" form depends on specific conditions affecting transactions and administration.

Although airlines had some flexibility to shape their route networks before deregulation, they were severely limited in doing so by law and regulatory policy. All decisions to add or drop a city from an airline's network and many of the most important decisions to add or abandon service between cities already on the network required the permission of the Civil Aeronautics Board. All agreements between airlines for complementary service and the interchange of passengers between networks were either prescribed by regulation or subject to regulatory approval. A firm's agreements with its labor force were largely but not entirely subject to pattern bargaining by unions across airlines reinforced by the protection from competition provided by entry regulation, by the firm- and aircraft-type specificity that safety regulation imposed on pilot service and by aircraft-related structural elements of compensation that had their origins in a government labor arbitration award and were completely in place when the regulated industry began to expand after World War II[14]. Finally, the regulatory regime contained incentives and constraints such that virtually all airlines of any

[13] For these purposes, firm "culture" can be viewed as the shared expectations within the firm as to how individuals and the firm should react to expenditure choices, to stochastic variation in operating conditions and across passenger service dimensions too complex to allow network contracts to be complete.

[14] *See* I. Cohen, "David L. Behncke, The Airline Pilots, and the New Deal: The Struggle for Federal Labor Legislation", 41*Labor History* 47*,* (2000). The arbitration award should be regarded as a second form of regulation, one that occurred before the Civil Aviation Act of 1938, but in a period in which the industry was only partially regulated but almost entirely dependent on airmail payments from the Post Office Department. Postal payments and this Federal imposition shaped the structure of the industry, and which was carried over into the regulatory environment after 1938. (*See* M. E. Levine, "Is Regulation Necessary? California Air Transportation and National Regulatory Policy, 74 *Yale Law Journal* 1416 (1965)

size elected to do most of their marketing distribution through a prescribed contractual relationship with travel agents.

Firms effectively were required to charge prices uniform with those of their competitors and to split those prices with other airlines sharing production of a passenger's itinerary (through mandated interline connections) using a formula prescribed by regulation. For much of the regulatory history, fares in each market were uniform but there was some variation in unit pricing among markets based on market history and density. This ultimately disappeared, and in the last years of regulation, the whole industry was governed by a simple, prescribed distance-based fare structure[15]. Since new entry was very difficult and price competition was effectively impossible, disequilibrium created by the emergence of new firms with different costs, different prices and different labor arrangements was nearly nonexistent.

Firms varied in size, but not much in their organization. Airlines were vertically integrated, mostly unionized[16], and responded nearly uniformly to uniform incentives to serve or shed less dense markets. "Trunk" carriers served dense markets and long-haul markets. Separately-certificated "local service" carriers (later renamed "regional carriers") served markets involving smaller cities and relatively short differences. Attempts to cross-subsidize service by regional carriers in less-dense markets by giving them access to service in "richer" markets created exceptions to this pattern over time, but these exceptions were themselves coordinated and strictly limited by regulation.

After deregulation, firms unilaterally determined their own U.S. domestic route structures and prices and were free to coordinate their route structures with other companies to create networks that could be marketed and priced to degrees that varied from separately-identified service with various levels of coordination to creation of one system with one brand. They took extensive advantage of the opportunity to do so[17]. Although labor agreements held over

---

[15] U.S. Civil Aeronautics Board. *Domestic Passenger Fare Investigation, January 1970 to December 1974*. U.S. Government Printing Office. Washington, D.C. 1975.

[16] Delta's pilots were unionized and subject to a contract very similar to those of other airlines but, unusually, the rest of its employees were not. Delta's wages, however, were largely determined by the wages paid by unionized airlines, although its work rules were more flexible.

[17] A summary of the extent and nature of the contractual model that was dominant as of 1986 is D.Pickrell and C. Oster, *A Study of the Regional Airline Industry: The Impact of Marketing Alliances*, U.S. Department of Transportation, May 1986. Since then, marketing alliances have become even more important and range in nature from contracts to mutually sell separately branded individual flights with limited financial transfers to comprehensive global networks like the Star Alliance, Skyteam and OneWorld, further containing within them large joint ventures combining financial results on many routes. In the United States, antitrust constraints prevent close coordination between airlines deemed to be domestic competitors, but various cooperative

from the regulated period affected these decisions by affecting the choice of aircraft size, the need to respond to competition not subject to legacy labor agreements became much more important[18]. In shaping their networks and aircraft fleets, airlines became free to contract with firms offering complementary service and could use these relationships to adapt to both the freedom to restructure their networks and to the continued influence of labor contracts. Agreements with other firms affecting domestic aviation were subject to antitrust limitations, but did not require regulatory approval.

Extensive "hub and spoke" route networks were created in which passengers originating at various (the "spokes") were flown to a central point (the "hub") and then combined with other passengers destined for the same airport, thus overcoming diseconomies of scale and allowing frequent service in airport-pair markets that couldn't support nonstop service at competitive frequencies (or at all). Passengers originating at the hub were offered much more nonstop service on many more itineraries than could be supported without combining connecting and local passengers. Aircraft of different sizes could be combined to offer service on different itineraries at different times. While the unit costs of carrying passengers through a hub were higher than of carrying them nonstop with all other things being equal, all other things were not equal -- relatively few markets were dense enough to support competitively-frequent nonstop service with large enough aircraft to produce acceptable costs. The hub could be regarded as a device to circumvent indivisibilities, the connection paradoxically *lowering* unit costs when compared with attempting to carry all passengers nonstop.

As noted, airlines operating large aircraft in extensive networks chose to abandon the cooperation with their peers on uniform terms that regulation had previously required while developing diverse new contractual relationships. Most importantly for this paper, they changed their terms of cooperation with airlines providing service in smaller aircraft, either to smaller cities or on lightly-traveled flights. Airlines were free to make their own production, marketing

---

arrangements between firms operating aircraft of different sizes or in different markets has emerged as detailed below.

[18] While the arbitration award (note 14 *supra*) and the contracts resulting from it were not addressed by the Airline Deregulation Act of 1978, the award has continued to influence very strongly the provisions of pilot labor contracts both of older firms and new entrants. Virtually all older firms have either had to find some way within this system of compensating pilots or to contract out the service. In the same way, although pattern bargaining was disrupted by deregulation and some new entrants have attempted to operate with contracts that didn't conform to that award by setting flat rates regardless of aircraft type flown, most of those "deviant" contracts have proved unsustainable in subsequent labor bargaining, although new-entrant competition has certainly affected compensation, "scope" (which elements of the network are covered by which contract) and productivity constraints.

and distribution arrangements and chose divergent paths. In short, the freer regulatory environment generated a great deal of experimentation in firm structure and contracting and new competition forced previously existing airlines to adapt or die.

Pricing was also deregulated. Since there were many common costs, a complex fare structure with price discrimination emerged that didn't depend on market power[19], ultimately managed on each network by "O&D-based yield management" systems that maximized the incremental value to the network of each proposed itinerary and decided what price to offer. In the example described above, the program would analyze the expected value of the Dothan-Atlanta seat on that day at that time, considering the history of demand on that segment, do the same for Atlanta-Tokyo, and then decide at what price it would accept a Dothan-Tokyo passenger who would occupy both seats. If both seats were within the control of the network, this was a fairly straightforward exercise, but if a different airline were making the decision for each segment, an externality would be created because without further contractual intervention, each would only consider the value to it of the seat it was allocating to the passenger and not the value to its interline partner. And the value to each airline varied constantly with the flow of reservations and economic conditions and was closely-held for commercial reasons.

The advent of hubs also presented a new operating challenge. Operations became much more interdependent, since the success of the hub depended on getting all the aircraft in the "bank"[20] on the ground and then off again with sufficient precision to create the required connections within a length of time that made connecting journeys acceptable to passengers and used capital like gates and aircraft as efficiently as possible. A delay at a spoke meant missed opportunities at a hub or required expensive "slack" in the use of aircraft and gates to ensure the connections worked. Even if spoke flights were on time on average, variation in performance imposed expense on the system as it accommodated the variation. And finally, the passenger experience involved two boardings, a baggage transfer and planning around departure and arrival times.

As with pricing, airlines developed complex systems and institutions to coordinate and sequence arrivals and departures, maintaining the integrity of the hub. This was difficult enough when all the coordination was in one firm,

---

[19] M.E. Levine, "Price Discrimination Without Market Power", 19 *Yale Journal on Regulation* 36 (2002)
[20] A "bank" is each iteration of the connection opportunity afforded by having aircraft on the ground simultaneously.

but when more than one firm was involved, externalities were created as operating performance and decisions on one airline affected the performance of the other and the overall passenger experience. An airline considering whether to delay a departure or expend extra resources to depart on time might not consider the costs to the airline to which its passengers were connecting, nor could it consider cost-minimizing compensating adjustments on the other airline when it had no control over its operations. The costs of service recovery (substitute flights, hotel rooms, goodwill compensation) were also created jointly but unless they were under the control of a single firm, they would not necessarily be taken into account by the operator who could most effectively control costs and benefit from goodwill.

All of this meant that deregulation drastically changed the economics of coordination. It created new incentives and a need for new strategies in network design (cities served, routes, schedules and aircraft assignments), airport operations, network control systems, pricing systems and for systems that coordinated the efforts of differing elements making up the network. It also created new opportunities for distribution and promotion that also exhibited common costs and had impacts on each transaction in the network. Airlines started the deregulated era with independent fragmented firms being reassembled into integrated networks. At first, airlines were separately branded and separately scheduled and operated, sharing revenue by variations of the prorate formula[21] that had been operative under regulation. Contracts between them involved these prorates, general language that connections would be coordinated and, often, agreements about distribution and marketing. As frequent flyer programs and pricing promotions became important, those were often covered by agreement as well.

At the same time, big unionized network airlines were subject to labor constraints that affected their ability either to fly the smaller aircraft needed for many network spokes or to cooperate with other airlines who could fly such aircraft more economically. Pilots especially were paid using contracts that enshrined the historic arbitration award [22] that varied pay according to the size and speed of the aircraft they flew. The variation in size and speed from the smallest and slowest aircraft on the network (for example, a turboprop with 19

[21] A "prorate formula" shares revenue between two airlines for a jointly-produced journey. Historically, this was done either by a ratio of local fares (which were set by regulation and hence not subject to individual manipulation), by mileage, or by some combination of both. The deregulation of fares greatly complicated these negotiations both by proliferating the number of fares offered for any given journey and by making them subject to unilateral manipulation in response to any agreed general formula.

[22] *See supra* notes 16 and 18 and accompanying text.

seats) and the biggest and fastest (a 300-500 seat jet) was enormous. Pilots who flew the biggest and fastest aircraft over long distances received very substantial rents (such flying was usually *less* demanding[23] than flying smaller aircraft on shorter hops) and such flying was allocated within airlines by seniority, regulated by union agreements.

Pilot unions are political democracies whose economic power depends on maintaining cohesion. In order to maintain cohesion, get voting approval for proposed contracts and to avoid replacement during strikes by more junior pilots, senior pilots flying more lucrative aircraft transfer some of their rents to more junior pilots in the form of pay scales that either are not fully proportional or "cut off" at a certain minimum aircraft size that will produce acceptable compensation for junior pilots. In turn, seniority assignment means that most junior pilots will get to fly larger aircraft and earn more rents later in their careers provided that they stay employed by the same airline.

This system creates incentives for airlines to outsource flying in smaller aircraft to less-experienced pilots outside the contract labor force, or to airlines with lower-cost contracts. To preserve junior members' jobs[24], pilot unions negotiate contracts that include "scope" clauses that limit an airline's ability to outsource work to other airlines. The limits vary a bit from airline to airline, but are some combination of restrictions on the extent of use of the network's "code"[25] on other airlines, size limits on the size and number of aircraft that can be flown by other airlines using the network's code, on the proportion of flying that can be outsourced to partners and quotas and limits on ownership by the large network airlines of other airlines to which work might be outsourced or with which marketing might be coordinated.

The interdependencies, externalities and labor contracts described above soon rendered arms-length piecework contracts between airlines obsolete for most network flying, although some such contracts remain even today on routes where interdependencies are minimal and the impact on labor is small or nonexistent. For coordination between flying on dense routes supporting large aircraft at most times of day and on those requiring the use of smaller aircraft, a

[23] The aircraft were more highly automated, flew in less congested airspace and required fewer takeoffs and landings to accomplish a month's work.
[24] Pilot contracts preserve strict seniority to determine who flies which flights and who gets laid off.
[25] The alphanumeric that identifies a flight as belonging to a network.

variety of organizational options have emerged in their place as experience in the deregulated market has accumulated[26].

These have included:

1) All flying, all operations, network management, revenue management and marketing are done directly inside a single firm. This arrangement has been tried but is very rare, for two reasons: the union politics of having all pilots in one local become very difficult and the overhead structure suitable for the regional operation is usually quite a bit "leaner" than that required to coordinate a large global network.

2) Direct management of many functions is done inside the "main" firm (network, schedule, pricing and inventory management, operational coordination), while others (operations, maintenance, some station staffing, FAA certification and supervision) are outsourced to a wholly owned firm with its own labor arrangements. This arrangement is fairly common for at least part of most large networks and has been used, abandoned and re-used since deregulation by almost every large network airline. As of 2011 it is in use by American, Delta, U.S. Airways, and Alaska for varying proportions of their operations. In this arrangement, pilots usually belong to a different union local from that of the large network airline and have different internal politics. Some versions of this organization put the subsidiary's pilots in on different contractual terms but a single seniority list or give pilots at the subsidiary first preference in hiring at the parent. In pilot terms[27] this preserves subgroup polities and politics, but ultimately reduces the distinction between 1) and 2). There are few or no externalities or holdup problems between firms in this arrangement, although there may be management and coordination problems related to firm culture and management of quality and service regularity.

3) Direct management of some functions inside the main network, others governed by contract with a firm partly or wholly owned by shareholders other than the network firm. For example, scheduling and fleet decisions may be made by the network partner and all revenue kept by the network partner, with payment to the regional partner made on the basis of production output. This

[26] Similar organizational options exist between networks involving aircraft of similar size, but they are limited by different factors, including international constraints. Many of the principles, influences and options are the same as those in this Article, but my focus here is on domestic networks involving large and small aircraft.
[27] Labor arrangements for pilots are almost always on different terms from those of other employees because since pilots are much more difficult to replace, they have much greater financial leverage. At a number of airlines, only the pilots are unionized and where other groups are unionized, pilots are always unionized separately.

eliminates many externalities, but doesn't eliminate holdups or opportunistic behavior on the part of either partner. Those are addressed by a variety of other terms in the contract: This form is used to varying degrees by virtually all of the major network carriers. A number of firms, such as Skywest, specialize in providing services using this arrangement to multiple major network customers who do not own shares in them.

In recent years, the arrangements in paragraph 1) above have virtually disappeared, in part as a response to the difficulty of maintaining efficient labor arrangements for very different kinds of flying and aircraft within a single firm, in part as a reflection of the difference in management infrastructure and overhead most efficient to manage the activity involvedThe arrangements of paragraph 2) and 3) have come to be dominant, with a tendency most recently to move from 2) to 3)[28], although each has its own advantages and disadvantages. The diverse ownership arrangements have the advantage allowing a credible threat of complete separation to enforce goals, but of course preserve incentive asymmetry. They are particularly useful in labor bargaining, since the labor force of the regional subsidiary must face the possibility of loss of the contract if they cannot come to terms efficient enough to allow financial success. (The option of closing down a wholly-owned subsidiary also exists, but is much more complicated.) On the other hand, diverse ownership arrangements create holdup incentives, as noted below.

In these arrangements, passenger revenue is the responsibility of the major airline network, which does all the pricing, marketing, "yield management" (a sophisticated form of price discrimination to allocated jointly-produced capacity to different customers on different itineraries on different terms and conditions) and compensation of third parties for directing passengers to the flights. The regional airline is compensated for its costs of those production inputs that are not provided by the major network. These always include labor and generally include operating maintenance, and may or may not include facilities rental, flight equipment capital costs, ancillary ground equipment, etc., depending on the comparative advantages of the firms, locations and the possibility of joint purchase and production. Compensation is usually in the form of an agreed rate per unit of output (available seat-miles, departures, etc.),

[28] The largest operator of a wholly-owned regional airline at the time of this writing is American, which recently announced its intention to move to long-term contracting with a firm owned by others (model 3) above. *See* AMR Says It May Initiate Spinoff of American Eagle, *The New York Times,* August 12, 2011, Sec. B, p.6.

with incentives for efficiency and some buffers against economic shock from external cost or political factors.

Since there is much relational specificity in the functional arrangements, the contracts are typically for long terms (ten years is common) to try to align incentives. Price terms are usually partly indexed to observable costs, like fuel (although fuel can also be provided at many locations by the major airline), inflation, etc., but tension exists and conditional price renegotiation is usually provided for. The long life of the contract provides an incentive for relationship-building behavior. It allows some assurance that both parties, but especially the regional airline, will be motivated to make complementary investments not directly provided for in the contract like labor practices designed to develop human capital in the workforce, operating practices designed to enhance customer loyalty, and capital investments in maintenance and ground equipment designed to lower long-term costs. In addition, long-term contracts give both parties incentives to invest to adapt operations to suit changing circumstances in a dynamic business greatly affected by external economic and political factors.

Fleets are often leased to the regional by the large network or the leases from third parties to the regional are guaranteed by the network. Under those circumstances, the large network defends itself from asset-specific opportunism by retaining the right to remove the fleet if the relationship is terminated and place it with another regional firm. Of course, withdrawal of the fleet and placement of the aircraft with another partner is a complicated and rather drastic measure creating real risks of operational failure, so it cannot be casually used by the major network. And the ability to do so simultaneously reinforces and mitigates some of the relational incentives described in the previous paragraph.

Termination of the relationship is used as a quality control mechanism with agreed-on quality metrics creating a right, after notice and an opportunity to cure, to terminate the relationship. Since interruption of service violates the quality metrics, these clauses also help protect both the major network and the regional airline from labor holdups. Again, there is an inherent tension between the use of such incentives and the difficulty -- impossibility really -- of specifying completely metrics that capture all of the salient features of the relationship. This contradiction between the need for long-term contracts to provide incentives to invest in the relationship and the critical importance to its success of factors which cannot be completely specified (such as day-to-day compatibility of organizations and personnel in a fluid environment that

requires close cooperation) underlies and underlines the complexity of the contracting problem. The larger network partner is vulnerable both to quality failures that will damage its reputation the loss of complementary business. The smaller partner, which is dependent on the larger partner for all or a large part (it may have multiple partners) of its revenue may find termination an existential threat.

Information technology (IT) presents this dilemma in particularly acute form. Almost all IT systems exhibit very large economies of scale, so that absent holdup problems network systems would use the same technology for both revenue and production. On the other hand, they are very difficult and expensive to replace, so that contract termination involving IT controlled by the large network partner becomes particularly costly for the smaller airline and severely limits its options in the event of dispute. Since revenue generation and coordination is an important feature of these contracts, has the largest impact by far on the larger network and involves information most or all of which is the hands of the network partner, these contracts almost always place revenue entirely in the hands of the network partner and revenue systems remain entirely under their control. Operating IT often remains in the hands of the regional partner, even though in principle it might be less costly to allow the larger partner to operate this system as well.

Finally, separate incorporation as in 2) and more arms-length arrangements as in 3) allow much more flexibility in bankruptcy proceedings, because it is not necessary to put the whole network into bankruptcy if one of the entities fails. In the case of 2), this allows more flexibility in capital structure. In the case of 3), a regional partner doesn't necessarily have to reorganize or liquidate if a network partner goes into bankruptcy, nor does the bankruptcy of the regional partner necessarily expose assets of the network partner to its creditors.

Major network airlines have largely been motivated to move from 2) to 3) for several reasons:

First, as a way of raising capital, because the form of the contract makes the revenue stream of the regional partner less variable than that of the network partner. This attracts investors with a different risk profile, allowing capital to be raised from some sources that are unwilling to invest in the network partner and who regard the termination risk as minimal or tolerable (because their capital contribution is secured by marketable assets).

Second, as a way of emphasizing the separation between pilot union locals at the network and regional airlines.

Third, as a way of allowing the network partner to contract with several regional airlines, credibly using potential replacement by one of the others as bargaining leverage with the unions and shareholders of the regional firm.

In the same way, regional partners consent to such arrangements such as 3) because independent ownership very often entails the right to provide similar services to another major network, using its trademarks, pricing and scheduling systems for that part of the regional airline's operation. This gives the regional airline portfolio diversification and bargaining leverage against the large network partner. On the other hand, one of the attractions of 2) to the major network airline is that it can be the exclusive contractual partner. This provides it with leverage over the management and labor at the regional carrier, but of course all the investment in the game ultimately comes from the large network partner.

At various times since deregulation, individual network airlines have used both 2) and 3), depending on labor relations, the need for capital and their ability to get acceptable quality of operation from the regional airline. As of 2011, every major network uses some form of 3), but only United uses it more or less exclusively and that arrangement has emerged only since its last Chapter 11 reorganization.

In another, most "distant", variant of the firm/contract tradeoff, an independent regional airline manages all of its operational functions, negotiates its schedules and negotiates its revenue sharing arrangement with the large network airline. Marketing identification can be independent of the major network, but codesharing (identification in automated systems of the regional airlines" flights with the code of the large network) and frequent-flyer arrangements are essential to providing the network benefits that motivate the arrangement in the first place.

The relationship is controlled by contract, often with bargaining leverage apportioned on both sides by aircraft provision, contract length and termination clauses. These arrangements can be very difficult to manage and usually exist only where a portion of the large network can be limited in its geographic and operational impact. Examples have included services to defined leisure destinations, such as Cape Cod and its islands, small cities and islands in the Pacific Northwest, the Hawaiian Islands, the Florida Keys and

Alaska. The success of these arrangements is usually determined by informal credibility and the mutual exchange of hostages, rather than the terms of the contract. Reputation matters a great deal.

Perhaps the most interesting real-world example of the influence of regulation on the evolution and differences in relationships and their consequences can be found in the history of Alaska Airlines, a medium-sized network that does not have either the scope and scale of the larger network carriers but has a network that consists of two distinct collections of smaller markets – those in the Pacific Northwest and those in Alaska. Its history is instructive. It started as a separate regulatory category, with scheduled passenger service limited to flights to, from and within Alaska. It served routes from Seattle to the major Alaska cities with aircraft as large as those being used by "trunk" airlines and ultimately competed with three of them on routes to Anchorage, Fairbanks and Juneau. It connected those cities to many smaller cities, towns and hamlets in Alaska on routes that had relatively few passengers but often involved significant amounts of cargo. At various times, it experimented with cargo and charter service outside Alaska with large aircraft in an effort to find a successful financial formula, but its historic focus was on service to and within Alaska.

For service within Alaska, it used smaller aircraft of the type used by regional airlines, as well as jet aircraft that carried both cargo and passengers. These were type 1) arrangements as described above. Alaska Airlines was subsidized, as were other regional airlines in the United States but according to somewhat different arrangements. After statehood, the political influence of Alaska's Congressional delegation allowed subsidies to continue and supported the expansion of jet service into much smaller markets than supported jet service in the "Lower 48". Its Congressional clout assured that such service continued to be subsidized even after deregulation.

Because Alaska Airlines was relatively small, heavily subsidized and flew relatively big aircraft on routes whose indivisibility inefficiencies were supported by regulation and subsidy, it was expensively unionized. As long as its ambitions and possibilities were constrained, this was not a major concern as labor costs could be passed along to customers and taxpayers. It was frequently in financial difficulty and experienced two dramatic changes in management, but its regional role and Congressional protection after statehood meant that there was never a serious possibility that its service would be terminated. This greatly limited its leverage with labor

After deregulation, Alaska began to expand, first into the Pacific Northwest, with a system with hubs at Seattle and Portland fed from Alaska and the smaller cities in the Northwest. It also began expanding with limited service into California. As time went on, it discovered that the evolving network character of the industry required that it continue to expand its scale and scope in order to allow its large-jet service to remain competitive. Deregulation was producing low-cost competitors on its principal Lower 48-Alaska routes, as well as on its new routes in the "Lower 48". To lower its costs to compete, it needed to expand its scale and to use more relatively large aircraft on more dense routes. To feed its hubs at Seattle and Portland, it expanded service into the Lower 48, first more extensively into California, then into Arizona and Mexico (much of this to accommodate leisure demand from Alaska as well as local demand) and more recently into major cities in the Midwest and East. This expansion has also taken place inside the firm.

The regional service in the Pacific Northwest could not support the labor costs of the main network operation (which itself struggled to reform its labor contracts and reduce its costs to compete with deregulated Lower-48 competitors). That regional service competed with regional service offered by deregulated partners of other deregulated network airlines. Accordingly, the regional service was placed by contract in Horizon, an owned subsidiary acquired in 1986 and operated as a wholly-owned hybrid of type 2 and 3, with elements of type 4. Alaska was forced into the form of this arrangement by the need to separate lower-cost labor contracts for Horizon and in order to avoid having its higher costs (themselves being reduced in contentious negotiations with its unions) migrate down into wholly-owned Horizon, Alaska was forced to keep it at arm's length. A years-long struggle ensued to maintain the level of coordination involved in maximizing the value of the combined network while keeping marketing and operations separate enough to maintain independence under the labor laws.

In the meantime, Alaska discovered that there was considerable value in having its Alaska regional system fed by a much larger network than it was prepared to risk trying to create by itself. Accordingly, it developed relationships with much larger network airlines such as Delta (originally as Northwest) and American, with most of the connections at Seattle. To preserve its negotiating leverage, it has maintained its own brand and identity, using codeshares[29] , along with frequent flyer program agreements (that benefit from economies of

[29] Flights on other airlines with multiple IT identity, either Alaska's IT code on other airline flights or other airlines' code on Alaska's flights.

scope) and marketing agreements. It has used its near-monopoly dominance of its "home" markets to and in Alaska to force large network airlines to accept less-than-exclusive arrangements and less-than-optimal revenue coordination in order to protect itself from vulnerability to holdups and to preserve its own strategic independence. These arrangements are all of the last type in the typology introduced above.

Over time, the arrangements with Horizon became increasingly awkward from a coordination standpoint and competitive pressure made the marketing integration of its network imperative. In August of 2010, Alaska converted its relationship with Horizon to a type 2 arrangement[30], and in January 2011, it announced that it was retiring the Horizon brand and labeling all of its flights "Alaska" [31]

Alaska's history is introduced here to highlight the absence of a single optimum airline network form and the influence of regulation on the relationships between large network airlines and their regional complements. What is especially striking is that all the forms outlined above have been tried by many airlines and virtually all except the single firm have persisted. They have co-existed and do coexist as of 2011 simultaneously in the industry and in individual networks, which are very often a mix of 2), 3) and arms-length arrangements. Some methods have been dominant at some points in time, but there has never been a time in which all networks have converged on any one method, and networks that have experimented with using only one form have always diversified over time.

For airline networks, the determining forces are the tension between the need for coordinating production and marketing on one hand to attract and retain revenue and the need for protecting the network from vulnerability in the supply of labor and capital inputs on the other. Union contracts for pilot labor remain a very important factor in influencing organizational form. Regulatory history has determined the degree of variation among firms and the starting point from which divergence has occurred. That history continues to influence labor pilot labor contracts as well.[32] A secondary but important factor is the

---

[30] "Horizon Air Changes Business Model". Aviation Week. 23 August 2010. http://www.aviationweek.com/aw/generic/story_generic.jsp?channel=aviationdaily&id=news/avd/2010/08/23/08.xml&headline=Horizon%20Air%20Changes%20Business%20Model. Accessed July 13, 2011.

[31] USA Today, January 26, 2011. http://travel.usatoday.com/flights/post/2011/01/horizon-air-alaska/140151/1 Accessed July 13, 2011.

[32] In that connection, it is interesting to note that the structure of pilot labor contracts has been so robust that even many new entrant or post-reorganization firms that have had the opportunity to start fresh with a new labor contract have been forced by pilots to adopt the old form as they grow. For example, Continental

difference in the degree of management overhead necessary to operate a regional airline as compared with a large network airline.

From the Coasean perspective, this is an extraordinarily interesting result. It suggests, as Coase has urged, that the particular circumstances surrounding any production requirement are of critical importance in determining the organizational form that production (and marketing, in the real world where price and product information is not costlessly available) will take. Coase's and Williamson's theories tell us where to look, but they do not tell us what we will find.

---

emerged from a bankruptcy with a single rate of pay, but has adopted the previous form. One fairly large new entrant, JetBlue, still does not have a pilot union on the property (it has defeated attempts to organize one), but it has adopted differential pay rates for different aircraft types. A few large networks like Southwest use a single aircraft type (albeit in units of different capacity) and a single pay rate, thus allowing its union politics to avoid the issue. One relatively large network, AirTran, has a union but a single rate of pay on two different aircraft types. But generally contracts that reflect the 1934 arbitration award are in use at virtually all U.S. networks of any size and indeed most airlines in Europe as well. Clearly pilot internal politics play an important role in determining the structure of pilot labor contracts.